

University of Denver

Digital Commons @ DU

---

Electronic Theses and Dissertations

Graduate Studies

---

1-1-2016

## Follower and Extender Sets in Symbolic Dynamics

Thomas Kelly French  
*University of Denver*

Follow this and additional works at: <https://digitalcommons.du.edu/etd>



Part of the [Mathematics Commons](#)

---

### Recommended Citation

French, Thomas Kelly, "Follower and Extender Sets in Symbolic Dynamics" (2016). *Electronic Theses and Dissertations*. 1152.

<https://digitalcommons.du.edu/etd/1152>

This Dissertation is brought to you for free and open access by the Graduate Studies at Digital Commons @ DU. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Digital Commons @ DU. For more information, please contact [jennifer.cox@du.edu](mailto:jennifer.cox@du.edu), [dig-commons@du.edu](mailto:dig-commons@du.edu).

# Follower and Extender Sets in Symbolic Dynamics

---

A Dissertation

Presented to

the Faculty of Natural Sciences and Mathematics

University of Denver

---

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Thomas French

June 2016

Advisor: Dr. Ronnie Pavlov

©Copyright by Thomas French 2016

All Rights Reserved

Author: Thomas French  
Title: Follower and Extender Sets in Symbolic Dynamics  
Advisor: Dr. Ronnie Pavlov  
Degree Date: June 2016

# Abstract

Given a word  $w$  in the language of a one-dimensional shift space  $X$ , the follower set of  $w$ , denoted  $F_X(w)$ , is the set of all right-infinite sequences which follow  $w$  in some point of  $X$ . Extender sets are a generalization of follower sets (introduced in [5]) and are defined similarly. To a given shift space  $X$ , then, we may associate a follower set sequence  $\{|F_X(n)|\}$  which records the number of distinct follower sets in  $X$  corresponding to words of length  $n$ . Similarly, we may define an extender set sequence  $\{|E_X(n)|\}$ . The complexity sequence  $\{\Phi_X(n)\}$  of a shift space  $X$  records the number of  $n$ -letter words in the language of  $X$  for each  $n$ . This thesis explores the relationship between the class of achievable follower and extender set sequences of one-dimensional shift spaces and the class of their complexity sequences.

Some surprising similarities suggest a connection may exist, for instance, both the complexity sequence and the extender set sequence are bounded if and only if there exists some  $n$  such that the value of the  $n^{\text{th}}$  term of the sequence is at most  $n$ . This thesis, however, also demonstrates important differences among complexity sequences and follower and extender set sequences of one-dimensional shifts. In particular, we show that unlike complexity sequences, follower and extender set sequences need not be monotone increasing. Finally,

we use the classical  $\beta$ -shifts to demonstrate that, while many follower set sequences may not be realized as complexity sequences, up to possible increase by 1, any complexity sequence may be realized as a follower set sequence of some shift space.

# Acknowledgements

I would like to thank my advisor, Dr. Ronnie Pavlov. This thesis, and indeed my entire graduate career, could not have been possible without his expert help, guidance, and tireless dedication. I would also like to thank the other members of my thesis committee, Dr. Nic Ormes, Dr. Mei Yin, Dr. Paul Horn, and Dr. Scott Leutenegger, whose comments and questions have helped to make this thesis more accurate, more precise, and more readable. Finally, I thank my family: My mother Emma encouraged me to apply to graduate school in the first place, and so began the long chain of events that lead to this achievement. My father Skip has always believed in me and never once uttered a discouraging word. My wife Laura has tolerated many long nights of studying without complaint, and frequently assured me that I had good ideas, and was not wasting my time. Without any of them, I would be lost.

# Table of Contents

Acknowledgements . . . . .	iv
List of Figures . . . . .	vi
<b>Chapter 1: Introduction</b>	<b>1</b>
<b>Chapter 2: Definitions and Examples</b>	<b>9</b>
<b>Chapter 3: Follower and Extender Sets</b>	<b>22</b>
3.1 Follower Sets and Complexity . . . . .	22
3.2 Decreases in Follower and Extender Set Sequences . . . . .	36
3.3 Follower Sets, Extender Sets, and $\beta$ -shifts . . . . .	64
<b>Bibliography</b>	<b>83</b>

# List of Figures

2.1	The graph $\mathcal{G}$ presenting the even shift . . . . .	16
2.2	A graph which is right-resolving, but not left-resolving, primitive, or follower-separated . . . . .	18
3.1	Step 1 for constructing $\mathcal{G}_{\ell,S}$ . . . . .	40
3.2	Step 2 for constructing $\mathcal{G}_{\ell,S}$ . . . . .	41
3.3	Step 3 for constructing $\mathcal{G}_{\ell,S}$ . . . . .	41
3.4	$\mathcal{G}_{5,\{0,3\}}$ with $i^* = 3$ (where we use the convention that edges terminating at “Back to Start” return to the initial vertex of the edge $p$ .) . . . . .	42
3.5	The graph $\mathcal{G}$ constructed as in Theorem 3.2.5 . . . . .	47
3.6	$\mathcal{G}$ such that $X_{\mathcal{G}}$ has a different number of predecessor and follower sets . . . . .	64
3.7	$d_{\beta}^*(1)$ in the case of Lemma 3.3.10 . . . . .	72



# Chapter 1

## Introduction

The field of dynamical systems concerns itself with the study of the long-term behavior of systems which change over time. A **topological dynamical system** is a pair  $(X, T)$ , where  $X$  is a compact space and  $T$  is a homeomorphism of  $X$ . In symbolic dynamics, we let  $(X, T)$  be a **shift space**. Given a finite alphabet of symbols  $\mathcal{A}$ , we say that  $(X, T)$  is a **one-dimensional shift space on  $\mathcal{A}$**  if  $X \subseteq \mathcal{A}^{\mathbb{Z}}$  and  $T$  is the shift map  $\sigma$ , so that for any point  $x = \dots x_{-2}x_{-1}.x_0x_1x_2\dots$  in  $X$ ,  $\sigma(x) = \dots x_{-1}x_0.x_1x_2x_3\dots$ . In order for  $(X, \sigma)$  to be a dynamical system,  $(X, \sigma)$  must satisfy two properties: First,  $X$  must be closed under  $\sigma$ , so that  $\sigma(X) = X$  and second,  $X$  must be closed in the product topology. It is clear then that  $X$  is compact by Tychonoff's Theorem, and  $T$  is certainly a homeomorphism. The requirement that  $X$  be closed in the product topology means that  $X$  must be closed under limits, where we consider two points to be close if they agree on a very long word about the origin. So, then, if  $X \subseteq \{0, 1\}^{\mathbb{Z}}$  is the set of all sequences containing exactly one 1 symbol, then  $(X, \sigma)$  is not a shift space, because we can approach the

point  $0^\infty$  of all zeros by shifting the 1 symbol further and further from the origin, and since  $0^\infty$  is not in  $X$ ,  $X$  is not closed in the product topology. In contrast, if  $X \subseteq \{0, 1\}^{\mathbb{Z}}$  is the set of all sequences containing *at most* one 1 symbol, then  $(X, \sigma)$  is a shift space—no second 1 symbol may be introduced by shifting or by taking limits.

Symbolic dynamics may be done in higher dimensions, with  $X \subseteq \mathcal{A}^{\mathbb{Z}^d}$  for some  $d \in \mathbb{N}$ , but the results of this thesis are specific to one-dimensional shift spaces. For simplicity, we will often refer to the shift space  $(X, \sigma)$  only by  $X$ .

A one-dimensional shift space  $X$  is **sofic** if there exists a finite directed labeled graph  $\mathcal{G}$  such that  $X$  is exactly the set of sequences of labels of bi-infinite walks on  $\mathcal{G}$ . In this case, we say that  $X$  is the **edge shift presented by  $\mathcal{G}$** , and we may denote  $X$  by  $X_{\mathcal{G}}$ . A simple example of a sofic shift is the even shift, presented by the graph in Figure 2.1, on page 16. In that graph, we may only see the label 1 when leaving the left vertex, and any path from the left vertex to itself must feature an even number of 0 labels. Thus, in the even shift, any string of 0's between two 1's must have even length. It is a simple exercise to check that this single restriction is enough to characterize exactly the set of all labels of bi-infinite walks on the graph.

The set of all words appearing in some point of  $X$  is called the **language** of  $X$ , denoted  $L(X)$ . The complexity function of a shift  $X$ ,  $\Phi_X(n)$ , counts the number of words of a given length  $n$  in  $L(X)$ . Thus, the **complexity sequence**  $\{\Phi_X(n)\}_{n \in \mathbb{N}}$  records the number of distinct  $n$ -letter words occurring in  $X$  for each length  $n$ . This sequence is natural to study; in particular, it may be used to calculate topological entropy of symbolic shifts. Our goal for this thesis will be a comparison between complexity sequences and other

types of sequences that we might associate to a one-dimensional shift space. The Morse-Hedlund Theorem (see [8]) implies that if there exists  $n \in \mathbb{N}$  with  $\Phi_X(n) \leq n$ , then every sequence in  $X$  must be periodic. This result is tight, for Sturmian shifts have  $n+1$  words of length  $n$  for every  $n$ , and yet contain no periodic points (see [2]). Furthermore, it is clear that the complexity sequence must be monotone increasing, as for each  $n$ -letter word  $w$  in  $X$ , there exists at least one  $n+1$ -letter word appearing in  $X$  having  $w$  as a prefix. These two properties of complexity sequences—that they must stay above  $n$  at the  $n^{\text{th}}$  term in order to be unbounded, and that they are monotone increasing—will guide our discussion; the first of these properties will extend to the other sequences we will define in this thesis, but we will show that the second cannot.

Given a word  $w$  appearing in some point of the shift space  $X$ , we define the **follower set of  $w$** , denoted  $F_X(w)$ , to be the set of all right-infinite sequences  $u \in \mathcal{A}^{\mathbb{N}}$  such that the infinite word  $wu$  appears in some point of  $X$ . That is,  $u$  may legally follow  $w$ . Alternatively, the follower set of  $w$  may be viewed as the set of all *finite* words  $u$  such that  $wu$  appears in some point of  $X$ . These two definitions are analogous—if  $u$  is a finite word which may follow  $w$ , then it clearly occurs as a prefix of some right-infinite sequence which may follow  $w$ , and if  $u$  is a right-infinite sequence which may follow  $w$ , then all of its finite prefixes certainly may legally follow  $w$ . While this thesis will use the definition involving right-infinite sequences, all results will apply to either definition.

It is a well-known equivalence that a shift space  $X$  is sofic if and only if there are only a finite number of distinct follower sets of words in  $X$ . (See [7] for a proof). For instance, in the even shift, the follower set of a word  $w$  depends only on the parity of the number of 0's following the last 1 in the

word  $w$ . If  $w$  ends with a 1 followed by an even number of 0's, then  $w$  may be followed by any sequence beginning with an even number of 0's followed by a 1, or by  $0^\infty$ . If  $w$  ends with a 1 followed by an odd number of 0's, then  $w$  may be followed by any sequence beginning with an odd number of 0's followed by a 1, or by  $0^\infty$ . Finally, if the symbol 1 does not occur in the word  $w$ , then  $w$  may be followed by any legal right-infinite sequence. That the even shift has only three follower sets reinforces our earlier assertion that the even shift is sofic.

Similarly, we define the **predecessor set of  $w$** , denoted  $P_X(w)$ , to be the set of all left-infinite sequences  $s \in \mathcal{A}^{-\mathbb{N}}$  such that the infinite word  $sw$  appears in some point of  $X$ . That is,  $s$  may legally precede  $w$ . As before, a definition replacing infinite sequences with finite words is analogous.

The **extender set of  $w$**  is a generalization of the follower set, first introduced in [5]. The extender set of  $w$ , denoted  $E_X(w)$ , is defined to be the set of all pairs  $(s, u)$ , where  $s$  is a left-infinite sequence,  $u$  is a right-infinite sequence, and  $swu$  is a point of  $X$ . Simply put, given a finite word  $w$ , we may think of the extender set of  $w$  as the set of all possible ways to extend  $w$  into a complete point of  $X$ . Once again, a definition replacing infinite sequences with finite words is analogous. Unlike in the case of follower sets, the definition of extender sets may be generalized to shift spaces on higher dimensions. Furthermore, as with follower sets, a shift  $X$  is sofic if and only if there are only a finite number of distinct extender sets of words in  $X$ . (See [9] for a proof).

We now define, for any  $n \in \mathbb{N}$ , the set  $F_X(n)$  as the set of all distinct follower sets corresponding to words of length  $n$  in  $X$ . (Similarly,  $P_X(n)$  is

the set of all distinct predecessor sets corresponding to words of length  $n$  in  $X$  and  $E_X(n)$  is the set of all distinct extender sets corresponding to words of length  $n$  in  $X$ ). Then the sequence  $\{|F_X(n)|\}_{n \in \mathbb{N}}$  records the number of distinct follower sets of words of length  $n$  for every  $n$ . We call this the **follower set sequence** of the shift  $X$ . (The **predecessor set sequence** and **extender set sequence** of  $X$  are defined similarly). Chapter 2 of this thesis provides more details and examples of these definitions, along with other important definitions that we will need. Chapter 2 also includes more detailed explanations of many of the examples provided in this introduction.

Chapter 3 of this thesis explores the relationship between the class of achievable follower, predecessor, and extender set sequences and the class of achievable complexity sequences of one-dimensional shift spaces. Fixing a specific one-dimensional shift  $X$ , its complexity sequence may be completely unrelated to its follower, predecessor, and extender set sequences. Consider the full 2-shift  $X_{[2]} = \{0, 1\}^{\mathbb{Z}}$ . It is easy to see that for any  $n$ ,  $\Phi_{X_{[2]}}(n) = 2^n$ . Yet there are no restrictions on what words may be adjacent; any word on  $\{0, 1\}$  may legally follow or precede any other word on  $\{0, 1\}$ , and so every word in  $L(X)$  has the same follower, predecessor, and extender set as every other word in  $L(X)$ . Hence,  $\{|F_X(n)|\} = \{|P_X(n)|\} = \{|E_X(n)|\} = \{1, 1, 1, \dots\}$ . However, a result of Ormes and Pavlov ([9]) suggests that the class of all achievable complexity sequences and the class of all achievable extender set sequences may be connected. Their result states that if there exists any  $n \in \mathbb{N}$  such that  $|E_X(n)| \leq n$ , then the shift  $X$  is sofic. This result strongly mirrors the Morse-Hedlund theorem: while the Morse-Hedlund theorem concerns the complexity sequence, and the theorem of Ormes and Pavlov concerns the

extender set sequence, each result states that if the sequence ever falls below  $\{n + 1\}$ , then the sequence must be bounded. (Interestingly, Sturmian shifts once again prove tightness, having  $n + 1$  extender sets of words of length  $n$  for every  $n$ , and failing to be sofic). This suggests a connection between the complexity sequence and extender set sequence. Section 3.1 of this thesis further explores this connection, and whether this connection may exist with follower or predecessor set sequences as well. While a similar result for follower set sequences is not proved for general one-dimensional shift spaces, we establish several results supporting the conjecture that  $\{|F_X(n)|\} \leq n$  for any  $n$  implies soficity of  $X$ :

**Theorem 1.0.1.** *[3] For any shift space  $X$ , if there exists  $n \in \mathbb{N}$  such that  $|F_X(n)| \leq \log_2(n + 1)$ , then  $X$  is sofic.*

**Theorem 1.0.2.** *[3] Let  $X$  be a shift space. If  $|F_X(n)| \leq n$  for any  $n \leq 3$ , then  $X$  is sofic.*

The results in Section 3.1 are the outcome of joint work with Ormes and Pavlov in [3]. In Section 3.2, we pivot from focusing on the connections among complexity sequences and follower, predecessor, and extender set sequences, and instead we examine the differences among them. In particular, we demonstrate that while complexity sequences must be monotone increasing, this need not be true of follower, predecessor, or extender set sequences. Furthermore, we show that the follower, predecessor, and extender set sequences of sofic shifts must be eventually periodic.

**Theorem 1.0.3.** *[4] Let  $X$  be a one-dimensional sofic shift,  $p$  be one greater than the total number of extender sets in  $X$ , and  $p_0$  be one greater than the total*

number of follower sets in  $X$ . Then the extender set sequence  $\{|E_X(n)|\}_{n \in \mathbb{N}}$  is eventually periodic, where the periodicity must begin before the  $p(1+p)!$ <sup>th</sup> term, and the least eventual period is at most  $p!$ . The follower set sequence  $\{|F_X(n)|\}_{n \in \mathbb{N}}$  is eventually periodic, where the periodicity must begin before the  $p_0(1+p_0!)!$ <sup>th</sup> term, and the least eventual period is at most  $p_0!$ .

We also show that a wide class of eventually periodic sequences may be achieved, both as follower set sequences and as extender set sequences.

**Theorem 1.0.4.** [4] Let  $\ell \in \mathbb{N}$ , and  $\mathcal{A} = \{A_1, A_2, A_3, \dots, A_k\}$  be a nontrivial partition of  $\{0, 1, \dots, \ell - 1\}$ . Let  $0 = r_1 < r_2 < \dots < r_k$  be natural numbers. Then there exists  $m \in \mathbb{N}$  and an irreducible graph  $\mathcal{G}$  such that the number of follower sets in  $X_{\mathcal{G}}$  of words of length  $n$  where  $n \geq \ell + 2$  and  $n \pmod{\ell} \in A_j$  will be exactly  $m + r_j$  for all  $1 \leq j \leq k$ . Furthermore,  $m$  may be chosen such that  $m < (6\ell + 3)r_k$ .

**Theorem 1.0.5.** [4] Let  $\ell \in \mathbb{N}$ , and  $\mathcal{A} = \{A_1, A_2, A_3, \dots, A_k\}$  be a nontrivial partition of  $\{0, 1, \dots, \ell - 1\}$ . Let  $0 = r_1 < r_2 < \dots < r_k$  be natural numbers. Then there exists  $m \in \mathbb{N}$  and an irreducible graph  $\mathcal{G}$  such that the number of extender sets in  $X_{\mathcal{G}}$  of words of length  $n$  where  $n \geq 14r_k\ell - 1$  and  $n \pmod{\ell} \in A_j$  will be exactly  $m + r_j$  for all  $1 \leq j \leq k$ . Furthermore,  $m$  may be chosen such that  $m \leq 39\ell^2 r_k^2$ .

These results are from my own work in [4], and establish that many follower, predecessor, and extender set sequences may not be realizable as complexity sequences. In Section 3.3, we turn the question on its head, and address the question of whether all complexity sequences may be realizable as follower, predecessor, or extender set sequences. We use the classical  $\beta$ -shifts

to demonstrate that the set of all sequences which may be realized as a complexity sequences is nearly a subset of the class of all sequences which may be realized as a follower set sequence.

**Theorem 1.0.6.** *Let  $\{\Phi_d(n)\}$  be the complexity sequence of a right-infinite sequence  $d$  such that for all  $i \in \mathbb{N}$ ,  $\sigma^i(d) \preceq d$ . Then the sequence  $\{\Phi_d(n)\}$  is the predecessor set sequence of  $X_\beta$  for some  $\beta > 1$ .*

**Theorem 1.0.7.** *Let  $\{\Phi_d(n)\}$  be the complexity sequence of any right-infinite sequence  $d$ . Then the sequence  $\{\Phi_d(n) + 1\}$  is the predecessor set sequence of  $X_\beta$  for some  $\beta > 1$ .*

The  $\beta$ -shifts also provide examples of other interesting occurrences, including shifts with positive topological entropy for which the follower set sequence is  $\{n + 1\}$ , contrasting with other common zero-entropy examples with the same follower set sequence.  $\beta$ -shifts are also a rich class of examples of shifts for which the follower and predecessor set sequences display drastically different limiting behavior.



# Chapter 2

## Definitions and Examples

Let  $\mathcal{A}$  denote a finite set, which we will refer to as our alphabet. Elements of  $\mathcal{A}$  will be called letters.

**Definition 2.0.1.** A *shift space*  $X$  (or sometimes referred to as a *subshift*  $X$ ) on an alphabet  $\mathcal{A}$  is some subset of  $\mathcal{A}^{\mathbb{Z}}$  which is shift-invariant and closed in the product topology.

**Example 2.0.2.** Let  $X \subset \{0, 1\}^{\mathbb{Z}}$  such that  $X = \{\dots x_{-2}x_{-1}.x_0x_1x_2\dots \mid x_0 = 0\}$ . Then  $X$  is not a shift space, because  $X$  is not shift-invariant. In particular,  $\sigma(\dots 111.01111\dots) = \dots 1110.1111\dots \notin X$ .

**Example 2.0.3.** Let  $X \subset \{0, 1\}^{\mathbb{Z}}$  such that  $X$  is the set of all bi-infinite sequences containing the symbol 1 exactly once. Then  $X$  is not a shift space. Though  $X$  is shift-invariant, it is not closed in the product topology, because the sequence:

$$x_1 = \dots 000.1000000000000\dots$$

$$x_2 = \dots 000.0001000000000\dots$$

$$x_3 = \dots 000.00000001000000\dots$$

$$x_4 = \dots 000.00000000001000\dots$$

$$\vdots$$

approaches the limit  $x = \dots 000.0000000\dots = 0^\infty \notin X$  in the product topology.

**Example 2.0.4.** Let  $X \subset \{0, 1\}^{\mathbb{Z}}$  such that  $X$  is the set of all bi-infinite sequences containing the symbol 1 at most once. Then  $X$  is a shift space, as no second 1 symbol may be introduced by shifting or by taking limits, and therefore  $X$  is shift-invariant and closed in the product topology.

**Definition 2.0.5.** A **word**  $w$  over  $\mathcal{A}$  is a member of  $\mathcal{A}^n$  for some  $n \in \mathbb{N}$ , which we call the **length** of  $w$ . Occasionally the length of  $w$  will be denoted by  $|w|$ . We use  $\emptyset$  to denote the **empty word**, the word of length zero.

**Definition 2.0.6.** Suppose we have an order on the alphabet  $\mathcal{A}$ . Then for two words of the same length (or two right-infinite sequences)  $x$  and  $y$ , we say that  $x$  is **lexicographically less than**  $y$  if, for the first place that  $x$  and  $y$  disagree,  $x$  takes a smaller value than  $y$ . We denote this by  $x \prec y$ . In this thesis,  $\mathcal{A}$  will usually consist of non-negative integers, and the order on  $\mathcal{A}$  will be the usual one.

**Definition 2.0.7.** For any words  $v \in \mathcal{A}^n$  and  $w \in \mathcal{A}^m$ , we define the **concatenation**  $vw$  to be the word in  $\mathcal{A}^{n+m}$  whose first  $n$  letters are the letters forming  $v$  and whose next  $m$  letters are the letters forming  $w$ .

**Definition 2.0.8.** For a word  $u \in \mathcal{A}^n$ , if  $u$  can be written as the concatenation of two words  $u = vw$  then we say that  $v$  is a **prefix** of  $u$  and that  $w$  is a **suffix** of  $u$ . When there is no risk of confusion, we denote the  $n$ -letter prefix of a word (or right-infinite sequence)  $u$  by  $(u)_n$ .

**Definition 2.0.9.** The **language** of a shift space  $X$ , denoted by  $L(X)$ , is the set of all words which appear in points of  $X$ . For any finite  $n \in \mathbb{N}$ , define  $L_n(X) = L(X) \cap \mathcal{A}^n$ , the set of words in the language of  $X$  of length  $n$ . We will sometimes informally refer to words in the language as being **legal**.

**Definition 2.0.10.** The **complexity function** of a shift space  $X$ ,  $\Phi_X(n) = |L_n(X)|$ , sends a length  $n$  to the number of words of that length in  $L(X)$ . The **complexity sequence** of a shift space  $X$  is  $\{\Phi_X(n)\}_{n \in \mathbb{N}}$ .

**Example 2.0.11.** The **full 2-shift** is  $X_{[2]} = \{0, 1\}^{\mathbb{Z}}$ . Clearly,  $X_{[2]}$  is a shift space. For any length  $n$ , any combination of  $n$ -many 0's and 1's is legal. Thus the complexity sequence of the full 2-shift is  $\{\Phi_{X_{[2]}}(n)\} = \{2^n\}$ .

**Example 2.0.12.** Let  $X$  be the shift space consisting only of two points,  $x = \dots 0101.0101\dots = (01)^\infty$  and its shift  $\sigma(x) = \dots 1010.1010\dots = (10)^\infty$ . The reader may check that  $X$  is a shift space. For any length  $n$ , only 2  $n$ -letter words occur in  $X$ , one beginning with the symbol 1 and alternating between 0 and 1, and another beginning with the symbol 0 and alternating between 1 and 0. Thus, the complexity sequence of  $X$  is  $\{\Phi_X(n)\} = \{2, 2, 2, \dots\}$ .

**Definition 2.0.13.** The **topological entropy**  $h(X)$  of a one-dimensional shift space  $X$  is a measure of the exponential growth rate of the number of words in  $X$ , and is given by

$$h(X) = \lim_{n \rightarrow \infty} \frac{1}{n} \log(\Phi_X(n)).$$

**Example 2.0.14.** It is a simple calculation to see that the topological entropy of the full 2-shift of Example 2.0.11 is  $\log(2)$ , while the topological entropy of the shift space in Example 2.0.12 is zero.

**Definition 2.0.15.** For any shift space  $X$  on an alphabet  $\mathcal{A}$ , and any word  $w$  in the language of  $X$ , we define the **follower set of  $w$  in  $X$** ,  $F_X(w)$ , to be the set of all right-infinite sequences  $u \in \mathcal{A}^{\mathbb{N}}$  such that the infinite word  $wu$  occurs in some point of  $X$ . (Note that  $F_X(\emptyset)$  is simply the set of all right-infinite sequences appearing in any point of  $X$ .) Some texts define the follower set of  $w$  to be the set of all finite words which may follow  $w$ ; the two definitions are analogous. Similarly, we define the **predecessor set of  $w$  in  $X$** ,  $P_X(w)$ , to be the set of all left-infinite sequences  $s \in \mathcal{A}^{-\mathbb{N}}$  such that  $sw$  occurs in some point of  $X$ . As before, a definition replacing left-infinite sequences with finite words is analogous.

**Example 2.0.16.** The **golden mean shift**  $X_\varphi$  is the set of all bi-infinite sequences on the alphabet  $\mathcal{A} = \{0, 1\}$  such that the symbol 1 never appears adjacent to another 1. Thus the follower set  $F(0)$  of the word 0 is equal to  $F(\emptyset)$ , because any legal right-infinite sequence may follow a 0 without introducing the word 11. In contrast, the follower set  $F(1)$  of the word 1 is equal to  $\{.x_0x_1x_2\dots \mid x_0 = 0\} \cap F(\emptyset)$ , that is, any legal right-infinite sequence beginning with the symbol 0, because the word 1 may not be followed by any sequence beginning with a 1, as this would create the forbidden word 11.

**Definition 2.0.17.** For any word  $w \in L(X)$ ,  $w$  is **follower-shortenable** if there exists  $v \in L(X)$  with strictly shorter length than  $w$  such that  $F_X(w) = F_X(v)$ .

**Example 2.0.18.** In the golden mean shift from example 2.0.16, only the last letter of a word determines its follower set—words ending with a 0 may be followed by any legal right-infinite sequence, while words ending with a 1 may

only be followed by legal right-infinite sequences which begin with 0. Thus, any word of length at least 2 is follower-shortenable. In particular, the word 01 is follower-shortenable to 1, while 00 is follower-shortenable to 0, and so on.

**Definition 2.0.19.** For any shift space  $X$  over the alphabet  $A$ , and any word  $w$  in the language of  $X$ , we define the **extender set of  $w$  in  $X$** ,  $E_X(w)$ , to be the set of all pairs  $(s, u)$  where  $s$  is a left-infinite sequence of symbols in  $\mathcal{A}$ ,  $u$  is a right-infinite sequence of symbols in  $\mathcal{A}$ , and  $swu$  is a point of  $X$ . Once again, a definition replacing infinite sequences with finite words is analogous.

**Remark 2.0.20.** For any word  $w \in L(X)$ , define a projection function  $f_w : E_X(w) \rightarrow F_X(w)$  by  $f_w(s, u) = u$ . Such a function sends the extender set of  $w$  onto the follower set of  $w$ . Any two words  $w, v$  with the same extender set would have the property then that  $f_w(E_X(w)) = f_v(E_X(v))$ , that is, that  $w$  and  $v$  have the same follower set. Similarly, any two words with the same extender set must also have the same predecessor set.

**Remark 2.0.21.** We may informally think of the extender set of a word  $w$  as the set of all possible ways to complete  $w$  into an entire bi-infinite point of  $X$ . Using this perspective, the definition of extender set may be generalized to multi-dimensional symbolic settings, unlike the definitions of follower and predecessor sets.

**Definition 2.0.22.** For any positive integer  $n$ , define the set  $F_X(n) = \{F_X(w) \mid w \in L_n(X)\}$ . Thus the cardinality  $|F_X(n)|$  is the number of distinct follower sets of words of length  $n$  in  $X$ . Similarly, define  $P_X(n) = \{P_X(w) \mid w \in L_n(X)\}$  and  $E_X(n) = \{E_X(w) \mid w \in L_n(X)\}$ , so that  $|P_X(n)|$

and  $|E_X(n)|$  are the numbers of distinct predecessor sets of words of length  $n$  in  $X$  and extender sets of words of length  $n$  in  $X$  respectively.

**Definition 2.0.23.** Given a shift space  $X$ , let the **follower set sequence** of  $X$  be  $\{|F_X(n)|\}_{n \in \mathbb{N}}$ , the sequence that records, for each  $n$ , the number of distinct follower sets in  $X$  corresponding to words in  $L_n(X)$ . Similarly, define the **predecessor set sequence** of  $X$  to be  $\{|P_X(n)|\}_{n \in \mathbb{N}}$  and the **extender set sequence** of  $X$  to be  $\{|E_X(n)|\}_{n \in \mathbb{N}}$ .

**Example 2.0.24.** In the full 2-shift as defined in Example 2.0.11, any right-infinite sequence on  $\{0, 1\}$  may legally follow any word in  $L(X_{[2]})$ , and so for any word  $w$ ,  $F(w) = F(\emptyset)$ . Thus every word in  $L(X_{[2]})$  has the same follower set, and so there is only one follower set in  $X_{[2]}$ . Similarly, there is only one predecessor and one extender set in  $X_{[2]}$ . Therefore  $\{|F_{X_{[2]}}(n)|\} = \{|P_{X_{[2]}}(n)|\} = \{|E_{X_{[2]}}(n)|\} = \{1, 1, 1, 1, \dots\}$ .

**Example 2.0.25.** In the golden mean shift from Example 2.0.16, the follower set of a word  $w$  is determined only by the last letter of  $w$ , so there are only two follower sets in  $X_\varphi$ . The follower set sequence of the golden mean shift is  $\{|F_{X_\varphi}(n)|\} = \{2, 2, 2, 2, \dots\}$ . Similarly, the predecessor set of a word  $w$  is determined only by the first letter of  $w$ , so there are only two predecessor sets in  $X_\varphi$ . The predecessor set sequence of the golden mean shift is  $\{|P_{X_\varphi}(n)|\} = \{2, 2, 2, 2, \dots\}$ . Finally, the extender set of a word  $w$  is determined by the first and last letters of the word  $w$ . A word beginning and ending with 0 will have a different extender set from a word beginning with 0 and ending with 1, for instance, because those two words will have different follower sets. Furthermore, a word beginning and ending with 0 will have a different

extender set from a word beginning with 1 and ending with 0, because those two words will have different predecessor sets, and so on. Thus there are four distinct extender sets in  $X_\varphi$ . However, not all four extender sets may be realized for every length. For length 1, there are only two words, and thus only two extender sets may be realized. For length 2, the extender set corresponding to words which both begin and end with 1 may not be realized, because 11 is not a legal word in  $X_\varphi$ . The extender set sequence of the golden mean shift is  $\{|E_{X_\varphi}(n)|\} = \{2, 3, 4, 4, 4, \dots\}$ .

**Definition 2.0.26.** A shift space  $X$  is a **shift of finite type** if it may be described by a finite list of forbidden words. That is, there exists  $m \in \mathbb{N}$  and a subset  $\mathcal{F} \subseteq \mathcal{A}^m$  such that  $X = \{x \in \mathcal{A}^{\mathbb{Z}} \mid x \text{ does not contain any word in } \mathcal{F} \text{ as a subword}\}$ . If  $X$  can be described by a finite list of forbidden words  $\mathcal{F} \subseteq \mathcal{A}^2$ , then  $X$  is a **nearest-neighbor shift of finite type**. In this case, the only restrictions concern which letters may and may not sit adjacent to one another.

**Remark 2.0.27.** Every shift space may be described by some collection  $\mathcal{F}$  of forbidden words, in fact, a space  $X \subseteq \mathcal{A}^{\mathbb{Z}}$  is closed in the product topology if and only if it can be described by a countable (not necessarily finite) collection  $\mathcal{F}$  of forbidden words.

**Example 2.0.28.** The golden mean shift of Example 2.0.16 is a nearest-neighbor shift of finite type, because  $X_\varphi$  is defined by the list of forbidden words  $\mathcal{F} = \{11\} \subseteq \{0, 1\}^2$ .

**Definition 2.0.29.** A shift space  $X$  is **sofic** if it is the image of a shift of finite type under a continuous shift-commuting map.

Equivalently, sofic shifts are those with only finitely many follower sets, that is, a shift  $X$  is sofic iff  $\{F_X(w) \mid w \in L(X)\}$  is finite ([7]). The same equivalence exists for extender sets:  $X$  is sofic iff  $\{E_X(w) \mid w \in L(X)\}$  is finite ([9]). This necessarily implies that for a sofic shift  $X$ , the follower set sequence and extender set sequence of  $X$  are bounded. In fact, the converse is also true: if the follower set or extender set sequence of a shift  $X$  is bounded, then  $X$  is necessarily sofic. (See [9]). Another well-known equivalence is that a shift  $X$  is sofic iff there exists a finite directed labeled graph  $\mathcal{G}$  such that  $X$  is exactly the set of sequences of labels of all bi-infinite walks on  $\mathcal{G}$  ([7]). In such a case, we say that  $X$  is the **edge shift presented by  $\mathcal{G}$** , and may denote  $X$  by  $X_{\mathcal{G}}$ .

**Example 2.0.30.** *Let  $X$  be the set of all bi-infinite sequences on alphabet  $\{0, 1\}$  such that whenever a run of 0's appears between two 1's, that run of 0's has even length. Then  $X$  is called the **even shift**. Then  $X$  is not a shift of finite type, as describing  $X$  by a list of forbidden words requires an infinite list, for example  $\mathcal{F} = \{101, 10001, 1000001, \dots\}$ . Yet  $X$  is sofic, as  $X$  is exactly the set of sequences of labels of all bi-infinite walks on the graph  $\mathcal{G}$  given in Figure 2.1.*

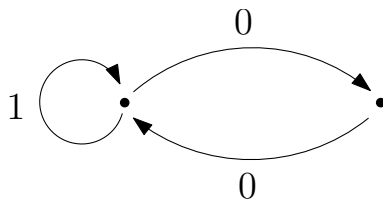


Figure 2.1: The graph  $\mathcal{G}$  presenting the even shift

*Additionally, we may determine that  $X$  is sofic because  $X$  is the image of the golden mean shift of example 2.0.16 under a continuous shift-commuting map*



$\psi : X_\varphi \rightarrow X$ , where  $\psi(\dots x_{-2}x_{-1}.x_0x_1x_2\dots) = \dots y_{-2}y_{-1}.y_0y_1y_2\dots$  where

$$y_n = \begin{cases} 0 & \text{if } x_nx_{n+1} = 10 \text{ or } 01 \\ 1 & \text{if } x_nx_{n+1} = 00. \end{cases}$$

**Definition 2.0.31.** A directed labeled graph  $\mathcal{G}$  is **irreducible** if for every ordered pair  $(I, J)$  of vertices in  $\mathcal{G}$ , there exists a path in  $\mathcal{G}$  from  $I$  to  $J$ . A shift space  $X$  is **irreducible** if for any two words  $w$  and  $v$  in  $L(X)$ , there exists a word  $u$  such that  $wuv \in L(X)$ . Note that a sofic shift is irreducible if and only if it can be presented by a graph  $\mathcal{G}$  which is irreducible.

Results about shifts presented by graphs which are not irreducible may often be found by considering the reducible graph's irreducible components; for this reason, results in this thesis will largely focus on the irreducible case.

**Definition 2.0.32.** A directed labeled graph  $\mathcal{G}$  is **primitive** if  $\exists N \in \mathbb{N}$  such that for every  $n \geq N$ , for every ordered pair  $(I, J)$  of vertices in  $\mathcal{G}$ , there exists a path in  $\mathcal{G}$  from  $I$  to  $J$  of length  $n$ . The least such  $N$  is the **primitivity distance** for  $\mathcal{G}$ .

**Example 2.0.33.** The reader may check that the graph given in Figure 2.1 is primitive with primitivity distance 2 (there is no path of length 1 from the right vertex to itself). On the other hand, the graph given in Figure 2.2 is not primitive, because all paths from vertex  $I$  to vertex  $K$  must have even length.

**Remark 2.0.34.** Any irreducible finite graph containing a self-loop is necessarily primitive. Let  $\mathcal{G}$  be an irreducible graph with a self-loop, and let  $K$  be the vertex at which the self-loop is anchored. Let  $N \in \mathbb{N}$  be such that for any

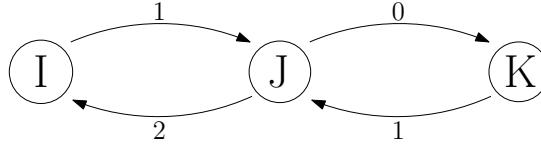


Figure 2.2: A graph which is right-resolving, but not left-resolving, primitive, or follower-separated

vertices  $I, J$ , there exists a path from  $I$  to  $J$  of length less than or equal to  $N$ . (At most, we can take  $N$  to be the total number of edges in the graph). Then for any pair of vertices  $I$  and  $J$ , and any length  $n > 2N$ , a path from  $I$  to  $J$  of length  $n$  may be created by traveling from  $I$  to  $K$ , then following the self-loop an appropriate number of times to inflate the length of the path, then traveling from  $K$  to  $J$ .

**Definition 2.0.35.** A directed labeled graph  $\mathcal{G}$  is **right-resolving** if for each vertex  $I$  of  $\mathcal{G}$ , all edges starting at  $I$  carry different labels. Similarly,  $\mathcal{G}$  is **left-resolving** if for each vertex  $I$  of  $\mathcal{G}$ , all edges ending at  $I$  carry different labels.

**Example 2.0.36.** The graph in Figure 2.2 is right-resolving, but fails to be left-resolving as the two edges ending at the vertex  $J$  are both labeled 1.

**Definition 2.0.37.** A directed labeled graph  $\mathcal{G}$  is **follower-separated** if distinct vertices in  $\mathcal{G}$  correspond to distinct follower sets. That is, for all vertices  $I, J$  in  $\mathcal{G}$ , there exists a one-sided infinite sequence  $s$  of labels which may follow one vertex but not the other.

**Example 2.0.38.** The graph in Figure 2.1 is follower-separated. Words ending at the left vertex may be followed by sequences beginning with 1, while words ending at the right vertex may not, and thus two vertices correspond to dis-

tinct follower sets. However, the graph in Figure 2.2 is not follower-separated; the reader may check that the right-infinite sequences which may follow words ending at vertex  $I$  are exactly the right-infinite sequences which may follow words ending at  $K$ .

**Definition 2.0.39.** A directed labeled graph  $\mathcal{G}$  is **extender-separated** if distinct pairs of vertices correspond to distinct extender sets. That is, for any two distinct pairs of initial and terminal vertices  $\{I \rightarrow I'\}$  and  $\{J \rightarrow J'\}$  such that there exist paths in  $\mathcal{G}$  from  $I$  to  $I'$  and from  $J$  to  $J'$ , there exists some word  $w$  which is the label of a path in  $\mathcal{G}$  beginning and ending with one pair of vertices, and pair  $(s, u)$ ,  $s$  a left-infinite sequence,  $u$  a right-infinite sequence, such that  $swu$  is a point of  $X_{\mathcal{G}}$ , but for every word  $v$  which is the label of some path beginning and ending with the other pair of vertices,  $svu$  is not a point of  $X_{\mathcal{G}}$ .

Informally, a graph  $\mathcal{G}$  is extender-separated if two distinct pairs  $\{I \rightarrow I'\}$  and  $\{J \rightarrow J'\}$  of initial and terminal vertices correspond to distinct extender sets whenever they correspond to non-empty extender sets. In a graph which is not irreducible, many pairs of initial and terminal vertices may correspond to an empty extender set, because there is no path between the initial and terminal vertex in those pairs. This does not prevent the graph from being extender-separated.

**Definition 2.0.40.** Given a directed labeled graph  $\mathcal{G}$ , a word  $w$  is **right-synchronizing** if all paths in  $\mathcal{G}$  labeled  $w$  terminate at the same vertex. The word  $w$  is **left-synchronizing** if all paths in  $\mathcal{G}$  labeled  $w$  begin at the same vertex. The word  $w$  is **bi-synchronizing** if  $w$  is both left- and right-

synchronizing. A **bi-synchronizing letter** is a bi-synchronizing word of length 1.

If a word  $w$  is right-synchronizing, left-synchronizing, or bi-synchronizing in  $\mathcal{G}$ , then  $w$  has the property that whenever  $u$  and  $v$  are words in  $L(X_{\mathcal{G}})$  such that  $uw$  and  $wv$  are in  $L(X_{\mathcal{G}})$ , then  $uwv$  is in  $L(X_{\mathcal{G}})$  as well.

**Example 2.0.41.** Consider the graph in Figure 2.1. In this graph, the word 1 is bi-synchronizing because the label 1 is applied to only one edge in the graph. In contrast, the word 0 is neither left- nor right-synchronizing, because edges labeled 0 begin at both vertices, and edges labeled 0 terminate at both vertices. Conversely, consider the graph in Figure 2.2. Every word on this graph is right-synchronizing, but the word 1 fails to be left-synchronizing because edges labeled 1 begin at both  $I$  and  $K$ .

In fact, every one-dimensional sofic shift is presented by some graph  $\mathcal{G}$  which is right-resolving, follower-separated, and contains a right-synchronizing word (see [7]).

**Definition 2.0.42.** If  $w$  is a word with the property that whenever  $u$  and  $v$  are words in  $L(X)$  such that  $uw$  and  $wv$  are in  $L(X)$ , then  $uwv$  is in  $L(X)$  as well, then  $w$  is said to be **intrinsically synchronizing**. Unlike the terms in Definition 2.0.40, the definition of intrinsically synchronizing makes no reference to a labeled graph, and so even words in non-sofic shifts may be intrinsically synchronizing.

Finally, for some parts of this thesis we will need to consider one-sided shift spaces:

**Definition 2.0.43.** A *one-sided shift space* is a subset  $X \subseteq \mathcal{A}^{\mathbb{N}}$ , paired with the shift map  $\sigma$ , which we now think of as removing the  $0^{\text{th}}$  digit, rather than shifting it left of the origin, so  $\sigma(.x_0x_1x_2x_3\dots) = .x_1x_2x_3\dots$

It is fairly easy to move back and forth between one-sided and two-sided shift spaces. A one-sided shift may be constructed from a two-sided shift by simply ignoring any digits left of the origin. A two-sided shift may be constructed from a one-sided shift in the following way:

**Definition 2.0.44.** *Two-sided shifts may be constructed from one-sided shifts using the natural extension:* Given a one-sided shift  $X$ , create a two-sided shift  $\widehat{X}$  by asserting that a bi-infinite sequence  $x$  is in  $\widehat{X}$  if and only if every finite subword of  $x$  is in  $L(X)$ .

Note that this construction implies  $L(\widehat{X}) \subseteq L(X)$ . Strict equality may not be possible, for instance, suppose there exists a word  $w$  and  $N \in \mathbb{N}$  such that the word  $w$  never appears after the  $N^{\text{th}}$  digit of any sequence in  $X$ . Then clearly  $w$  cannot be in  $L(\widehat{X})$ , as no finite word longer than  $N + |w|$  could end with  $w$  in  $L(X)$ , and so no bi-infinite sequence  $x$  containing  $w$  could have the property that every finite subword of  $x$  is in  $L(X)$ .

# Chapter 3

## Follower and Extender Sets

### 3.1 Follower Sets and Complexity

We begin by exploring the connections between follower, predecessor, and extender set sequences and complexity sequences of a one-dimensional shift space  $X$ . An immediate relation is that for a fixed one-dimensional shift  $X$ , for any  $n \in \mathbb{N}$ ,  $|E_X(n)| \leq \Phi_X(n)$ , with equality only if each  $n$ -letter word has a distinct extender set.  $\Phi_X(n)$  is also an upper bound for  $|F_X(n)|$  and  $|P_X(n)|$ . However, the follower, predecessor, and extender set sequences may be substantially smaller than the complexity sequence if many words have the same follower, predecessor, or extender set. For instance, recall examples 2.0.11 and 2.0.24, which show that the full 2-shift  $X_{[2]}$  has a complexity sequence which grows exponentially in  $n$ , yet a constant follower, predecessor, and extender set sequence.

The following result of Ormes and Pavlov regarding extender sets closely mirrors the Morse-Hedlund theorem for complexity, and suggests a relation-

ship between the class of achievable complexity sequences and the class of achievable extender set sequences of one-dimensional shift spaces:

**Theorem 3.1.1.** [9] *For a shift space  $X$ , if there exists an  $n \in \mathbb{N}$  such that  $|E_X(n)| \leq n$ , then  $X$  is sofic.*

This theorem tells us that extender set sequences, like complexity sequences, are bounded if and only if they fall as low as  $n$  for some  $n$ . It is natural to wonder, then, whether follower set sequences follow the same rule, or whether unbounded follower set sequences may grow to infinity slower than linearly. Ormes and Pavlov have conjectured that follower set sequences must behave like complexity and extender set sequences:

**Conjecture 1.** [Ormes-Pavlov] *For a shift space  $X$ , if there exists an  $n \in \mathbb{N}$  such that  $|F_X(n)| \leq n$ , then  $X$  is sofic.*

**Remark 3.1.2.** *To prove this conjecture for follower set sequences would be equivalent to proving it for predecessor set sequences as well. If  $X$  is a shift space, define  $-X$  to be the shift space made from “flipping” points of  $X$ . That is, if  $\dots x_{-2}x_{-1}.x_0x_1x_2\dots \in X$ , then  $\dots x_2x_1.x_0x_{-1}x_{-2}\dots \in -X$ . Then  $-X$  is a shift space and  $\{|F_X(n)|\} = \{|P_{-X}(n)|\}$  and  $\{|P_X(n)|\} = \{|F_{-X}(n)|\}$ . For this reason, we will ignore predecessor set sequences and focus on follower set sequences for the majority of this thesis.*

In this section we will present results toward the conjecture of Ormes and Pavlov. We begin with some simple facts about follower sets, which will repeatedly be useful. The proofs are simple and left to the reader.

**Lemma 3.1.3.** For any shift space  $X$ , any  $w \in L_n(X)$ , and any  $m \in \mathbb{N}$ ,  $F_X(w) = \bigcup_v F_X(vw)$ , where the union is taken over those  $v \in L_m(X)$  for which  $vw \in L_{m+n}(X)$ .

**Lemma 3.1.4.** For any shift space  $X$ , and  $w \in L_n(X)$  and any  $m < n$ , there exists a  $v \in L_m(X)$  for which  $F(v) \supseteq F(w)$ .

**Lemma 3.1.5.** Let  $X$  be a shift space. If for two words  $w, u \in L(X)$ ,  $F_X(w) = F_X(u)$ , then for any  $v \in L(X)$ ,  $F_X(wv) = F_X(uv)$ .

The following will be our main tool for proving soficity of a shift space via the sets  $F_X(n)$ .

**Theorem 3.1.6.** For any shift space  $X$ , if there exists  $n \in \mathbb{N}$  such that  $F_X(n) \subseteq \bigcup_{\ell \leq n-1} F_X(\ell)$ , then  $X$  is sofic.

*Proof.* If there exists  $n \in \mathbb{N}$  such that  $F_X(n) \subseteq \bigcup_{\ell \leq n-1} F_X(\ell)$ , then for any word  $w \in L_n(X)$ , the follower set  $F_X(w)$  is also the follower set of a strictly shorter word, so  $w$  is follower-shortenable to a word of length strictly less than  $n$ . Now, let  $v \in L(X)$  of length greater than  $n$ , say  $v = v_1v_2\dots v_nv_{n+1}\dots v_k$  where  $k > n$ . Then  $v_1v_2\dots v_n \in L_n(X)$ , and so is follower-shortenable to some word  $v' \in L(X)$  of length less than  $n$ . But  $F_X(v_1v_2\dots v_n) = F_X(v')$  implies  $F_X(v_1v_2\dots v_nv_{n+1}\dots v_k) = F_X(v'v_{n+1}\dots v_k)$  by Lemma 3.1.5, so  $v$  is follower-shortenable to a word  $v'v_{n+1}\dots v_k$ . If  $v'v_{n+1}\dots v_k$  has length at least  $n$ , we may apply the above process again and shorten repeatedly, getting shorter and shorter words with the same follower set until we find one with length less than  $n$ . So  $v$  is follower-shortenable to a word of length less than  $n$ . But this means that  $\bigcup_{\ell \leq n-1} F_X(\ell)$  contains all follower sets in  $X$ , so  $X$  has only finitely many follower sets, and thus,  $X$  is sofic.  $\square$



We can now show that  $|F_X(n)| = 1$  for any  $n$  always implies soficity of  $X$ .

**Theorem 3.1.7.** *For any shift space  $X$ , if there exists  $n$  for which  $|F_X(n)| = 1$ , then  $X$  is a full shift.*

*Proof.* We prove the contrapositive. Without loss of generality, assume that the alphabet  $\mathcal{A}$  of  $X$  consists entirely of letters which actually appear in points of  $X$ , and assume that  $X$  is not the full shift on  $\mathcal{A}$ . Then there exists a word  $w = w_1w_2\dots w_k \in \mathcal{A}^k$  which is not in the language of  $X$ ; suppose that the length  $k$  of  $w$  is minimal. It must be the case that  $k$  is at least 2, since we assumed that all letters of  $\mathcal{A}$  are in  $L(X)$ . Then since we assumed  $k$  to be minimal,  $w_2\dots w_k \in L(X)$ , so we can choose some one-sided infinite sequence  $s$  appearing in  $X$  which begins with  $w_2\dots w_k$ . Similarly,  $w_1$  is in  $L(X)$ , so for any  $n \in \mathbb{N}$ , we may choose an  $n$ -letter word  $v$  ending with  $w_1$ . Then  $vs$  contains  $w \notin L(X)$ , so  $s \notin F_X(v)$ . However, since  $s$  appears in  $X$ , there exists some  $n$ -letter word  $u$  which can be followed by  $s$  in  $X$ , and so  $s \in F_X(u)$ . Hence  $F_X(u) \neq F_X(v)$ , so  $|F_X(n)| \geq 2$ , and since  $n$  was arbitrary, this is true for all  $n$ .  $\square$

We can now prove a version of Conjecture 1 for unions of the sets  $F_X(n)$ , rather than the sets themselves.

**Theorem 3.1.8.** *For any shift space  $X$ , if there exists  $n \in \mathbb{N}$  so that  $\left| \bigcup_{\ell \leq n} F_X(\ell) \right| \leq n$ , then  $X$  is sofic.*

*Proof.* We prove the contrapositive, and so assume that  $X$  is nonsocic. By Theorem 3.1.7,  $|F_X(1)| \geq 2$ . Then, by Theorem 3.1.6, for every  $n > 1$ , we have  $F_X(n) \setminus \bigcup_{\ell < n} F_X(\ell) \neq \emptyset$ , and so  $\left| \bigcup_{\ell \leq n} F_X(\ell) \right| > \left| \bigcup_{\ell \leq n-1} F_X(\ell) \right|$ . Therefore, by induction, for each  $n$ ,  $\left| \bigcup_{\ell \leq n} F_X(\ell) \right| \geq n + 1$ .  $\square$

We may now prove Theorem 1.0.1, which establishes a logarithmic lower bound for the growth rate of  $|F_X(n)|$  for nonsofic shifts.

*Proof of Theorem 1.0.1.* Suppose that for some  $n$ ,  $F_X(n) = \{F_1, F_2, \dots, F_k\}$  where  $k \leq \log_2(n+1)$ . By Lemma 3.1.3, for each length  $\ell < n$ , every follower set of a word in  $L_\ell(X)$  is a union of follower sets of words of length  $n$ . Therefore, every element of  $\bigcup_{\ell \leq n} F_X(\ell)$  is a non-empty union of elements of  $F_X(n)$ . There are at most  $2^k - 1 \leq 2^{\log_2(n+1)} - 1 = n$  such unions, so  $|\bigcup_{\ell \leq n} F_X(\ell)| \leq n$ , which implies that  $X$  is sofic by Theorem 3.1.8.  $\square$

Our next result shows that under the additional assumption that some non-empty word  $w$  has the same follower set as the empty word, Conjecture 1 is true.

**Lemma 3.1.9.** *For any shift space  $X$ , if there exists a non-empty word  $w \in L(X)$  such that  $F_X(w) = F_X(\emptyset)$  and  $n \in \mathbb{N}$  such that  $|F_X(n)| \leq n$ , then  $X$  is sofic.*

*Proof.* The follower set of the empty word is the set of all right-infinite sequences appearing in any point of  $X$ . If there exists a word  $w$  such that any legal right-infinite sequence may appear after  $w$ , then by Lemma 3.1.4, there is a letter with this property as well. So we may assume that  $F_X(a) = F_X(\emptyset)$  where  $a$  is a single letter.

The fact that  $F_X(\emptyset) = F_X(a)$  implies by Lemma 3.1.5 that for every  $w \in L(X)$ ,  $F_X(w) = F_X(aw) = F_X(aaw) = \dots$ . Therefore, every follower set of a word of length  $n$  is also a follower set of a word of any length greater than  $n$ . In other words,  $F_X(1) \subseteq F_X(2) \subseteq F_X(3) \subseteq \dots$ . Then, for every  $n$ ,

$F_X(n) = \bigcup_{\ell \leq n} F_X(\ell)$ , and so if  $|F_X(n)| \leq n$  for some  $n$ , clearly  $|\bigcup_{\ell \leq n} F_X(\ell)| \leq n$ , implying that  $X$  is sofic by Theorem 3.1.8.  $\square$

**Theorem 3.1.10.** *For any shift space  $X$ , if there exists  $n \geq 2$  for which  $|F_X(n)| \leq 2$ , then  $X$  is sofic.*

*Proof.* The case where  $|F_X(n)| = 1$  is treated by Theorem 3.1.7, so we choose any  $n \geq 2$  and suppose that there are exactly 2 follower sets in  $X$  of words of length  $n$ , say  $F_1$  and  $F_2$ . We consider the sets in  $F_X(1)$ . By Lemma 3.1.3, every element of  $F_X(1)$  is either  $F_1$ ,  $F_2$ , or  $F_1 \cup F_2$ . If  $|F_X(1)| = 1$ ,  $X$  is sofic by Theorem 3.1.7, so assume that  $|F_X(1)| \geq 2$ , that is, at least two of the above sets must appear in  $F_X(1)$ . Note that  $F(\emptyset) = \bigcup_{w \in L_X(n)} F(w) = F_1 \cup F_2$ , so by Lemma 3.1.9, if  $F_1 \cup F_2$  is an element of  $F_X(1)$ , then  $X$  is sofic. The only remaining case is that  $F_X(1) = \{F_1, F_2\} = F_X(n)$ , and then  $X$  is sofic by Theorem 3.1.6.  $\square$

We are now prepared to prove Conjecture 1 for  $n \leq 3$ , as in Theorem 1.0.2. Our proof is much more complicated than the cases where  $n = 1, 2$ .

*Proof of Theorem 1.0.2.* Clearly, for  $n < 3$ , Theorems 3.1.7 and 3.1.10 imply this result. We can then restrict to the case where  $n = 3$ . If  $|F_X(3)| < 3$ , then  $X$  is again sofic by either Theorem 3.1.7 or Theorem 3.1.10. We therefore suppose that  $|F_X(3)| = 3$ , say  $F_X(3) = \{F_1, F_2, F_3\}$ . We also note that  $F(\emptyset) = F_1 \cup F_2 \cup F_3$ , and if any of  $F_X(1)$ ,  $F_X(2)$ , or  $F_X(3)$  contains  $F_1 \cup F_2 \cup F_3$  as an element, then  $X$  is sofic by Lemma 3.1.9. Therefore, in everything that follows, we assume that  $F_1 \cup F_2 \cup F_3$  is not contained in  $F_X(i)$  for  $i \leq 3$ .

We first show that if any  $F_i$  is contained entirely within another, then  $X$  is sofic. Suppose for a contradiction that some  $F_i$  is contained in another, and so without loss of generality, we say that  $F_2 \subseteq F_1$ . By Lemma 3.1.3, all elements of  $F_X(2)$  are nonempty unions of  $F_1, F_2$ , and  $F_3$ . However,  $F_1 \cup F_3 = F_1 \cup F_2 \cup F_3$ , and so  $F_1 \cup F_3 = F_1 \cup F_2 \cup F_3$  are not in  $F_X(2)$  as assumed above. Also,  $F_1 \cup F_2 = F_1$ . Therefore, the only possible elements of  $F_X(2)$  are  $F_1, F_2, F_3$ , and  $F_2 \cup F_3$ . If fewer than three of these four sets are part of  $F_X(2)$ , then  $X$  is sofic by Theorem 3.1.10. Thus we may assume at least three of the four sets appear. If  $F_1, F_2$ , and  $F_3$  are all in  $F_X(2)$ , then  $F_X(3) \subseteq F_X(2)$ , implying that  $X$  is sofic by Theorem 3.1.6. Therefore,  $F_2 \cup F_3 \in F_X(2)$ . We note that by Lemma 3.1.4, some element of  $F_X(2)$  must contain  $F_1$ . If  $F_3$  contained  $F_1$ , then  $F_3 = F_1 \cup F_2 \cup F_3$  is in  $F_X(3)$ , which we assumed not to be the case above. Similarly,  $F_2 \cup F_3$  cannot contain  $F_1$ . Therefore,  $F_1$  is the only set of  $F_1, F_2, F_3$ , and  $F_2 \cup F_3$  to contain  $F_1$ , and so  $F_1 \in F_X(2)$ . Therefore  $F_X(2)$  consists of  $F_1, F_2 \cup F_3$ , and exactly one of  $F_2$  and  $F_3$ . We note that if  $F_2 \cup F_3$  is equal to any of  $F_1, F_2$ , or  $F_3$ , then either  $|F_X(2)| = 2$  or  $F_X(3) \subseteq F_X(2)$ , in either case implying soficity by either Theorem 3.1.10 or Theorem 3.1.6. So from now on we assume  $F_2 \cup F_3$  is not equal to  $F_1, F_2$ , or  $F_3$ .

Now, let us consider  $F_X(1)$ . By Lemma 3.1.3,  $F_X(1)$  can only consist of unions of sets in  $F_X(2)$ . The set  $F_X(1)$  cannot contain  $F_1 \cup F_3 = F_1 \cup F_2 \cup F_3$ , and since  $F_1 \cup F_2 = F_1$  we see that  $F_X(1) \subseteq F_X(2)$ . There exists some word  $ab \in L_2(X)$  such that  $F_X(ab) = F_2 \cup F_3$ . Clearly  $F_X(a)$  is an element of  $F_X(1)$  and therefore  $F_X(a) = F_X(xy)$  for some  $xy \in L_2(X)$ . But then by Lemma 3.1.5,  $F_X(xyb) = F_X(ab) = F_2 \cup F_3$ , a contradiction since we above noted that  $F_2 \cup F_3$  does not equal any of  $F_1, F_2$ , or  $F_3$ . We have then shown that if any

of the follower sets  $F_1, F_2$ , and  $F_3$  are contained in one another,  $X$  is sofic, and so for the rest of the proof assume that no such containments exist. Note that this also implies that if any of  $F_1 \cup F_2$ ,  $F_1 \cup F_3$ , or  $F_2 \cup F_3$  contain each other, then the containing set is  $F_1 \cup F_2 \cup F_3$ , which we have assumed is not in  $F_X(1)$ ,  $F_X(2)$ , or  $F_X(3)$ .

We break the remainder of the proof into cases by how many of the sets  $F_1, F_2$ , and  $F_3$  are elements of  $F_X(2)$ . If all three of the sets are elements of  $F_X(2)$ , then  $X$  is sofic by Theorem 3.1.6. We then have three remaining cases.

*Case 1: none of  $F_1, F_2, F_3$  are in  $F_X(2)$ .* By Lemma 3.1.3,  $F_X(2)$  consists of nonempty unions of  $F_1, F_2$ , and  $F_3$ , and we have assumed that  $F_1 \cup F_2 \cup F_3$  is not in  $F_X(2)$ . If  $|F_X(2)| \leq 2$ , then  $X$  is sofic by Theorem 3.1.10. The only possibility is then that  $F_X(2) = \{F_1 \cup F_2, F_1 \cup F_3, F_2 \cup F_3\}$ . Then by Lemma 3.1.4,  $F_X(1)$  must contain supersets of each of these sets, and it cannot contain  $F_1 \cup F_2 \cup F_3$ . This forces  $F_X(1)$  to also be  $\{F_1 \cup F_2, F_1 \cup F_3, F_2 \cup F_3\}$ , meaning that  $F_X(2) = F_X(1)$ , and so  $X$  is sofic by Theorem 3.1.6.

*Case 2: exactly one of  $F_1, F_2, F_3$  is in  $F_X(2)$ .* Without loss of generality, suppose that  $F_1 \in F_X(2)$  and  $F_2, F_3 \notin F_X(2)$ . At least two other sets must be elements of  $F_X(2)$  or else  $X$  is sofic by Theorem 3.1.10, and they must be unions of  $F_1, F_2$ , and  $F_3$  by Lemma 3.1.3. Therefore,  $F_X(2)$  contains at least two of the sets  $F_1 \cup F_2$ ,  $F_1 \cup F_3$ , and  $F_2 \cup F_3$ . By Lemma 3.1.4, some superset of any such union must also be present in  $F_X(1)$ . In this case the superset must be the set itself since we've assumed that  $F_1 \cup F_2 \cup F_3 \notin F_X(1)$ . If  $F_1$

is also in  $F_X(1)$ ,  $F_X(2) \subseteq F_X(1)$ , and  $X$  would be sofic by Theorem 3.1.6, so  $F_1 \notin F_X(1)$ .

Now, let  $abc$  be some word such that  $F_X(abc) = F_2$ . What, then, is the follower set of  $ab$ ? If it is any set in  $F_X(1)$ , then there would exist  $d$  so that  $F_X(ab) = F_X(d)$ , and then  $F_X(abc)$  would equal  $F_X(dc)$  by Lemma 3.1.5, meaning that  $F_2 \in F_X(2)$ , a contradiction. So the only choice for  $F_X(ab)$  is  $F_1$ . Since at least two of  $F_1 \cup F_2$ ,  $F_1 \cup F_3$ , and  $F_2 \cup F_3$  are in  $F_X(2)$ ,  $F_X(2)$  contains a set of the form  $F_1 \cup F_i$ . Say that  $F_X(xy) = F_1 \cup F_i$ . Then,  $F_X(xy) \supseteq F_X(ab)$ , meaning that  $F_X(xyc) \supseteq F_X(abc) = F_2$ . Since none of the  $F_i$  contain each other, this means that  $F_X(xyc) = F_2$ . But then since  $F_1 \cup F_i$  also is a member of  $F_X(1)$ , there exists  $z$  so that  $F_X(z) = F_1 \cup F_i$ , and then by Lemma 3.1.5,  $F_X(zc) = F_2$ , a contradiction since  $F_2 \notin F_X(2)$ . Hence,  $X$  is sofic in this case as well.

*Case 3: exactly two of  $F_1, F_2, F_3$  are in  $F_X(2)$ .* Without loss of generality, suppose that  $F_1, F_2 \in F_X(2)$  and  $F_3 \notin F_X(2)$ . By Lemma 3.1.4,  $F_X(2)$  must contain some superset of  $F_3$  which is not  $F_1 \cup F_2 \cup F_3$ , so it is of the form  $F_3 \cup F_i$  for  $i = 1$  or  $2$ . As in Case 2, any of the sets  $F_1 \cup F_2$ ,  $F_1 \cup F_3$ , or  $F_2 \cup F_3$  which is an element of  $F_X(2)$  must be in  $F_X(1)$  as well. This means that if  $F_1$  and  $F_2$  are both in  $F_X(1)$ , then  $F_X(2) \subseteq F_X(1)$  and  $X$  would be sofic by Theorem 3.1.6, so we restrict to the case where at least one of these sets is not in  $F_X(1)$ .

Now, let  $abc$  be some word such that  $F_X(abc) = F_3$ . As in Case 2, the follower set of  $ab$  must be some set which occurs in  $F_X(2)$  but not  $F_X(1)$ , which must be either  $F_1$  or  $F_2$  (depending on which is not part of  $F_X(1)$ ). Without

loss of generality, we say that  $F_X(ab) = F_2$ . We now show that neither  $F_1 \cup F_2$  nor  $F_2 \cup F_3$  is in  $F_X(2)$ . Suppose for a contradiction that there is a word  $xy \in L(X)$  for which  $F_X(xy) = F_2 \cup F_i$ ,  $i = 1$  or  $3$ . Then, since  $F(xy) \supseteq F(ab) = F_2$ ,  $F(xyc) \supseteq F(abc) = F_3$ . Again, since no  $F_i$  contains another, this implies that  $F(xyc) = F_3$ . Finally, we note that  $F(y) \supseteq F(xy) = F_2 \cup F_i$ , so  $F(y) = F_2 \cup F_i$ . Therefore, by Lemma 3.1.5,  $F(yc) = F(xyc) = F_3$ , but this is a contradiction since  $F_3 \notin F_X(2)$ . We now know that neither  $F_1 \cup F_2$  nor  $F_2 \cup F_3$  is in  $F_X(2)$ . By Lemma 3.1.3, all sets in  $F_X(2)$  are nonempty unions of  $F_1$ ,  $F_2$ , and  $F_3$ , and if  $|F_X(2)| < 3$ , then  $X$  is sofic by Theorem 3.1.10. The only remaining case is then that  $F_X(2) = \{F_1, F_2, F_1 \cup F_3\}$ .

We now consider the sets in  $F_X(1)$ . Recall that  $F_2 \notin F_X(1)$  and that  $F_1 \cup F_3 \in F_X(1)$  since  $F_1 \cup F_3 \in F_X(2)$ . If  $|F_X(1)| = 1$ , then  $X$  is sofic by Theorem 3.1.7, so we can assume that  $F_X(1)$  contains at least one other set, which must be a nonempty union of the elements of  $F_X(2)$  by Lemma 3.1.3. The only possibilities are  $F_1$  and  $F_1 \cup F_3$ , since we assumed earlier that  $F_1 \cup F_2 \cup F_3 \notin F_X(2)$ . Therefore, every set in  $F_X(1)$  is a superset of  $F_1$ .

Our final step will involve considering what happens when a word with follower set  $F_1$  is extended on the right by a letter. Suppose for a contradiction that there exists a word  $w \in L(X)$  with  $F_X(w) = F_1$  and a letter  $i$  for which  $F_X(wi) = F_2$ . Then, for any letter  $j$ , since  $F_X(j) \in F_X(1)$ ,  $F_X(j) \supseteq F_X(w) = F_1$ . Therefore,  $F_X(ji) \supseteq F_X(wi) = F_2$ . However, the only superset of  $F_2$  in  $F_X(2)$  is  $F_2$  itself, and so for every  $j \in A$ ,  $F_X(ji) = F_2$ . Finally, note that, by Lemma 3.1.3,  $F_X(i) = \bigcup_j F_X(ji) = F_2$ , a contradiction since  $F_2 \notin F_X(1)$ .

Similarly, let's assume for a contradiction that there exists a word  $w \in L(X)$  with  $F_X(w) = F_1$  and a letter  $i$  for which  $F_X(wi) = F_3$ . Then, choose

a letter  $j$  with  $F_X(j) = F_1 \cup F_3$ . Then, since  $F_X(j) \supseteq F_X(w) = F_1$ ,  $F_X(ji) \supseteq F_X(wi) = F_3$ . However, the only superset of  $F_3$  in  $F_X(2)$  is  $F_1 \cup F_3$ , so  $F_X(ji) = F_1 \cup F_3$ . Then, since  $F_X(j) = F_X(ji) = F_1 \cup F_3$ , by Lemma 3.1.5,  $F_X(jii) = F_X(ji) = F_1 \cup F_3$ , a contradiction since  $F_1 \cup F_3 \notin F_X(3)$ .

This means that for every word  $w \in L(X)$  with  $F_X(w) = F_1$  and any letter  $a$  for which  $wa \in L(X)$ ,  $F(wa) = F_1$ . But then, since the follower set of every letter contains  $F_1$ , the follower set of every legal 2-letter word contains  $F_1$ , a contradiction since  $F_X(2)$  contains  $F_2$ , and we assumed that none of the  $F_i$  contains another. Every case has either led to a contradiction or to the conclusion that  $X$  is sofic, and so we've proved that  $X$  is sofic.

□

Our final result for this section is a version of Conjecture 1 for a class of coded subshifts, which we define below.

**Definition 3.1.11.** *Given a set  $\mathcal{W}$  of finite words, the **coded subshift** with **code words**  $\mathcal{W}$  is the shift space generated by taking the closure of the set of all biinfinite sequences made from concatenating infinitely many words in  $\mathcal{W}$ .*

**Theorem 3.1.12.** *Given a sofic shift  $X$ , choose a subset  $\mathcal{V} \subseteq L(X)$  with the property that for any finite word  $v \in L(X)$ , there exists some  $w \in \mathcal{V}$  such that  $v$  is a suffix of  $w$ . Create a coded subshift  $Y$  with code words  $\mathcal{W} = \{wc \mid w \in \mathcal{V}\}$  where  $c$  is a letter not appearing in the alphabet of  $X$ . Then  $Y$  satisfies Conjecture 1 (That is, if  $|F_Y(n)| \leq n$  for any  $n \in \mathbb{N}$ , then  $Y$  is sofic).*

*Proof.* We begin with two preliminary observations. First,  $X \subseteq Y$ , since any point of  $X$  is a limit of finite words in  $L(X)$ , all of which are suffixes of code



words, which are themselves in  $L(Y)$ . We also note that any word in  $L(Y)$  without a  $c$  must be a subword of a code word, and therefore in  $L(X)$ .

Second, for any word  $ucv \in L(Y)$ , where  $c$  does not occur in  $u$  or  $v$ ,  $F_Y(ucv) = F_Y(cv)$ . Clearly  $F_Y(ucv) \subseteq F_Y(cv)$ . Let  $s \in F_Y(cv)$ . Because  $uc$  is the suffix of a code word and  $vs$  is the begins with a code word in  $Y$ ,  $ucvs$  occurs in  $Y$ , so  $s \in F_Y(ucv)$ , and therefore  $F_Y(ucv) \supseteq F_Y(cv)$ .

We begin our proof by claiming that there are only finitely many follower sets in  $Y$  of words not containing the letter  $c$ . There are only finitely many follower sets in  $X$ , so it is sufficient to show that for any  $w, v \in L(X)$ ,  $F_X(w) = F_X(v)$  implies  $F_Y(w) = F_Y(v)$ . To that end, let  $F_X(w) = F_X(v)$  and consider any  $s \in F_Y(w)$ . If  $s$  does not contain the letter  $c$ , then  $ws$  is a limit of longer and longer words in  $\mathcal{W}$ , and since all such words are in  $L(X)$ ,  $ws$  occurs in  $X$ , i.e.  $s \in F_X(w)$ . Since  $F_X(w) = F_X(v)$ ,  $s \in F_X(v)$ , i.e.  $vs$  also occurs in  $X$ . Since  $Y \supseteq X$ ,  $vs$  occurs in  $Y$  as well, and so  $s \in F_Y(v)$ .

On the other hand, if  $s$  contains the letter  $c$  and  $s \in F_Y(w)$ , then  $s = s'cs''$  for some  $s'$  not containing  $c$  ( $s'$  may be the empty word). By the same logic as above,  $ws' \in L(X)$ , therefore  $vs' \in L(X)$ , and so  $vs'$  occurs as a suffix of some word in  $\mathcal{W}$ . But then,  $vs'c$  is a suffix of some code word, and so  $vs'cs''$  occurs in  $Y$ .

We have shown that in both cases,  $s \in F_Y(w)$  implies  $s \in F_Y(v)$ , and so  $F_Y(w) \subseteq F_Y(v)$ . By the same argument,  $F_Y(v) \subseteq F_Y(w)$ , giving  $F_Y(w) = F_Y(v)$ . Therefore there are only finitely many follower sets in  $Y$  of words not containing  $c$ .

Now, we assume that  $n$  is such that  $|F_Y(n)| \leq n$ . Partition  $L_n(Y)$  into  $n + 1$  sets based on the last appearance of the letter  $c$  in the word—the first set

$S_0$  consists of words with no  $c$ , the second set  $S_1$  consists of words ending with  $c$ , the third  $S_2$  consists of words ending with  $c$  followed by another letter that is not  $c$ , and so on, up to the final set  $S_n$  which consists of words beginning with a  $c$  followed by  $n - 1$  other symbols which are not  $c$ . Since  $X \subseteq Y$ , there exist words in  $L(Y)$  of every length without any  $c$  symbols, implying that  $S_0 \neq \emptyset$ . Therefore, there must exist  $k > 0$  so that all follower sets (in  $Y$ ) of words in  $S_k$  are also follower sets (in  $Y$ ) of some word in  $S_i$  for some  $i < k$ ; else each of the  $n + 1$  sets  $S_i$  would contribute a follower set not in any previous one, contradicting  $|F_Y(n)| \leq n$ .

Let  $w$  be a word in  $L(Y)$  of length at least  $k$ . Our goal is to show that  $F_Y(w)$  is either equal to one of the finitely many follower sets of words without a  $c$  or to the follower set of a word of length less than  $k$ . Clearly, if  $w$  does not contain a  $c$ , we are done, so suppose  $w$  contains the letter  $c$ . As noted earlier,  $F_Y(w)$  is unchanged if all letters before the last occurrence of  $c$  are removed from  $w$ . If this removal results in a word of length less than  $k$ , then again we are done. So let us proceed under the assumption that  $w$  begins with  $c$ , has length  $k$  or greater and contains no other  $c$  symbols.

Let  $p$  denote the  $k$ -letter prefix of  $w$ . Since  $p$  begins with  $c$ ,  $p$  can be arbitrarily extended backwards in any legal way to yield an  $n$ -letter word  $p'$  which has the same follower set as  $p$ . Note that  $p' \in S_k$ , and so there exists  $i < k$  and  $p'' \in S_i$  so that  $F_Y(p) = F_Y(p') = F_Y(p'')$ . There are two cases. If  $i \neq 0$ , then we may again remove the letters of  $p''$  before the final  $c$  symbol to yield a word  $p'''$  of length  $i < k$  for which  $F_Y(p) = F_Y(p''')$ . Then, we replace the prefix  $p$  of  $w$  by  $p'''$  to yield a new word  $w'$  with strictly smaller length, which still begins with a  $c$  and contains no other  $c$  symbols, and for which

$F_Y(w) = F_Y(w')$  by Lemma 3.1.5. We then repeat the above steps. If at each step,  $i \neq 0$ , then eventually  $w$  will be shortened to a word of length less than  $k$  with the same follower set in  $Y$ , of which there are clearly only finitely many.

The only other case is that at some point, the prefix of length  $k$  has the same follower set in  $Y$  as a word in  $S_0$ . Then, that prefix can be replaced by the word in  $S_0$ , yielding a word with no  $c$  symbols with the same follower set in  $Y$  as  $w$ . Again, we note that there are only finitely many follower sets in  $Y$  of words not containing  $c$ . We have then shown that  $F_Y(w)$  (for arbitrary  $w$  of length at least  $k$ ) has follower set in  $Y$  from a finite collection (namely all follower sets in  $Y$  of words with no  $c$  and all follower sets in  $Y$  of words with length at most  $k - 1$ ), which implies that  $Y$  is sofic.

□

**Remark 3.1.13.** *Though the hypotheses of Theorem 3.1.12 may seem strong, the class of subshifts which satisfy them is large, including all so-called  $S$ -gap shifts [7] and the reverse context-free shift of [11] (with  $X = \{a, b\}^{\mathbb{Z}}$  and  $c = c$ ).*

**Example 3.1.14.**  *$S$ -gap shifts are sofic if and only if the set  $S$  is the union of a finite set and an arithmetic progression. Using  $X = 0^{\infty}$  and  $c = 1$ , every  $S$ -gap shift can be constructed as in Thm. 3.1.12, and so every  $S$ -gap shift satisfies Conjecture 1.*

## 3.2 Decreases in Follower and Extender Set Sequences

While Section 3.1 focused on possible connections between complexity sequences and follower set sequences, the primary goal of this section is to demonstrate an important difference between the two—while complexity sequences must be monotone increasing, we will construct a class of examples which show that follower, predecessor, and extender set sequences may decrease. However, in the interest of classifying what sequences may appear as follower, predecessor, and extender set sequences, we will first show that for a sofic shift  $X$ , the follower and extender set sequences must be eventually periodic. (Of course this is also true of the predecessor set sequence, though as suggested by Remark 3.1.2, we neglect the argument). We begin with a lemma which is reminiscent of the pumping lemma for regular languages ([6]).

**Lemma 3.2.1.** *Let  $X$  be a sofic shift, define  $p$  to be one greater than the total number of extender sets in  $X$ , and define  $p_0$  to be one greater than the total number of follower sets in  $X$ . Then all words  $w$  in  $L(X)$  of length  $n \geq p$  may be written as  $w = xyz$ , where  $|y| \geq 1$  and the word  $xy^iz$  has the same extender set as the word  $w$  for all  $i \in \mathbb{N}$ . Furthermore, all words  $w$  in  $L(X)$  of length  $n \geq p_0$  may be written as  $w = xyz$ , where  $|y| \geq 1$  and the word  $xy^iz$  has the same follower set as the word  $w$  for all  $i \in \mathbb{N}$ .*

*Proof.* Since  $X$  is sofic,  $X$  has only finitely many extender sets. Let  $p$  be one greater than the number of extender sets in  $X$ . Since  $X$  is presented by a finite labeled graph  $\mathcal{G}$ , and there are only  $|V(\mathcal{G})|^2$  possible pairs of vertices in

$\mathcal{G}$ , and each extender set must correspond to a non-empty set of pairs, we have that  $p \leq 2^{(|V(\mathcal{G})|^2)}$ . Let  $w$  be a word in  $L(X)$  of length  $n \geq p$ . Consider the prefixes of  $w$ . Since  $w$  is of length at least  $p$ , there exist two prefixes of  $w$  (one necessarily a strict subword of the other) with the same extender set, say  $x$  and  $xy$ , where  $|y| \geq 1$ . Then for any pair  $(s, u)$  of infinite sequences,  $sxu$  is a point of  $X$  if and only if  $sxyu$  is a point of  $X$  also. Call the remaining portion of  $w$  by  $z$  so that  $w = xyz$ . (We may have  $|z| = 0$ ).

Now, let  $(s, u)$  be in the extender set of  $w$ , that is, that  $swu$  is a point of  $X$ . But  $swu = sxyz u$ , so  $(s, yzu)$  is in the extender set of  $x$ . By above, then,  $(s, yzu)$  is in the extender set of  $xy$  also, that is, that  $sxyyzu$  is a point of  $X$ . Hence,  $(s, u)$  is in the extender set of  $xyz$ . So  $E_X(xyz) \subseteq E_X(xyyz)$ . On the other hand, if  $(s, u)$  is in the extender set of  $xyyz$ , then  $sxyyzu$  is a point of  $X$ , and so  $(s, yzu)$  is in the extender set of  $xy$ , and therefore the extender set of  $x$ . Thus  $sxyz u$  is a point of  $X$ , and so  $(s, u)$  is in the extender set of  $xyz = w$ . Therefore  $E_X(xyz) = E_X(xyyz)$ . Applying this argument repeatedly gives that  $E_X(xyz) = E_X(xy^i z)$  for any  $i \in \mathbb{N}$ .

Letting  $p_0$  be one greater than the total number of follower sets in  $X$ , an identical argument gives the corresponding result for follower sets.  $\square$

With this lemma, we may prove Theorem 1.0.3, which states that the follower and extender set sequences of sofic shift spaces are eventually periodic:

*Proof of Theorem 1.0.3.* Let  $X$  be a one-dimensional sofic shift. We prove that the sequences  $\{F_X(n)\}$  and  $\{E_X(n)\}$  (that is, the sequences which record not the *number* of follower sets of each length, but instead the identities of those sets) are eventually periodic, which will trivially imply our desired result,

eventual periodicity of the follower set sequence  $\{|F_X(n)|\}$  and the extender set sequence  $\{|E_X(n)|\}$ .

Let  $w$  be a word of length  $p$  in  $L(X)$ , where  $p$  is defined as in Lemma 3.2.1. Then by Lemma 3.2.1,  $w = xy_wz$  where  $|y_w| \geq 1$  and  $xy_w^i z$  has the same extender set as  $w$  for all  $i \in \mathbb{N}$ . Let  $k = \text{lcm}\{|y_w| \mid w \in L_p(X)\}$ . Since the longest such a word  $y_w$  could be is  $p$ , we have  $k \leq p!$ . Clearly for any word  $w \in L_p(X)$ , there is an  $i \in \mathbb{N}$  such that  $xy_w^i z \in L_{p+k}(X)$ . Therefore, every extender set in  $E_X(p)$  is also an extender set in  $E_X(p+k)$ , so  $E_X(p) \subseteq E_X(p+k)$ .

Now, let  $w$  be a word of length  $n > p$  in  $L(X)$ . Then  $w$  has some word  $w' = w_1 \dots w_p \in L_p(X)$  as a prefix. Applying Lemma 3.2.1 to  $w'$  as above, we get a word  $w''$  of length  $p+k$  with the same extender set as  $w'$ . If  $(s, u)$  is in the extender set of  $w = w'w_{p+1} \dots w_n$ , then  $sw'w_{p+1} \dots w_n u$  is a point of  $X$ , and so  $(s, w_{p+1} \dots w_n u)$  is in the extender set of  $w'$ . Since  $w'$  and  $w''$  have the same extender set,  $sw''w_{p+1} \dots w_n u$  is a point of  $X$ , and  $(s, u)$  is in the extender set of  $w''w_{p+1} \dots w_n$ . Similarly, if  $(s, u)$  is in the extender set of  $w''w_{p+1} \dots w_n$ , then  $(s, u)$  is in the extender set of  $w$  as well. Therefore  $w''w_{p+1} \dots w_n$  is a word in  $X$  of length  $n+k$  with the same extender set as  $w$ . Hence, every extender set in  $E_X(n)$  is an extender set in  $E_X(n+k)$ . So  $E_X(n) \subseteq E_X(n+k)$  for any  $n \geq p$ .

But sofic shifts only have finitely many extender sets, so eventually, the sequence  $\{|E_X(n+jk)|\}_{j \in \mathbb{N}}$  must stop growing. Thus, we have  $E_X(n) = E_X(n+k)$  for all sufficiently large  $n$ , and the sequence  $\{E_X(n)\}$  is eventually periodic with period  $k$ , where  $k \leq p!$ . Certainly, this implies that the extender set sequence is eventually periodic with period  $k$  as well.

Suppose  $n \geq p$  and  $E_X(n) = E_X(n+k)$ . Then we claim that  $E_X(n+k) = E_X(n+2k)$ : By above, we have  $E_X(n+k) \subseteq E_X(n+2k)$ . Suppose  $w$  is a word of length  $n+2k$ , say  $w = w_1w_2\dots w_{n+k}w_{n+k+1}\dots w_{n+2k}$ . Then, because  $E_X(n) = E_X(n+k)$ , there exists a word  $z$  of length  $n$  such that  $z$  and  $w_1w_2\dots w_{n+k}$  have the same extender set. Let  $(s, u)$  be in the extender set of  $w$ . Then  $sw_1w_2\dots w_{n+k}w_{n+k+1}\dots w_{n+2k}u$  is a point of  $X$ , so  $(s, w_{n+k+1}\dots w_{n+2k}u)$  is in the extender set of  $w_1w_2\dots w_{n+k}$ , and thus in the extender set of  $z$ . So  $szw_{n+k+1}w_{n+k+2}\dots w_{n+2k}u$  is a point of  $X$ . Hence  $(s, u)$  is in the extender set of  $zw_{n+k+1}w_{n+k+2}\dots w_{n+2k}$ , a word of length  $n+k$ . Similarly, if  $(s, u)$  is in the extender set of  $zw_{n+k+1}w_{n+k+2}\dots w_{n+2k}$ , then  $(s, u)$  is in the extender set of  $w$ , giving  $E_X(w) = E_X(zw_{n+k+1}w_{n+k+2}\dots w_{n+2k})$ . Therefore we have  $E_X(n+2k) \subseteq E_X(n+k)$  and we may conclude that  $E_X(n+k) = E_X(n+2k)$ .

Now,  $\{|E_X(n)|\} < p$  for any given  $n \in \mathbb{N}$ . Moreover, we have proven that the sequence  $\{E_X(n+jk)\}_{j \in \mathbb{N}}$  is nondecreasing and nested by inclusion, and once two terms of the sequence are equal, it will stabilize for all larger  $j$ . The sequence  $\{E_X(n+jk)\}_{j \in \mathbb{N}}$  must grow fewer than  $p$  times, so the periodicity of the sequence  $\{E_X(n)\}$  (and thus of  $\{|E_X(n)|\}$ ) must begin before the  $p + pk^{\text{th}}$  term, and  $p + pk \leq p + p(p!) = p(1 + p!)$ .

Again, a similar argument using the follower set portion of Lemma 3.2.1 establishes the corresponding result for follower sets.  $\square$

Next we demonstrate the existence of sofic shifts with follower set sequences which are not eventually constant. The first example of a shift space whose follower and extender set sequence are not monotone increasing is due to Martin Delacourt (page 8 of [9]). The following construction is loosely based on his example.

Given  $\ell \in \mathbb{N}$  and  $S \subset \{0, 1, \dots, \ell - 1\}$ , construct an irreducible graph  $\mathcal{G}_{\ell, S}$  in the following way: First, place edges labeled  $p, q$  and  $b$  as below, followed by a loop of  $\ell$ -many edges labeled  $a$ . We will refer to the initial vertex of the edge  $p$  as “Start.”

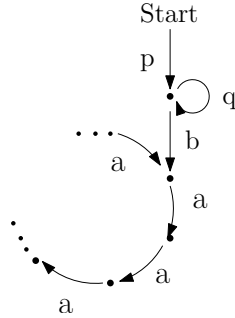


Figure 3.1: Step 1 for constructing  $\mathcal{G}_{\ell, S}$

Choose a fixed  $i^* \in S$ . Then for every  $i \in S$ ,  $i \neq i^*$ , place two consecutive edges  $c_i$  and  $d_i$ , such that the initial vertex of  $c_i$  is the  $(i - 2 \pmod{\ell})^{th}$  vertex of the loop of edges labeled  $a$  (where we make the convention that the terminal vertex of the edge  $b$  is the  $0^{th}$  vertex of the loop, the next vertex the  $1^{st}$  vertex of the loop, and so on) and the terminal vertex of  $d_i$  is “Start.” For  $i^*$ , we still add the edge  $c_{i^*}$ , but follow it instead by another edge labeled  $b$  and another loop of  $\ell$ -many edges labeled  $a$ :



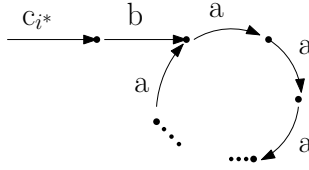


Figure 3.2: Step 2 for constructing  $\mathcal{G}_{\ell,S}$

Again,  $\forall i \in S$ , at the  $(i - 2 \pmod{\ell})^{\text{th}}$  vertex of the new loop, add an edge  $c_i$ , and if  $i \neq i^*$ , follow with an edge  $e_i$  returning to Start. After  $c_{i^*}$ , add a third loop of  $\ell$ -many edges labeled  $a$ :

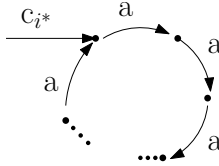


Figure 3.3: Step 3 for constructing  $\mathcal{G}_{\ell,S}$

Finally, for each  $i \in S$ , (including  $i^*$ ), add an edge  $c_i$  from the  $(i - 2 \pmod{\ell})^{\text{th}}$  vertex of the third loop returning to Start. The resulting graph is  $\mathcal{G}_{\ell,S}$ . Due to choice of  $i^*$ , there are  $|S|$ -many possible graphs  $\mathcal{G}_{\ell,S}$ ; the results of this paper will hold for any of them.

We first make some basic observations about the graphs  $\mathcal{G}_{\ell,S}$ . It is easy to check that for any  $\ell, S$ , the graph  $\mathcal{G}_{\ell,S}$  will be irreducible, right-resolving and left-resolving, follower-separated and extender-separated. We furthermore observe that the graph is primitive:

**Lemma 3.2.2.** *For any  $\ell \in \mathbb{N}$ ,  $S \subset \{0, 1, \dots, \ell - 1\}$ , the graph  $\mathcal{G}_{\ell,S}$  is primitive, and the primitivity distance of  $\mathcal{G}_{\ell,S}$  is at most  $3\ell + 3$ .*

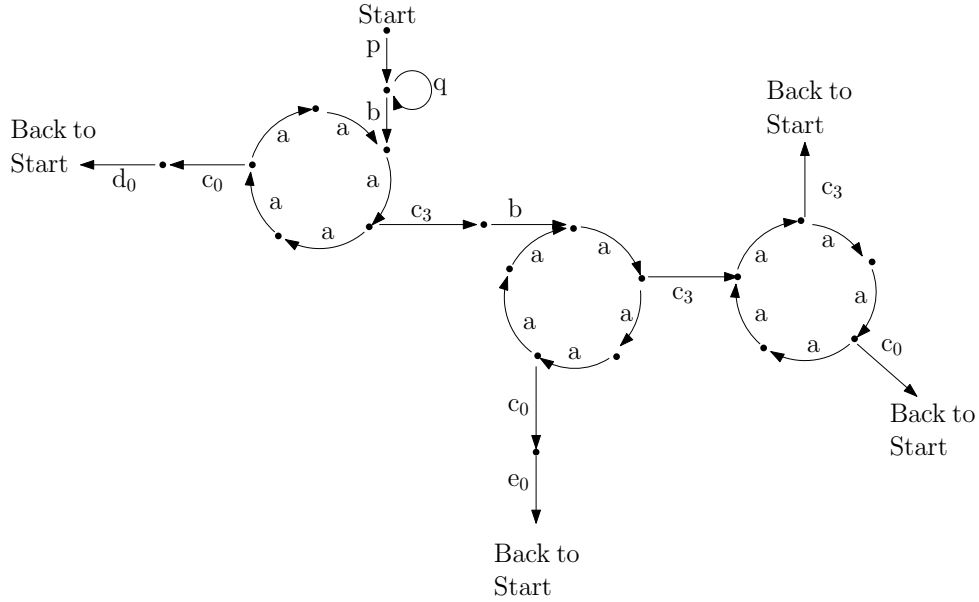


Figure 3.4:  $\mathcal{G}_{5, \{0,3\}}$  with  $i^* = 3$  (where we use the convention that edges terminating at “Back to Start” return to the initial vertex of the edge  $p$ .)

*Proof.* Any irreducible graph with a self-loop is primitive as in Remark 2.0.34, so due to the self-loop labeled  $q$ ,  $\mathcal{G}_{\ell,S}$  is primitive. So long as a path passes the vertex at which  $q$  is anchored, that path may be inflated to any greater length by following the self-loop  $q$  repeatedly. Given any two vertices  $I$  and  $J$  in  $\mathcal{G}_{\ell,S}$ , we may clearly get from  $I$  to  $J$ , while being certain to also pass the vertex anchoring  $q$ , by traveling through each loop of edges labeled  $a$  at most once (and following at most  $\ell - 1$  of the edges in each loop), and using no more than six letters total for the connecting paths between the loops. Thus, the longest path required to travel from one vertex to another in  $\mathcal{G}_{\ell,S}$ , requiring that such a path pass through the vertex which anchors  $q$ , is of length at most  $3(\ell - 1) + 6 = 3\ell + 3$ . For instance, if  $S = \{1\}$ , the shortest possible path from the initial vertex of  $p$  to itself is labeled by the word  $pb a^{\ell-1} c_1 b a^{\ell-1} c_1 a^{\ell-1} c_1$ , which clearly travels through the vertex anchoring  $q$ . This is a sort of “worst-

case” scenario, where, because  $|S| = 1$ , the shortest path requires traveling through all three loops of the graph, and with  $i^* = 1$ , the exit of each loop is as far away from the entry point as possible.  $\square$

Any graph  $\mathcal{G}_{\ell,S}$  created by the above construction will present a shift with a follower set sequence and extender set sequence having eventual period  $\ell$  with a difference in limsup and liminf of 1:

**Theorem 3.2.3.** *Let  $\ell \in \mathbb{N}$  and  $S \subset \{0, 1, \dots, \ell - 1\}$ . Then the shift  $X_{\mathcal{G}_{\ell,S}}$  has  $3\ell + 3|S| + 4$  follower sets for words of length  $n$  where  $n \geq \ell + 2$  and  $n \pmod{\ell} \in S$ , and only  $3\ell + 3|S| + 3$  follower sets for words of length  $n'$  where  $n' \geq \ell + 2$  and  $n' \pmod{\ell} \notin S$ . Furthermore, the shift  $X_{\mathcal{G}_{\ell,S}}$  has  $(3\ell + 2|S| + 1)^2 + |S| + 3$  extender sets for words of length  $n$  where  $n \geq 3\ell + 3$  and  $n \pmod{\ell} \in S$ , and only  $(3\ell + 2|S| + 1)^2 + |S| + 2$  extender sets for words of length  $n'$  where  $n' \geq 3\ell + 3$  and  $n' \pmod{\ell} \notin S$ .*

*Proof.* Let  $n \geq \ell + 2$ . Let  $\mathcal{G} = \mathcal{G}_{\ell,S}$  as defined above. We will find the number of follower sets and extender sets of words of length  $n$  in  $X_{\mathcal{G}}$  (though for one part of the extender set case, we will need to require  $n \geq 3\ell + 3$ ). Each follower set  $F_X(w)$  is uniquely determined by the set of terminal vertices of paths labeled  $w$  in  $\mathcal{G}$ , and each extender set  $E_X(w)$  is determined by the set of pairs  $\{I \rightarrow T\}$  of initial and terminal vertices of paths labeled  $w$  in  $\mathcal{G}$ .

Since in  $\mathcal{G}$  the words  $p$ ,  $q$ ,  $c_i^*b$ ,  $d_i$ ,  $e_i$ , and  $c_i^*a$  are right-synchronizing, the longest path required to get from a right-synchronizing word to any vertex of  $\mathcal{G}$  is  $\ell + 2$ . Because  $n \geq \ell + 2$ , and the graph  $\mathcal{G}$  is irreducible and right-resolving, every singleton represents the follower set of some word  $w$  of length  $n$ . There are  $3\ell + 2|S| + 1$  such follower sets, all distinct as  $\mathcal{G}$  is follower-separated.

It is easy to see that the right-synchronizing words listed above are in fact bi-synchronizing. Since the graph  $\mathcal{G}$  is both left- and right-resolving, if a legal word  $w$  contains any bi-synchronizing word, then only one pair  $\{I \rightarrow T\}$  can be the initial and terminal vertices of paths labeled  $w$ . If  $n \geq 3\ell + 3$ , by Lemma 3.2.2 every single pair  $\{I \rightarrow T\}$  corresponds to the extender set of some word  $w$  of length  $n$ . There are  $(3\ell + 2|S| + 1)^2$  such extender sets, all distinct as  $\mathcal{G}$  is extender-separated.

Note that for the graph  $\mathcal{G}$ , recording the labels of two edges beyond any loop of edges labeled  $a$ , whether before or after the loop, results in a bi-synchronizing word. Since any word which would be capable of having a follower set not corresponding to a singleton (or of having an extender set not corresponding to a single pair of initial and terminal vertices) must be one which avoids all bi-synchronizing words, and  $n \geq \ell + 2 > 2$ , any word of length  $n$  with such a follower or extender set must include a string of  $a$ 's, and no more than 1 letter on either side of such a string. So only words of the forms  $a^n$ ,  $ka^{n-1}$ ,  $a^{n-1}k'$ , and  $ka^{n-2}k'$  (where  $k$  and  $k'$  are labels appearing in  $\mathcal{G}$  not equal to  $a$ ) can terminate (or begin) at more than one vertex.

The word  $a^n$  has 1 follower set, corresponding to all  $3\ell$  vertices involved in loops of edges labeled  $a$ . This follower set is distinct from those corresponding to singletons, for which we have previously accounted. Similarly, the extender set of the word  $a^n$  is distinct from those for which we have previously accounted.

The label  $a$  is only followed in  $\mathcal{G}$  by the labels  $a$  and  $c_i$  for all  $i \in S$ . For each  $c_i$ , the word  $a^{n-1}c_i$  has a unique follower set corresponding to three terminal vertices, one for each loop in which the  $a^{n-1}$  may occur. Thus there

are  $|S|$ -many follower sets of this form, all distinct from previous follower sets. Similarly, there are  $|S|$ -many extender sets of this form, all distinct from previous extender sets as well.

The label  $a$  is only preceded in  $\mathcal{G}$  by the labels  $a$ ,  $b$ , and  $c_{i^*}$ . Since  $c_{i^*}a$  is bi-synchronizing, the follower set for the word  $c_{i^*}a^{n-1}$  corresponds to a singleton and has already been counted. The word  $ba^{n-1}$  has a follower set corresponding to two terminal vertices, one in each loop of edges labeled  $a$  which is preceded by  $b$ . Hence there is 1 additional distinct follower set of this form, and again this behavior is mirrored by the extender sets—there is 1 additional distinct extender set for the word  $ba^{n-1}$ .

Finally, based on our above observations, if a word of the form  $ka^{n-2}k'$  is to have a follower set corresponding to a greater number of vertices than one, that word must be of the form  $ba^{n-2}c_i$ . By construction, a path with this label only exists in  $\mathcal{G}$  if  $n \pmod{\ell} \in S$ . If such a path exists, it contributes a single new follower set corresponding to two terminal vertices, each one edge past a loop of edges labeled  $a$  that is preceded by the label  $b$ . This follower set cannot repeat one that we already found: if  $i \neq i^*$ , then the follower set of  $ba^{n-2}c_i$  is exactly the set of all legal sequences beginning with  $d_i$  or  $e_i$ , clearly not equal to the follower set of any other word of length  $\ell$ . If  $i = i^*$ , the follower set of  $ba^{n-2}c_i$  contains sequences beginning with each of the letters  $a$  and  $b$ , but no other letter, again setting it apart from any other follower set previously discussed. Similarly, the word  $ba^{n-2}c_i$  contributes a single new extender set for any length  $n$  for which a path with this label exists.

Therefore, in  $X_{\mathcal{G}}$ , if  $n \geq \ell + 2$  and  $n \pmod{\ell} \in S$ , there are  $3\ell + 2|S| + 1 + 1 + |S| + 1 + 1 = 3\ell + 3|S| + 4$  follower sets of words of length  $n$ , while

if  $n' \geq \ell + 2$  and  $n' \pmod{\ell} \notin S$ , there are only  $3\ell + 3|S| + 3$  follower sets of words of length  $n'$ . Moreover, if  $n \geq 3\ell + 3$  and  $n \pmod{\ell} \in S$ , there are  $(3\ell + 2|S| + 1)^2 + 1 + |S| + 1 + 1 = (3\ell + 2|S| + 1)^2 + |S| + 3$  extender sets of words of length  $n$ , while if  $n' \geq 3\ell + 3$  and  $n' \pmod{\ell} \notin S$ , there are only  $(3\ell + 2|S| + 1)^2 + |S| + 2$  extender sets of words of length  $n'$ .

□

**Example 3.2.4.** *For the shift presented by the graph in Figure 3.4, for  $n \geq 7$ , the follower set sequence oscillates between  $|F_X(n)| = 3(5) + 3(2) + 4 = 25$  if  $n \equiv 0$  or  $3 \pmod{5}$ , and  $|F_X(n)| = 24$  if  $n \equiv 1, 2,$  or  $4 \pmod{5}$ . Moreover, for  $n \geq 18$ , the extender set sequence oscillates between  $|E_X(n)| = (3(5) + 2(2) + 1)^2 + (2) + 3 = 405$  if  $n \equiv 0$  or  $3 \pmod{5}$ , and  $|E_X(n)| = 404$  if  $n \equiv 1, 2,$  or  $4 \pmod{5}$ .*

We have now demonstrated the existence of sofic shifts whose follower set sequences and extender set sequences eventually oscillate between two different (but adjacent) values. We may furthermore combine these graphs, forming new graphs presenting shifts whose follower set sequences and extender set sequences oscillate by more than 1:

**Theorem 3.2.5.** *Let  $\mathcal{G}_1$  and  $\mathcal{G}_2$  be two finite, irreducible, right-resolving, left-resolving, primitive, extender-separated labeled graphs with disjoint label sets, each containing a self-loop labeled by a bi-synchronizing letter  $q_1$  and  $q_2$  respectively. Let  $I_1$  be the anchoring vertex of  $q_1$  in  $\mathcal{G}_1$  and  $I_2$  be the anchoring vertex of  $q_2$  in  $\mathcal{G}_2$ . Let  $x, y$  be letters not in the label set of  $\mathcal{G}_1$  or  $\mathcal{G}_2$ . Construct a new graph  $\mathcal{G}$  by taking the disjoint union of  $\mathcal{G}_1$  and  $\mathcal{G}_2$  and adding an edge labeled  $x$  beginning at  $I_1$  and terminating at  $I_2$  and an edge labeled  $y$  begin-*

ning at  $I_2$  and terminating at  $I_1$ . Then  $\mathcal{G}$  is finite, irreducible, right-resolving, left-resolving, primitive, extender-separated, contains a self-loop labeled with a bi-synchronizing letter, and for any  $n \in \mathbb{N}$ ,

$$|F_{X_{\mathcal{G}}}(n)| = |F_{X_{\mathcal{G}_1}}(n)| + |F_{X_{\mathcal{G}_2}}(n)|.$$

Moreover, for any  $n$  greater than twice the maximum of the primitivity distances of  $\mathcal{G}_1$  and  $\mathcal{G}_2$ ,

$$|E_{X_{\mathcal{G}}}(n)| = |E_{X_{\mathcal{G}_1}}(n)| + |E_{X_{\mathcal{G}_2}}(n)| + 2|V(\mathcal{G}_1)| \cdot |V(\mathcal{G}_2)|.$$

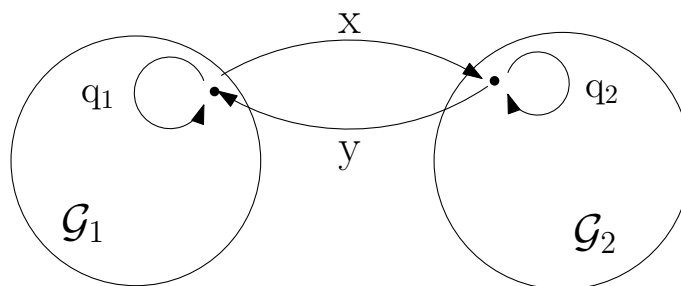


Figure 3.5: The graph  $\mathcal{G}$  constructed as in Theorem 3.2.5

*Proof.* The reader may check that  $\mathcal{G}$  is finite, irreducible, right-resolving, left-resolving, primitive, extender-separated and contains a self-loop labeled with a bi-synchronizing letter. We first check that this construction does not cause any collapsing of follower or extender sets. That is, if two words  $w$  and  $v$  had distinct follower or extender sets in  $X_{\mathcal{G}_1} \sqcup X_{\mathcal{G}_2}$ , then they have distinct follower or extender sets in  $X_{\mathcal{G}}$ . We present the argument for follower sets; the argument for extender sets is similar.

If for two words  $w$  and  $v$  in  $L_n(X_{\mathcal{G}_1}) \sqcup L_n(X_{\mathcal{G}_2})$ , we have  $F(w) \neq F(v)$ , then there exists a sequence  $s$  in  $X_{\mathcal{G}_1}$  or  $X_{\mathcal{G}_2}$  which may follow one word but not the other. Without loss of generality, say that  $s$  may follow  $w$  but not  $v$ . Such a sequence may still follow  $w$  in  $\mathcal{G}$  by following the same path labeled  $s$  which followed  $w$  in  $\mathcal{G}_1$  or  $\mathcal{G}_2$  to begin with. The sequence  $s$  still may not follow  $v$ , as no path existed in the parent graph  $\mathcal{G}_1$  or  $\mathcal{G}_2$  labeled  $s$  following  $v$ , and any new paths introduced by our construction may not be labeled  $s$ , as  $s$  did not contain the letters  $x$  or  $y$ . Hence  $|F_{X_{\mathcal{G}_1}}(n)| + |F_{X_{\mathcal{G}_2}}(n)| \leq |F_{X_{\mathcal{G}}}(\ell)|$ , and similarly,  $|E_{X_{\mathcal{G}_1}}(n)| + |E_{X_{\mathcal{G}_2}}(n)| \leq |E_{X_{\mathcal{G}}}(n)|$ .

Now we establish that no extra follower sets are introduced by this construction. If two words  $w$  and  $v$  had the same follower set in  $X_{\mathcal{G}_1} \sqcup X_{\mathcal{G}_2}$ , then they certainly exist in the same parent graph,  $\mathcal{G}_1$  or  $\mathcal{G}_2$ ; without loss of generality, say  $\mathcal{G}_1$ . Let  $s$  be some sequence following  $w$  in  $\mathcal{G}$ . If the path in  $\mathcal{G}$  labeled  $s$  is contained within  $\mathcal{G}_1$ , then  $s$  is part of the follower set of  $w$  in  $X_{\mathcal{G}_1}$  and thus, is part of the follower set of  $v$  in  $X_{\mathcal{G}_1}$ . So  $s$  may follow  $v$  in  $\mathcal{G}$ . On the other hand, if  $s$  is presented by a path traveling through both graphs, then  $s$  contains the letter  $x$ . Let  $z$  denote the maximal finite prefix of  $s$  without the letter  $x$ . A path labeled  $z$  follows  $w$  in  $\mathcal{G}_1$ , and since  $z$  must terminate at  $I_1$  in order to be followed by  $x$ , a path labeled  $zq_1$  follows  $w$  in  $\mathcal{G}_1$  as well. Then a path labeled  $zq_1$  must also follow  $v$  in  $\mathcal{G}_1$ , and since  $q_1$  is left-synchronizing, there must exist a path labeled  $z$  following  $v$  in  $\mathcal{G}_1$  terminating at  $I_1$ . Such a path may certainly, then, be followed by  $x$ , and indeed, the remaining portion of  $s$ , so  $s$  is in the follower set of  $v$ . Thus in  $X_{\mathcal{G}}$ , the follower sets of  $w$  and  $v$  remain the same.



Moreover, if a word  $v$  is not in either  $L_n(X_{\mathcal{G}_1})$  or  $L_n(X_{\mathcal{G}_2})$ , then  $v$  includes either the letter  $x$  or  $y$ . Since  $x$  and  $y$  are right-synchronizing, a path labeled  $v$  may terminate only at a single vertex. Due to the fact that  $\mathcal{G}_1$  and  $\mathcal{G}_2$  are each irreducible, right-resolving, and contain a right-synchronizing letter, there exists a word  $w$  in either  $L_n(\mathcal{G}_1)$  or  $L_n(\mathcal{G}_2)$  which terminates at the same unique vertex as paths labeled  $v$ . Since the right-synchronizing letter terminates at the same vertex as  $x$  or  $y$ , depending on the graph, we may construct  $w$  to be of the desired length  $n$  in the following way: Let  $v = v_1v_2\dots v_n$  and let  $v_i$  be the last occurrence of  $x$  or  $y$  in  $v$ . Then set  $w_{i+1}w_{i+2}\dots w_n = v_{i+1}v_{i+2}\dots v_n$ . Create  $w_1\dots w_i$  by replacing  $v_i$  by  $q_1$  or  $q_2$  (choosing the one which terminates at the same place as  $v_i$ ) and then following any path backward in that same parent graph ( $\mathcal{G}_1$  or  $\mathcal{G}_2$ ) to fill in  $i - 1$  labels before  $w_i$ . Then  $w$  is a right-synchronizing word contained entirely in either  $\mathcal{G}_1$  or  $\mathcal{G}_2$  of length  $n$  terminating at the same single vertex as  $v$ , and so  $w$  and  $v$  have the same follower set in  $X_{\mathcal{G}}$ . Therefore the construction introduced no extra follower sets, and  $|F_{X_{\mathcal{G}}}(n)| = |F_{X_{\mathcal{G}_1}}(n)| + |F_{X_{\mathcal{G}_2}}(n)|$ .

This construction also causes no splitting of extender sets: If  $w$  and  $v$  have the same extender set in  $X_{\mathcal{G}_1} \sqcup X_{\mathcal{G}_2}$ , then they certainly exist in the same parent graph,  $\mathcal{G}_1$  or  $\mathcal{G}_2$ ; without loss of generality, say  $\mathcal{G}_1$ . Let  $(s, u)$  be in the extender set of  $w$ . If  $s$  and  $u$  are both contained within  $\mathcal{G}_1$ , then  $(s, u)$  is certainly in the extender set of  $v$  as well. Otherwise, let  $z$  be the maximal suffix of  $s$  with no appearance of the letter  $y$  and  $z'$  be the maximal prefix of  $u$  with no appearance of  $x$ . (Note that in this case one of  $z$  and  $z'$  may be infinite, but not both.) Then  $zwz'$  is contained in  $\mathcal{G}_1$  and since  $w$  and  $v$  have the same extender set in  $\mathcal{G}_1$ , a path labeled  $zwz'$  exists in  $\mathcal{G}_1$  as well. If

$z = s$  but  $z'$  is finite, then as above,  $z'$  may be followed by  $q_1$  in  $\mathcal{G}_1$ , which is left-synchronizing, so there must exist a path labeled  $z'$  following  $sv$  in  $\mathcal{G}_1$  terminating at  $I_1$ . Such a path may certainly, then, be followed by  $x$ , and the remaining portion of  $u$ , and so  $(s, u)$  is in the extender set of  $v$ . Similarly, if  $z$  is finite but  $z' = u$ , then  $z$  may be preceded by  $q_1$ , which is right-synchronizing, so there must exist a path labeled  $z$  preceding  $vu$  in  $\mathcal{G}_1$  beginning at  $I_1$ . Such a path may then be preceded by  $y$ , and the preceding portion of  $s$ , and so  $(s, u)$  is in the extender set of  $v$ . If both  $z$  and  $z'$  are finite, then the path  $q_1 z v z' q_1$  exists in  $\mathcal{G}_1$ , and so a path labeled  $z v z'$  exists in  $\mathcal{G}_1$  beginning and ending at  $I_1$  which then may be extended to an infinite path labeled  $svu$ , so  $(s, u)$  is in the extender set of  $v$ . Thus  $E(w) = E(v)$ .

Finally, if a word  $v$  is not in either  $L_n(X_{\mathcal{G}_1})$  or  $L_n(X_{\mathcal{G}_2})$ , then  $v$  includes either the letter  $x$  or  $y$ . Since  $x$  and  $y$  are bi-synchronizing and  $\mathcal{G}$  is both left- and right-resolving, paths labeled  $v$  have exactly one pair  $\{I \rightarrow T\}$  of initial and terminal vertices. If  $I$  and  $T$  are in the same parent graph  $\mathcal{G}_i$ , we observe that if  $n$  is longer than twice the primitivity distance for  $\mathcal{G}_i$ , we can construct a path  $w$  from  $I$  to  $I_i$ , and a path  $u$  from  $I_i$  to  $T$ , such that the path labeled  $wq_iu$  has length  $n$ . Because  $q_i$  is bi-synchronizing, paths labeled  $wq_iu$  have only one pair of initial and terminal vertices; the exact same initial and terminal vertices as  $v$ , so  $E(v) = E(wq_iu)$ . On the other hand, if  $I$  and  $T$  are in different parent graphs, then  $E(v)$  is certainly not equal to the extender set of any word in  $L_n(X_{\mathcal{G}_1}) \sqcup L_n(X_{\mathcal{G}_2})$ , so the construction did introduce new extender sets of words of length  $n$ , but only at most  $2|V(\mathcal{G}_1)| \cdot |V(\mathcal{G}_2)|$  many of them. Furthermore, if  $n$  is longer than the primitivity distance of  $\mathcal{G}$ , then all  $2|V(\mathcal{G}_1)| \cdot |V(\mathcal{G}_2)|$  such extender sets will be realized, and since  $\mathcal{G}$  is extender-

separated, they will all be distinct. The primitivity distance for  $\mathcal{G}$  is at most one greater than the sum of the primitivity distances of  $\mathcal{G}_1$  and  $\mathcal{G}_2$ , so if  $n$  is greater than twice the maximum of the primitivity distances of  $\mathcal{G}_1$  and  $\mathcal{G}_2$ , we have  $|E_{X_{\mathcal{G}}}(n)| = |E_{X_{\mathcal{G}_1}}(n)| + |E_{X_{\mathcal{G}_2}}(n)| + 2|V(\mathcal{G}_1)| \cdot |V(\mathcal{G}_2)|$ .  $\square$

It is evident that the process outlined in Theorem 3.2.5 may be repeated an arbitrary number of times, and since the constant introduced (0 in the follower set case,  $2|V(\mathcal{G}_1)| \cdot |V(\mathcal{G}_2)|$  in the extender set case) does not depend on  $n$ , we may use this process to increase the oscillations in the follower and extender set sequences of the resulting shift. We formalize this idea in Theorems 1.0.4 and 1.0.5, which state that there exist sofic shifts with follower set sequences and extender set sequences of every eventual period and with any natural number as the difference in  $\limsup$  and  $\liminf$  of the sequence.

*Proof of 1.0.4.* Let  $\mathcal{G}_{\ell,S}$  denote the graph constructed from  $\ell$  and  $S \subset \{0, 1, \dots, \ell - 1\}$  as in Theorem 3.2.3. First construct  $\mathcal{G}_{\ell,A_2 \cup A_3 \cup \dots \cup A_k}$ . By Theorem 3.2.3, this graph will give one more follower set to words of length  $n \geq \ell + 2$  and  $n \pmod{\ell} \in A_2 \cup A_3 \cup \dots \cup A_k$  than to words of length  $n \geq \ell + 2$  and  $n \pmod{\ell} \in A_1$ .

Now, as  $\mathcal{G}_{\ell,A_2 \cup A_3 \cup \dots \cup A_k}$  is finite, irreducible, right-resolving, left-resolving, primitive, extender-separated, and contains a self-loop labeled with the bi-synchronizing letter  $q$ , we may use the process defined in Theorem 3.2.5 to join together  $r_2$  many copies of  $\mathcal{G}_{\ell,A_2 \cup A_3 \cup \dots \cup A_k}$  only by giving each copy a disjoint set of labels. Call the resulting graph  $\mathcal{G}_2$ . For each  $n$ , the number of follower sets of words of length  $n$  in  $\mathcal{G}_2$  is the sum of the number of follower sets of words of length  $n$  in each of the  $r_2$  copies of  $\mathcal{G}_{\ell,A_2 \cup A_3 \cup \dots \cup A_k}$ . Therefore, the

graph  $\mathcal{G}_2$  gives  $r_2$  more follower sets to words of length  $n \geq \ell + 2$  with  $n \pmod{\ell} \in A_2 \cup A_3 \cup \dots \cup A_k$  than to words of length  $n \geq \ell + 2$  with  $n \pmod{\ell} \in A_1$ .

Using the same process, we may now join onto  $\mathcal{G}_2$  another  $(r_3 - r_2)$  many copies of the graph  $\mathcal{G}_{\ell, A_3 \cup A_4 \cup \dots \cup A_k}$ , and call the resulting graph  $\mathcal{G}_3$ . Now, words of length  $n \geq \ell + 2$  where  $n \pmod{\ell} \in A_3 \cup A_4 \cup \dots \cup A_k$  will have  $(r_3 - r_2) + r_2 = r_3$  more follower sets than words of length  $n \geq \ell + 2$  where  $n \pmod{\ell} \in A_1$ , while words of length  $n \geq \ell + 2$  where  $n \pmod{\ell} \in A_2$  will have only  $r_2$  greater follower sets than words of length  $n \geq \ell + 2$  where  $n \pmod{\ell} \in A_1$ .

Continue on this way, adjoining next  $(r_4 - r_3)$  copies of  $\mathcal{G}_{\ell, A_4 \cup A_5 \cup \dots \cup A_k}$  to make  $\mathcal{G}_4$ , and so forth, terminating after constructing  $\mathcal{G}_k$ . The graph will clearly be irreducible, and in  $\mathcal{G}_k$ , for each  $1 \leq j \leq k$ , words of length  $n \geq \ell + 2$  where  $n \pmod{\ell} \in A_j$  will have  $r_j$  more follower sets than words of length  $n \geq \ell + 2$  where  $n \pmod{\ell} \in A_1$ . That is, if  $m$  is defined to be the number of follower sets of words of length  $n \geq \ell + 2$  where  $n \pmod{\ell} \in A_1$ , then words of length  $n \geq \ell + 2$  where  $n \pmod{\ell} \in A_j$  will have  $m + r_j$  many follower sets in  $X_{\mathcal{G}_k}$ .

Using the formula established in Theorem 3.2.3, we see that for each  $\mathcal{G}_{\ell, S}$ , words of lengths  $n \geq \ell + 2$  whose residue classes are not in  $S$  have  $3\ell + 3|S| + 3$  many follower sets. Since  $A_1 \subseteq S^c$  for every graph used in the construction of

$\mathcal{G}_k$ , words of length  $n \geq \ell + 2$  where  $n \pmod{\ell} \in A_1$  must have the following number of follower sets:

$$\begin{aligned}
m &= r_2(3\ell + 3|A_2 \cup A_3 \cup \dots \cup A_k| + 3) \\
&\quad + (r_3 - r_2)(3\ell + 3|A_3 \cup A_4 \cup \dots \cup A_k| + 3) + \dots \\
&\quad + (r_k - r_{k-1})(3\ell + 3|A_k| + 3) \\
&= r_2(3\ell + 3 \sum_{j=2}^k |A_j| + 3) + (r_3 - r_2)(3\ell + 3 \sum_{j=3}^k |A_j| + 3) + \dots \\
&\quad + (r_k - r_{k-1})(3\ell + 3 \sum_{j=k}^k |A_j| + 3) \\
&= \sum_{i=2}^k (r_i - r_{i-1})(3\ell + 3 \sum_{j=i}^k |A_j| + 3).
\end{aligned}$$

Furthermore, since for all  $i \geq 2$ , we have  $\sum_{j=i}^k |A_j| < \ell$ , we get that

$$\begin{aligned}
m &< \sum_{i=2}^k (r_i - r_{i-1})(6\ell + 3) \\
&= (6\ell + 3) \sum_{i=2}^k (r_i - r_{i-1}) \\
&= (6\ell + 3)r_k.
\end{aligned}$$

□

This theorem shows that we may construct a sofic shift whose follower set sequence follows any desired oscillation scheme, increasing or decreasing by specified amounts at specified lengths  $n$ , and repeating with any desired

eventual period. A similar result, Theorem 1.0.5, holds for extender set sequences, though the bounds for  $m = \liminf\{|E_X(n)|\}$  and for the start of the periodicity of the sequence are different:

*Proof of Theorem 1.0.5.* We follow the same construction as in the proof of Theorem 1.0.4, combining  $r_k$ -many graphs using Theorem 3.2.5 to construct the graph  $\mathcal{G}_k$ . Then, in  $\mathcal{G}_k$ , for each  $1 \leq j \leq k$ , sufficiently long words of length  $n$  where  $n \pmod{\ell} \in A_j$  will have  $r_j$  more extender sets than sufficiently long words of length  $n$  where  $n \pmod{\ell} \in A_1$ . That is, if  $m$  is defined to be the number of extender sets of sufficiently long words of length  $n$  where  $n \pmod{\ell} \in A_1$ , then words of sufficient length  $n$  where  $n \pmod{\ell} \in A_j$  will have  $m + r_j$  many extender sets in  $X_{\mathcal{G}_k}$  for all  $1 \leq j \leq k$ .

To discover what length is sufficient for periodicity of the extender set sequence to begin, we observe that for every graph  $\mathcal{G}$  used in the construction of  $\mathcal{G}_k$ , the primitivity distance of  $\mathcal{G}$  is less than or equal to  $3\ell + 3$  as in Lemma 3.2.2. Note that, since  $\ell \geq 1$ , we have  $3\ell + 3 \leq 7\ell - 1$ . (In fact, in any interesting case,  $\ell \geq 2$ , so  $3\ell + 3 \leq 5\ell - 1$ , but  $\ell = 1$  certainly may be chosen as a trivial case, where  $S = \{0\}$  necessarily). By Theorem 3.2.5, the eventual periodicity of the extender set sequence of the combination of two graphs begins before the  $2z + 1^{\text{st}}$  term, where  $z$  is the maximum of the primitivity distances of the two graphs. So, when adding two graphs together in this construction, we get that the eventual periodicity begins before  $(3\ell + 3) + (3\ell + 3) + 1 \leq (7\ell - 1) + (7\ell - 1) + 1 = 14\ell - 1$ . (We observe that the primitivity distance of the resulting graph will also be less than  $14\ell - 1$ ). Since  $\ell$  is the same for each graph involved in the construction, it does not matter which type of graph we are adding at each step, whether a copy of  $\mathcal{G}_{\ell, A_2 \cup \dots \cup A_k}$ ,  $\mathcal{G}_{\ell, A_3 \cup \dots \cup A_k}$ , up to  $\mathcal{G}_{\ell, A_k}$ .

Though we have performed the same construction as in Theorem 1.0.4, we add our graphs in a more efficient order to minimize the effect of the constant  $2|V(\mathcal{G}_1)| \cdot |V(\mathcal{G}_2)|$ . First consider the case where  $r_k$  is a power of 2. Then we may choose to construct  $\mathcal{G}_k$  in such a way that at each step, we add two graphs each made up of the same number of components. (2 graphs each consisting of 2 components to make 4, 2 graphs each consisting of 4 components to make 8, and so on). Then if  $a_i$  is an upper bound for the start of primitivity at the  $i^{\text{th}}$  step, an upper bound for the primitivity at the  $i + 1^{\text{st}}$  step is  $2(a_i) + 1$ . Then letting  $a_1 = 7\ell - 1$ , the value of the sequence  $a_i = 2(a_{i-1}) + 1$  at  $i = \log_2(r_k) + 1$  will give an upper bound for the start of primitivity for  $\mathcal{G}_k$ , since we must add together two peices of equal components  $\log_2(r_k)$  times to construct  $\mathcal{G}_k$ .

We claim that for all  $i$ ,  $a_i = 2^{i-1}(7\ell) - 1$ . This is trivially true for the base case,  $i = 1$ . By induction, suppose  $a_{i-1} = 2^{i-2}(7\ell) - 1$ . Then

$$\begin{aligned}
a_i &= 2(a_{i-1}) + 1 \\
&= 2(2^{i-2}(7\ell) - 1) + 1 \\
&= 2^{i-1}(7\ell) - 2 + 1 \\
&= 2^{i-1}(7\ell) - 1.
\end{aligned}$$

Thus, when  $r_k$  is a power of 2, the primitivity of the extender set sequence of  $\mathcal{G}_k$  begins before  $a_{\log_2(r_k)+1} = 2^{\log_2(r_k)}(7\ell) - 1 = 7\ell r_k - 1$ .

Now, the upper bound for the beginning of the periodicity of the extender set sequence certainly increases as  $r_k$  increases—increasing  $r_k$  means adding more graphs to construct  $\mathcal{G}_k$ —and so, since  $r_k \leq 2^{\lceil \log_2(r_k) \rceil}$  for all  $r_k \in \mathbb{N}$ , and

$2^{\lceil \log_2(r_k) \rceil}$  is a power of 2, for any  $r_k$ , the primitivity of  $\mathcal{G}_k$  must begin at the latest when  $n = 7(2^{\lceil \log_2(r_k) \rceil})\ell - 1 \leq 7(2^{\log_2(r_k)+1})\ell - 1 = 14(r_k)\ell - 1$ .

Finally, it remains to show that in this construction,  $m \leq 39\ell^2 r_2^2$ . We first show that  $\mathcal{G}_2$  (that is,  $r_2$  combined copies of  $\mathcal{G}_{n, A_2 \cup \dots \cup A_k}$ ) has at most  $6\ell r_2$  vertices and will have  $m \leq 39\ell^2 r_2^2$ . The bound on the number of vertices is clear—for any graph  $\mathcal{G}_{\ell, S}$  constructed by the method defined at the beginning of this section,  $\mathcal{G}_{\ell, S}$  has  $3\ell + 2|S| + 1$  vertices, and since  $|S| \leq \ell$  and  $\ell \geq 1$ , we have  $3\ell + 2|S| + 1 \leq 6\ell$ . With each graph having at most  $6\ell$  vertices, it is trivial that  $\mathcal{G}_2$  has at most  $6\ell r_2$  vertices. As discussed in Theorem 3.2.3, the number of extender sets for  $n \geq 3\ell + 3$  and  $n \pmod{\ell} \notin S$  (that is,  $m$  for  $\mathcal{G}_{\ell, S}$ ) is  $(3\ell + 2|S| + 1)^2 + |S| + 2 \leq 36\ell^2 + \ell + 2 \leq 39\ell^2$ . This proves the base case, when  $r_2 = 1$ . Suppose for an induction that after joining together  $i$  copies of  $\mathcal{G}_{\ell, A_2 \cup \dots \cup A_k}$  to make a graph  $\mathcal{G}$ , we get  $m \leq 39i^2\ell^2$ , and we then adjoin a single copy of  $\mathcal{G}_{\ell, A_2 \cup \dots \cup A_k}$  to  $\mathcal{G}$ . Then, by Theorem 3.2.5, words of sufficient length  $n$  in the new graph where  $n \pmod{\ell} \in A_1$  will have a number of extender sets equal to the number of extender sets for words of such length in  $\mathcal{G}$  (bounded above by  $39i^2\ell^2$ ) plus the number of extender sets for words of such length in  $\mathcal{G}_{\ell, A_2 \cup \dots \cup A_k}$  (bounded above by  $39\ell^2$ ) plus twice the product of the number of vertices in  $\mathcal{G}$  and  $\mathcal{G}_{\ell, A_2 \cup \dots \cup A_k}$  (bounded above by  $2(6i\ell)(6\ell) < 2i(39\ell^2)$ ).

Thus, for the resulting graph containing  $i + 1$  copies of  $\mathcal{G}_{\ell, A_2 \cup \dots \cup A_k}$ , we have:

$$m < 39i^2\ell^2 + 39\ell^2 + 2i(39\ell^2) = 39\ell^2(i^2 + 1 + 2i) = 39\ell^2(i + 1)^2,$$

giving the result for  $\mathcal{G}_2$  when  $i = r_2$ .



We next consider adding  $(r_3 - r_2)$  many copies of  $\mathcal{G}_{\ell, A_3 \cup \dots \cup A_k}$  to  $\mathcal{G}_2$  to make  $\mathcal{G}_3$ . By the same argument as above, the graph consisting of  $(r_3 - r_2)$  many copies of  $\mathcal{G}_{\ell, A_3 \cup \dots \cup A_k}$  will have at most  $6\ell(r_3 - r_2)$  vertices and  $m \leq 39\ell^2(r_3 - r_2)^2$ . Again using Theorem 3.2.5 to combine the two graphs, the resulting graph  $\mathcal{G}_3$  will have at most  $6\ell r_2 + 6\ell(r_3 - r_2) = 6\ell r_3$  vertices, and will have

$$\begin{aligned}
m &\leq 39\ell^2 r_2^2 + 39\ell^2 (r_3 - r_2)^2 + 2(6\ell r_2)(6\ell(r_3 - r_2)) \\
&< 39\ell^2 r_2^2 + 39\ell^2 (r_3 - r_2)^2 + 39\ell^2 (2(r_2)(r_3 - r_2)) \\
&= 39\ell^2 (r_2^2 + (r_3 - r_2)^2 + 2((r_2)(r_3 - r_2)) \\
&= 39\ell^2 (r_2 + (r_3 - r_2))^2 \\
&= 39\ell^2 r_3^2.
\end{aligned}$$

Continuing inductively, we can see that for  $\mathcal{G}_k$ , we will have  $m \leq 39\ell^2 r_k^2$ .  $\square$

While we may achieve any desired oscillation scheme, we cannot achieve any eventually periodic sequence we like— $m$  must be sufficiently large. For instance,  $\{1, 5, 1, 5, \dots\}$  is not achievable as the follower set sequence of any shift space. An oscillation of 4 with period 2 is achievable, but by Theorem 3.1.7, we know that if 1 occurs anywhere in the follower set sequence of  $X$ , then  $X$  is a full shift, which has follower set sequence  $\{1, 1, 1, 1, \dots\}$ .

**Theorem 3.2.6.** *Let  $X$  be an irreducible sofic shift with  $\liminf\{|F_X(n)|\} = m$  and  $\limsup\{|F_X(n)|\} = m + r$  with least eventual period  $\ell$ . Then we have that  $m > \log_2(r)$  and  $m > \frac{1}{2} \log_2(\log_2(\ell))$ . If  $\liminf\{|E_X(n)|\} = m'$  and*

$\limsup\{|E_X(n)|\} = m' + r'$  with least eventual period  $\ell'$ , then  $m' > \sqrt{\log_2(r')}$  and  $m' > \sqrt{\frac{1}{2} \log_2(\log_2(\ell'))}$ .

*Proof.* Let  $\mathcal{G}$  be an irreducible right-resolving presentation of  $X$  which contains a right-synchronizing word. Let  $|V(\mathcal{G})|$  be denoted by  $V$ . Since a follower set of a word  $w$  in a sofic shift is determined by the non-empty set of terminal vertices of paths labeled  $w$  in a presentation  $\mathcal{G}$ ,  $X$  has less than  $|\mathcal{P}(V(\mathcal{G}))| = 2^V$  follower sets, and so  $m + r < 2^V$ . Because  $\mathcal{G}$  is irreducible, right-resolving, and contains a right-synchronizing word, for large enough  $n$ , each singleton will correspond to the follower set of some word for any length greater than  $n$ . Because  $m$  occurs infinitely often in the follower set sequence,  $m$  must be greater than or equal to the number of singletons in  $\mathcal{G}$ , that is,  $m \geq V$ . Thus, we have:

$$2^m \geq 2^V > m + r > r$$

$$m > \log_2(r).$$

Moreover, as there are less than  $2^V$  follower sets, the least eventual period  $\ell$  of the follower set sequence is less than or equal to  $(2^V)!$ , as in Theorem 1.0.3. Thus  $(2^V)! \geq \ell$ . So we have:

$$(2^m)^{(2^m)} > (2^m)! \geq (2^V)! \geq \ell$$

$$(2^m)^{(2^m)} > \ell$$

$$\log_2((2^m)^{(2^m)}) > \log_2(\ell)$$

$$2^m \log_2(2^m) > \log_2(\ell)$$

$$2^m(m) > \log_2(\ell)$$

$$\log_2(2^m) + \log_2(m) > \log_2(\log_2(\ell))$$

$$m + \log_2(m) > \log_2(\log_2(\ell)).$$

Since  $m > \log_2(m)$ , we have:

$$2m > \log_2(\log_2(\ell))$$

$$m > \frac{1}{2} \log_2(\log_2(\ell)).$$

Now, an extender set of a word  $w$  in a sofic shift is determined by the non-empty set of pairs of initial and terminal vertices of paths labeled  $w$  in  $\mathcal{G}$ , so  $X$  has less than  $2^{V^2}$  extender sets, that is,  $m' + r' < 2^{V^2}$ . Since two words with the same extender set have the same follower set,  $m' \geq m \geq V$ . Thus we have:

$$2^{m'^2} \geq 2^{V^2} > m' + r' > r'$$

$$m'^2 > \log_2(r')$$

$$m' > \sqrt{\log_2(r')}.$$

Finally, as there are less than  $2^{V^2}$  extender sets, by Theorem 1.0.3,  $(2^{V^2})! \geq \ell'$ , giving:

$$(2^{m'^2})^{(2^{m'^2})} > (2^{m'^2})! \geq (2^{V^2})! \geq \ell'$$

$$(2^{m'^2})^{(2^{m'^2})} > \ell'$$

$$(2^{m'^2})(m'^2) > \log_2(\ell')$$

$$m'^2 + \log_2(m'^2) > \log_2(\log_2(\ell')).$$

Since  $m'^2 > \log_2(m'^2)$ , we have:

$$2m'^2 > \log_2(\log_2(\ell'))$$

$$m' > \sqrt{\frac{1}{2} \log_2(\log_2(\ell'))}.$$

□

**Remark 3.2.7.** For the examples discussed here, we take  $r$  or  $r' = r_k$ .

Finally, we demonstrate the existence of a non-sofic shift whose follower set sequence and extender set sequence are not monotone increasing. The construction uses Sturmian shifts, and so we give a brief definition:

**Definition 3.2.8.** Given  $0 < \alpha < 1$ , let  $T_\alpha : [0, 1) \rightarrow [0, 1)$  by  $T_\alpha(x) = x + \alpha \pmod{1}$ . Then  $T_\alpha$  is the **circle rotation by  $\alpha$** , and  $([0, 1), T_\alpha)$  is a dynamical system. If  $\alpha$  is irrational, the orbit of any point  $x$ ,  $\mathcal{O}(x) = \{x + n\alpha \pmod{1} \mid n \in \mathbb{Z}\}$ , is dense in  $[0, 1)$ . For each point  $x \in [0, 1)$  we define a symbolic coding  $\phi(x)$  of  $x$ , such that  $\phi(x) = \dots x_{-2}x_{-1}.x_0x_1x_2\dots$  where  $x_n = 1$  if  $T_\alpha^n(x) \in [0, \alpha)$  and  $x_n = 0$  if  $T_\alpha^n(x) \in [\alpha, 1)$ . Let  $Y = \overline{\{\phi(x) \mid x \in [0, 1)\}}$ . Then  $Y$  is a dynamical system when paired with the shift map  $\sigma$ . For rational  $\alpha$ ,  $Y$  contains only periodic points, so we will assume from now on that  $\alpha$  is irrational. For irrational  $\alpha$ ,  $Y$  is a **Sturmian Shift**. Sturmian shifts are non-sofic and contain no periodic points, and furthermore,  $\Phi_Y(n) = n + 1$  for all  $n$  (See [2]).

Given that Sturmian shifts are non-sofic, Theorem 3.1.1 shows that  $|E_Y(n)| \geq n + 1$ . Thus we can deduce from the fact that  $\Phi_Y(n) = n + 1$  that  $|E_Y(n)| =$

$n + 1$ . The following Lemma proves the result for both extender and follower sets:

**Lemma 3.2.9.** *If  $Y$  is a Sturmian shift, then for any  $n \in \mathbb{N}$ ,  $|F_Y(n)| = |E_Y(n)| = n + 1$ .*

*Proof.* For a fixed length  $n$ , Sturmian shifts have exactly  $n + 1$  words in  $L_n(Y)$ , so it is sufficient to show that any two words of length  $n$  in  $Y$  have distinct follower and extender sets. Sturmian shifts are symbolic codings of irrational circle rotations, say by  $\alpha \notin \mathbb{Q}$ . We may take  $\alpha < \frac{1}{2}$  by simply switching the labels 0 and 1 whenever  $\alpha > \frac{1}{2}$ . Furthermore, the cylinder sets of words of length  $n$  correspond to a partition of the circle into  $n + 1$  subintervals, so for two words of length  $n$  in  $Y$ , each corresponds to a subinterval of the circle, and the two subintervals are disjoint. Let  $w$  and  $v$  be two distinct words in  $L_n(Y)$  corresponding to disjoint intervals  $I_w$  and  $I_v$ ,  $[0, \alpha)$  be the interval coded with 1, and  $T_\alpha$  be the rotation by  $\alpha$ . We claim that there exists an  $N \in \mathbb{N}$  such that  $T_\alpha^{-N}[0, \alpha)$  intersects one of  $I_w$  and  $I_v$  but not the other: Since  $\alpha < \frac{1}{2}$ , and since  $\{n\alpha \mid n \in \mathbb{N}\}$  is dense in the circle, if one of  $I_w$  and  $I_v$  has length at least  $\frac{1}{2}$ , there exists  $N \in \mathbb{N}$  such that  $T_\alpha^{-N}[0, \alpha)$  is contained entirely inside that large interval, and thus completely disjoint from the other. Otherwise, take  $I_w^c$ , which clearly has length at least  $\frac{1}{2}$ , and find an  $N \in \mathbb{N}$  such that  $T_\alpha^{-N}[0, \alpha)$  is contained inside  $I_w^c$  and intersects  $I_v \subseteq I_w^c$ , again possible due to denseness of  $\{n\alpha \mid n \in \mathbb{N}\}$ . Hence we have proved our claim, that  $\exists N \in \mathbb{N}$  such that  $T_\alpha^{-N}[0, \alpha)$  intersects one of  $I_w$  and  $I_v$  but not the other, and therefore, that the symbol 1 may follow one of the words  $w$  and  $v$  exactly  $N$  units later, but not the other. Therefore  $w$  and  $v$  have distinct follower sets, and thus, distinct extender sets, completing the proof.  $\square$

We use Lemma 3.2.9 and the sofic shifts constructed earlier in this section to build a non-sofic shift with follower and extender set sequences which occasionally decrease:

**Theorem 3.2.10.** *There exists an irreducible non-sofic shift  $X$  such that  $\{|F_X(n)|\}$  and  $\{|E_X(n)|\}$  are not monotone increasing.*

*Proof.* Take a sofic shift  $X = X_{\mathcal{G}_{\ell,S}}$  for any  $\ell, S$  as defined earlier in this section. Take the direct product of  $X$  and a Sturmian shift  $Y$ . Two words in  $X \times Y$  have the same extender set if and only if the projection of those words to both their first and second coordinates have the same extender set in  $X$  and  $Y$ , respectively. That is, if two words  $w$  and  $v$  have different extender sets in  $X$ , then any two words whose projections to their first coordinate are  $w$  and  $v$  will have different extender sets in  $X \times Y$ , and similarly for words  $w'$  and  $v'$  with different extender sets in  $Y$ . Therefore  $|E_{X \times Y}(n)| = |E_X(n)| \cdot |E_Y(n)|$ . By similar logic,  $|F_{X \times Y}(n)| = |F_X(n)| \cdot |F_Y(n)|$ .

Thus, if we let  $m = \liminf_{n \in \mathbb{N}} \{|E_X(n)|\}$ , then for any  $n \geq 3\ell + 3$  with  $n \pmod{\ell} \notin S$ , we have  $|E_{X \times Y}(n)| = m \cdot (n + 1)$ , and if  $n \pmod{\ell} \in S$ , then  $|E_{X \times Y}(n)| = (m + 1)(n + 1)$ . As  $m$  is fixed and  $n$  approaches infinity, it is clear that  $\{|E_{X \times Y}(n)|\}$  is unbounded, and thus the shift  $X \times Y$  is nonsocic. Furthermore, as the direct product of a mixing shift ( $X$  is primitive by Lemma 3.2.2, and therefore mixing) with an irreducible shift,  $X \times Y$  is irreducible.

Choose  $n$  large enough that  $n > m - 1$ , and such that  $n \pmod{\ell} \in S$  and  $n + 1 \pmod{\ell} \notin S$ . Then

$$\begin{aligned}
|E_{X \times Y}(n)| &= (m + 1)(n + 1) \\
&= mn + m + n + 1 \\
&> mn + m + (m - 1) + 1 \\
&= mn + 2m \\
&= m(n + 2) \\
&= |E_{X \times Y}(n + 1)|.
\end{aligned}$$

Therefore the extender set sequence of  $X \times Y$  is not monotone increasing. A similar argument shows that the follower set sequence of  $X \times Y$  is not monotone increasing as well.  $\square$

**Example 3.2.11.** *Let  $X = X_{\mathcal{G}_5, \{0, 3\}}$  as in Figure 3.4. Then*

$m = \liminf_{n \in \mathbb{N}} \{|E_X(n)|\} = (3\ell + 2|S| + 1)^2 + |S| + 2 = 404$ , so for  $n = 405$  (since  $405 > 3\ell + 3$ ,  $405 > m - 1$ ,  $405 \pmod{5} \in \{0, 3\}$ , and  $406 \pmod{5} \notin \{0\}$ ),  $|E_{X \times Y}(n)| > |E_{X \times Y}(n + 1)|$ . In particular,  $|E_{X \times Y}(405)| = (405)(406) = 164,430$  while  $|E_{X \times Y}(406)| = (404)(407) = 164,428$ .

**Remark 3.2.12.** *The reader may observe that once  $n$  is sufficiently large for the follower or extender set sequence of  $X \times Y$  to decrease, these decreases will happen for exactly the same lengths  $n$  as the decreases in the follower or extender set sequence of  $X = X_{\mathcal{G}_{\ell, S}}$ . Thus there are infinitely many lengths for which the follower or extender set sequence of  $X \times Y$  decreases.*

### 3.3 Follower Sets, Extender Sets, and $\beta$ -shifts

We have now established that not all follower, predecessor, and extender set sequences may be achieved as complexity sequences, for follower, predecessor, and extender set sequences may fail to be monotone increasing. We now ask the opposite question: may all complexity sequences be achieved as follower, predecessor, or extender set sequences? We use the classical  $\beta$ -shifts to answer this question. In the process, we will also show that  $\beta$ -shifts provide a valuable class of examples on a related topic: the realizable differences in limiting behavior among follower, predecessor, and extender set sequences of non-sofic shifts. It is easy to show that the follower and predecessor set sequences of a sofic shift (while certainly both bounded) may approach different limits.

**Example 3.3.1.** Consider  $X_{\mathcal{G}}$ , where  $\mathcal{G}$  is the graph shown in Figure 3.6.

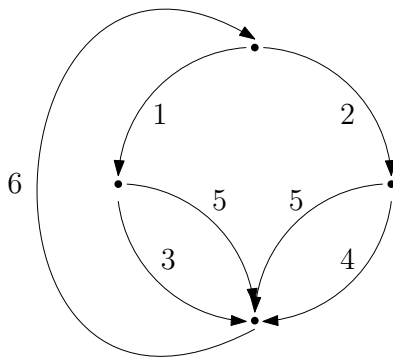


Figure 3.6:  $\mathcal{G}$  such that  $X_{\mathcal{G}}$  has a different number of predecessor and follower sets

Observe that every word in  $L(X_{\mathcal{G}})$  is right-synchronizing. Hence, for any word  $w$  in  $L(X_{\mathcal{G}})$ , all paths labeled  $w$  end at the same single vertex in  $\mathcal{G}$ . The follower set corresponding to the uppermost vertex is  $F(6)$  and consists of all legal right-infinite sequences beginning with 1 or 2. The follower set of the leftmost



vertex is  $F(1)$  and consists of all legal right-infinite sequences beginning with 3 or 5. The follower set corresponding to the rightmost vertex is  $F(2)$  and consists of all legal right-infinite sequences beginning with 4 or 5. Finally, the follower set corresponding to the bottom vertex is  $F(3) = F(4) = F(5)$ , and consists of all legal right-infinite sequences beginning with 6. Thus, these four follower sets are distinct, and  $\mathcal{G}$  is follower-separated. Moreover, the follower set of any word  $w$  depends only on the last letter of  $w$ , and so there are only four follower sets in  $X_{\mathcal{G}}$ , one corresponding to each vertex of the graph, and all four correspond to words of every possible length  $n$ . Hence, the follower set sequence of  $X_{\mathcal{G}}$  is  $\{|F_{X_{\mathcal{G}}}(n)|\} = \{4, 4, 4, 4, \dots\}$ . On the other hand, there are five predecessor sets in  $X_{\mathcal{G}}$ —the predecessor sets corresponding to each vertex, as well as the predecessor set of the word 5, which corresponds to 2 vertices in  $\mathcal{G}$ . Since the predecessor set of 5 contains sequences ending with both 1 and 2, it is clearly not equal to any of the other four predecessor sets. Thus there are 5 total predecessor sets in  $X_{\mathcal{G}}$ , and again all five correspond to words of every possible length  $n$ . Hence, the predecessor set sequence of  $X_{\mathcal{G}}$  is  $\{|P_{X_{\mathcal{G}}}(n)|\} = \{5, 5, 5, 5, \dots\}$ .

**Remark 3.3.2.** While the graph  $\mathcal{G}$  in Figure 3.6 does not satisfy the hypotheses of Theorem 3.2.5, (in particular,  $\mathcal{G}$  is not left-resolving and does not contain a self-loop), the same construction applied to  $k$  copies of  $\mathcal{G}$  will nonetheless yield a graph presenting a shift whose total number of predecessor sets is  $k$  greater than its total number of follower sets, where  $I_1, I_2$ , and so on, may be chosen to be any of the vertices of the graph  $\mathcal{G}$ . We omit the details of the proof as it is similar to the proof of Theorem 3.2.5.

For the non-sofic case, since the follower, predecessor, and extender set sequences are unbounded, we may see far more dramatic differences in limiting behavior of those sequences.

We begin our exploration of the follower, predecessor, and extender set sequences of  $\beta$ -shifts with a brief definition of the  $\beta$ -shift. The  $\beta$ -shift is traditionally defined as a one-sided shift space (so the shift map  $\sigma$  now removes the first letter, rather than shifting it left of the origin, that is,  $\sigma(.x_0x_1x_2\dots) = .x_1x_2x_3\dots$ ), but we will end the section with a discussion about using the natural extension to translate our results to the two-sided setting.

**Definition 3.3.3.** *Given  $\beta > 1$ , let  $d_\beta : [0, 1) \rightarrow \{\lfloor \beta \rfloor + 1\}^{\mathbb{N}}$  be the map which sends each point  $x \in [0, 1)$  to its expansion in base  $\beta$ . That is, if  $x = \sum_{n=1}^{\infty} \frac{x_n}{\beta^n}$ , then  $d_\beta(x) = .x_1x_2x_3\dots$  (In the case where  $x$  has more than one  $\beta$  expansion, we take the lexicographically largest expansion.) The closure of the image,  $\overline{d_\beta([0, 1))}$ , is a one-sided symbolic dynamical system called the  $\beta$ -**shift**, denoted  $X_\beta$ . (Introduced in [12]). Then if  $T_\beta : [0, 1) \rightarrow [0, 1)$  is given by  $T_\beta(x) = \beta x \pmod{1}$ , then the  $\beta$ -shift  $X_\beta$  is a symbolic coding of  $T_\beta$ : for any  $i \in \mathbb{N}$ ,  $x \in [0, 1)$ ,  $\sigma^i(d_\beta(x)) = k$  if and only if  $T_\beta^i(x) \in [\frac{k}{\beta}, \frac{k+1}{\beta})$ . Therefore the  $\beta$ -shift must have alphabet  $\{0, 1, \dots, \lfloor \beta \rfloor\}$ .*

An equivalent characterization of the  $\beta$ -shift is given by a right-infinite sequence  $d_\beta^*(1) = \lim_{x \nearrow 1} d_\beta(x)$ . For any sequence  $x$  on the alphabet  $\{0, \dots, \lfloor \beta \rfloor\}$ ,  $x \in X_\beta$  if and only if every shift of  $x$  is lexicographically less than or equal to  $d_\beta^*(1)$  (see [10]). Then clearly, the sequence  $d_\beta^*(1)$  has the property that every shift of  $d_\beta^*(1)$  is lexicographically less than or equal to  $d_\beta^*(1)$ . The sequence  $d_\beta(1)$  only terminates with an infinite string of 0's in the case that  $X_\beta$  is a shift

of finite type, however, the sequence  $d_\beta^*(1)$  never terminates with an infinite string of 0's. (Moreover,  $d_\beta^*(1) \neq d_\beta(1)$  if and only if  $X_\beta$  is a shift of finite type ([1])).

**Example 3.3.4.** *The golden mean shift  $X_\varphi$  of Example 2.0.16, considered as a one-sided shift, is the  $\beta$ -shift corresponding to  $\beta = \varphi = \frac{1+\sqrt{5}}{2}$ . Since the golden mean shift is a shift of finite type, we should have that  $d_\varphi(1)$  terminates in an infinite string of 0's, and  $d_\varphi^*(1) \neq d_\varphi(1)$ . Indeed,  $d_\varphi(1) = .11000000\dots$ , and  $d_\varphi^*(1) = .10101010\dots$  The reader may check that the requirement that every shift of a right infinite sequence be lexicographically less than or equal to  $d_\varphi^*(1)$  is equivalent to the requirement that the right-infinite sequence never see the word 11.*

In fact, any right-infinite sequence  $d$  satisfying the two properties we have discussed must be equal to  $d_\beta^*(1)$  for some  $\beta$ :

**Lemma 3.3.5.** *Let  $d$  be a one-sided right-infinite sequence on the alphabet  $\mathcal{A} = \{0, 1, \dots, k\}$  (with  $k$  occurring in  $d$ ) such that every shift of  $d$  is lexicographically less than or equal to  $d$  and  $d$  does not end with an infinite string of zeros. Then there exists  $k < \beta \leq k + 1$  such that  $d = d_\beta^*(1)$ .*

*Proof.* Suppose  $d = .d_0d_1d_2\dots$  be a sequence on the alphabet  $\mathcal{A} = \{0, 1, \dots, k\}$  with the symbol  $k$  occurring in  $d$ . Let  $1 = \sum_{i=0}^{\infty} \frac{d_i}{\beta^{i+1}}$ . Because not all  $d_i$  are equal to zero, the equation has some solution  $\beta$  by the Intermediate Value Theorem. Because  $d$  does not end in an infinite string of 0's, it cannot be the case that  $d$  is an expansion which is equal to  $d_\beta(1) \neq d_\beta^*(1)$  as in the case where  $X_\beta$  is a shift of finite type. Furthermore, the requirement that every shift of  $d$  is lexicographically less than or equal to  $d$  necessitates that  $d_0 = k$ ,

for if  $d_0 < k$  but  $d_i = k$  for some  $i \neq 0$ , then  $\sigma^i(d) \succ d$ , a contradiction. Then, multiplying both sides of the equation by  $\beta$ , we get

$$\beta = \sum_{i=0}^{\infty} \frac{d_i}{\beta^i} = d_0 + \sum_{i=1}^{\infty} \frac{d_i}{\beta^i} = k + \sum_{i=1}^{\infty} \frac{d_i}{\beta^i} > k$$

with the final inequality being strict because not all of  $\{d_i\}_{i=1}^{\infty}$  may be equal to 0. On the other hand,

$$\beta = \sum_{i=0}^{\infty} \frac{d_i}{\beta^i} \leq \sum_{i=0}^{\infty} \frac{k}{\beta^i} = k \sum_{i=0}^{\infty} \left(\frac{1}{\beta}\right)^i = k \left(\frac{1}{1 - \frac{1}{\beta}}\right) = k \left(\frac{\beta}{\beta - 1}\right)$$

as  $\beta > 1$ . But then

$$\begin{aligned} \beta &\leq k \left(\frac{\beta}{\beta - 1}\right) \\ \beta(\beta - 1) &\leq k\beta \\ \beta - 1 &\leq k \\ \beta &\leq k + 1. \end{aligned}$$

□

Moreover,  $X_\beta$  is sofic if and only if  $d_\beta^*(1)$  is eventually periodic (see [1]). We use these facts to characterize the follower set sequences of all one-sided  $\beta$ -shifts:

**Lemma 3.3.6.** *If  $w, v \in L(X_\beta)$  and  $w$  does not terminate with any prefix of the sequence  $d_\beta^*(1)$ , then  $wv \in L(X_\beta)$ . Thus,  $F_{X_\beta}(w) = X_\beta = F(\emptyset)$ .*

*Proof.* If  $v \in L(X_\beta)$ , then  $v0^\infty \in X_\beta$  trivially. Then we claim  $wv0^\infty \in X_\beta$  as well. Given  $\sigma^i(wv0^\infty)$ , if  $i \geq |w|$ ,  $\sigma^i(wv0^\infty) \preceq d_\beta^*(1)$  as  $v0^\infty$  is legal. If  $i < |w|$ ,

then the first  $|w| - i$  letters of  $\sigma^i(wv0^\infty)$  are not a prefix of  $d_\beta^*(1)$ , but since they are contained in the legal word  $w$ , must be lexicographically less than the first  $|w| - i$  letters of  $d_\beta^*(1)$ , and thus,  $\sigma^i(wv0^\infty) \preceq d_\beta^*(1)$ . Thus,  $wv0^\infty \in X_\beta$  and  $wv \in L(X_\beta)$ .  $\square$

We now prove the useful fact that  $\beta$ -shifts satisfy Conjecture 1.

**Theorem 3.3.7.** *For any  $\beta$ -shift,  $|F_{X_\beta}(n)| \leq n$  for any  $n$  if and only if  $X_\beta$  is sofic.*

*Proof.* Since  $\beta > 1$ , we have  $\lfloor \beta \rfloor \geq 1$ . Since  $d_\beta^*(1)$  is a sequence on  $\{0, 1, \dots, \lfloor \beta \rfloor\}$  satisfying the conditions of Lemma 3.3.5,  $d_\beta^*(1)$  must begin with  $\lfloor \beta \rfloor$  by an identical argument as presented in the proof of Lemma 3.3.5.  $\lfloor \beta \rfloor$  is lexicographically larger than 0, and so by Lemma 3.3.6,  $F(0) = F(\emptyset)$  for any  $\beta$ -shift. Then every  $\beta$ -shift  $X_\beta$  features a word  $w$  with  $F(w) = F(\emptyset)$ , and thus by Lemma 3.1.9,  $|F_{X_\beta}(n)| \leq n$  for some  $n$  if and only if  $X_\beta$  is sofic.  $\square$

**Theorem 3.3.8.** *For any  $\beta$ -shift  $X_\beta$ , and for any  $n \in \mathbb{N}$ , we have  $|F_{X_\beta}(n)| \leq n + 1$ .*

*Proof.* Fix  $n \in \mathbb{N}$  and partition the words  $w \in L_n(X_\beta)$  into  $n + 1$  classes  $\{S_0, S_1, \dots, S_n\}$ , where the index represents the length of the maximal prefix of  $d_\beta^*(1)$  appearing as a suffix of the word  $w$ . So the class  $S_0$  will contain words which do not contain any prefix of  $d_\beta^*(1)$  as a suffix, while the words in the class  $S_1$  are words ending with the first letter of  $d_\beta^*(1)$ , but containing no larger prefix of  $d_\beta^*(1)$  as a suffix, and so on, up to  $S_n$ , which consists of the single word formed by the first  $n$  letters of  $d_\beta^*(1)$ . We claim that for two words  $w$  and  $v \in L_n(X_\beta)$  such that  $w$  and  $v$  are in the same class,  $F(w) = F(v)$ . This will imply the result.

By Lemma 3.3.6, any word  $w \in S_0$  will have  $F(w) = F(\emptyset)$ , and so the claim is certainly true for  $S_0$ .  $S_n$  contains only one word, so the claim holds trivially for  $S_n$  as well. So let  $0 < k < n$  and consider two words  $v, w \in S_k$ . Then the last  $k$  letters of  $v$  and  $w$  are identical, and equal to the first  $k$  letters of  $d_\beta^*(1)$ , and so  $w$  and  $v$  may only differ on their first  $n - k$  letters. Let  $s$  be some right-infinite sequence in  $F(w)$ , so  $ws$  is a point of  $X_\beta$ . Then every shift of  $ws$  is lexicographically less than or equal to  $d_\beta^*(1)$ . Then, certainly, every shift of  $vs$  beyond the first  $n - k$  shifts is lexicographically less than or equal to  $d_\beta^*(1)$ , because after the first  $n - k$  letters,  $vs$  and  $ws$  are equal. Furthermore, for  $0 \leq i < n - k$ ,  $\sigma^i(vs)$  does not begin with a prefix of  $d_\beta^*(1)$  by definition of the class  $S_k$ , and so  $\sigma^i(v)$  is already strictly lexicographically less than the first  $n - i$  letters of  $d_\beta^*(1)$  because  $v$  is a legal word. Then, any sequence  $s$  may follow  $\sigma^i(v)$  and  $\sigma^i(vs)$  will be less lexicographically than  $d_\beta^*(1)$ . So in fact, the first  $n - k$  letters place no restrictions whatsoever on the sequences which may follow  $v$ , and indeed,  $s \in F(v)$ . So  $F(w) \subseteq F(v)$  and similarly,  $F(v) \subseteq F(w)$ , so  $F(v) = F(w)$ . Hence, for any  $\beta$ -shift  $X_\beta$ ,  $|F_{X_\beta}(n)| \leq n + 1$  for all  $n \in \mathbb{N}$ .  $\square$

**Corollary 1.** *For any non-sofic  $\beta$ -shift  $X_\beta$ ,  $\{|F_{X_\beta}(n)|\}_{n \in \mathbb{N}} = \{n + 1\}_{n \in \mathbb{N}}$ .*

*Proof.* By Thm. 3.3.8,  $|F_{X_\beta}(n)| \leq n + 1$  for all  $n \in \mathbb{N}$ . If  $X_\beta$  is non-sofic, by Theorem 3.3.7,  $|F_{X_\beta}(n)| \geq n + 1$  for all  $n \in \mathbb{N}$ . Thus, for all  $n \in \mathbb{N}$  we have  $|F_{X_\beta}(n)| = n + 1$ , that is, the follower set sequence  $\{|F_{X_\beta}(n)|\}_{n \in \mathbb{N}}$  of  $X_\beta$  is equal to  $\{n + 1\}_{n \in \mathbb{N}}$ .  $\square$

**Remark 3.3.9.** *We remind the reader that the shift  $X_\beta$  contains an intrinsically synchronizing word if and only if there exists some word  $w \in L(X_\beta)$  such*

that  $w$  is not a subword of the sequence  $d_\beta^*(1)$  (see [1]), and so this provides examples of non-sofic shifts which contain an intrinsically synchronizing word and still have the property that  $|F_X(n)| = n + 1$  for all  $n \in \mathbb{N}$ . Non-sofic  $\beta$ -shifts are furthermore remarkable because, unlike Sturmian shifts and other examples in which  $|F_X(n)| = n + 1$  for all  $n$ , non-sofic  $\beta$ -shifts have positive topological entropy. ( $\beta > 1$  and the topological entropy of a  $\beta$ -shift is  $\log(\beta)$ ).

Though this is sufficient to characterize the follower set sequences of non-sofic beta-shifts, we could also prove that non-sofic beta-shifts have  $n + 1$  follower sets of words of length  $n$  for every  $n$  using the following lemma, which will be useful later:

**Lemma 3.3.10.** *Let  $X_\beta$  be a  $\beta$ -shift. Let  $w \in L_n(X_\beta)$  such that  $w$  contains multiple different prefixes of  $d_\beta^*(1)$  as suffixes (all necessarily nested subwords). Say the longest prefix of  $d_\beta^*(1)$  contained as a suffix of  $w$  has length  $j$ . Then  $\sigma^j(d_\beta^*(1)) \preceq \sigma^k(d_\beta^*(1))$  for every  $k$  such that  $w$  has a prefix of  $d_\beta^*(1)$  of length  $k$  as a suffix, and therefore the follower set of  $w$  consists of all sequences which are lexicographically less than or equal to  $\sigma^j(d_\beta^*(1))$ .*

*Proof.* Suppose  $k < j$  and  $w$  contains a prefix of  $d_\beta^*(1)$  of length  $j$  and a prefix of  $d_\beta^*(1)$  of length  $k$  as suffixes. We will denote the prefix of length  $j$  by  $(d_\beta^*(1))_j$ , and similarly for  $k$ . Note that the first  $k$  and the last  $k$  letters of  $(d_\beta^*(1))_j$  must be equal to  $(d_\beta^*(1))_k$ . So  $d_\beta^*(1)$  might appear as in Figure 3.7. (It is also possible that no letters exist in the middle portion between the two occurrences of  $(d_\beta^*(1))_k$ , or even that they overlap; the proof still holds in such cases).

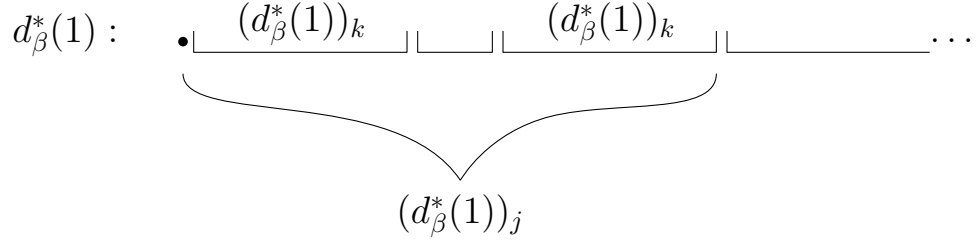


Figure 3.7:  $d_\beta^*(1)$  in the case of Lemma 3.3.10

If  $\sigma^j(d_\beta^*(1)) \succ \sigma^k(d_\beta^*(1))$ , then it is easy to see that  $\sigma^{j-k}(d_\beta^*(1)) \succ d_\beta^*(1)$ , a contradiction, as every shift of  $d_\beta^*(1)$  must be lexicographically less than or equal to  $d_\beta^*(1)$ . Therefore  $\sigma^j(d_\beta^*(1)) \preceq \sigma^k(d_\beta^*(1))$ , establishing the result. Since the right-infinite sequences which may legally follow  $w$  consist of the sequences  $s$  st. every shift of  $ws$  is lexicographically less than or equal to  $d_\beta^*(1)$ , we have  $F(w) = \{s \mid s \preceq \sigma^k(d_\beta^*(1)) \forall k \text{ s.t. } w \text{ has a } k\text{-letter prefix of } d_\beta^*(1) \text{ as a suffix}\}$ . Since  $\sigma^j(d_\beta^*(1))$  is the smallest such shift, the follower set of  $w$  consists exactly of those sequences lexicographically less than or equal to  $\sigma^j(d_\beta^*(1))$ .  $\square$

**Corollary 2.** *Let  $X_\beta$  be a  $\beta$ -shift,  $n \in \mathbb{N}$ , and the classes  $S_0, S_1, \dots, S_n$  be as defined above. Then the common follower set of words in class  $S_j$  is equal to the follower set of words in class  $S_k$  if and only if  $\sigma^j(d_\beta^*(1)) = \sigma^k(d_\beta^*(1))$ .*

*Proof.* Lemma 3.3.10 shows that we may distinguish between follower sets of words by only looking at  $\sigma^j(d_\beta^*(1))$ , where  $j$  is the length of the longest prefix of  $d_\beta^*(1)$  appearing as a suffix of the word. The corollary follows immediately from that observation.  $\square$

It is easy to characterize the follower sets of a non-sofic  $\beta$ -shift from this corollary: since  $d_\beta^*(1)$  is not eventually periodic,  $j \neq k$  implies  $\sigma^j(d_\beta^*(1)) \neq \sigma^k(d_\beta^*(1))$ , and therefore the follower sets in each class  $S_0, S_1, \dots, S_n$  must all be



distinct. Thus, for each  $n$ ,  $|F_{X_\beta}(n)| = n + 1$  (confirming a result we already established in Corollary 1). On the other hand, this corollary will also aid in characterizing the follower set sequences of sofic  $\beta$ -shifts:

**Theorem 3.3.11.** *For any sofic  $\beta$ -shift, let  $p = \min\{j \in \mathbb{N} \mid \exists k < j, \sigma^k(d_\beta^*(1)) = \sigma^j(d_\beta^*(1))\}$ . Then for any  $n \in \mathbb{N}$ , we have:  $|F_{X_\beta}(n)| = \begin{cases} n + 1 & n < p \\ p & n \geq p \end{cases}$*

*Proof.* If  $X_\beta$  is sofic, the sequence  $d_\beta^*(1)$  is eventually periodic. Let  $p = \min\{j \in \mathbb{N} \mid \exists k < j, \sigma^k(d_\beta^*(1)) = \sigma^j(d_\beta^*(1))\}$ . Then for every length  $n < p$ , the shifts  $\sigma^i(d_\beta^*(1))$ ,  $0 \leq i \leq n$ , are all distinct, and so  $|F_{X_\beta}(n)| = n + 1$  by Cor. 2. However, at length  $p$ , the word in  $S_p$  will have the same follower set as words in length  $k$  for some  $k < j$  by Cor. 2, and so there will only be  $p$  (rather than  $p + 1$ ) follower sets of words of length  $p$ . But if  $\sigma^p(d_\beta^*(1)) = \sigma^k(d_\beta^*(1))$ , then  $\sigma^{p+\ell}(d_\beta^*(1)) = \sigma^{k+\ell}(d_\beta^*(1))$  for all  $\ell \in \mathbb{N}$ , and so after length  $p$ , no lengths will contribute any new follower sets. Indeed,  $|F_{X_\beta}(n)| = p$  for all  $n \geq p$ .  $\square$

We next explore the predecessor set sequences of  $\beta$ -shifts. Because for now we are working with one-sided shifts, rather than considering the predecessor set of  $w$  to be the set of all left-infinite sequences which may precede  $w$ , we will instead consider the predecessor set of  $w$  to be the set of all finite words which may precede  $w$ . For now we have no choice but to use this definition; when we later apply these results to two-sided  $\beta$ -shifts, the reader may recall that the results will hold no matter which definition of predecessor set is used, as asserted in Definition 2.0.15.

**Theorem 3.3.12.** *For any  $\beta$ -shift  $X_\beta$ , the predecessor set sequence  $\{|P_{X_\beta}(n)|\}$  of  $X_\beta$  is equal to the complexity sequence  $\{\Phi_{d_\beta^*(1)}(n)\}$  of  $d_\beta^*(1)$ .*

*Proof.* First, partition all finite words into classes  $S_0, S_1, \dots$  as before, indexed by the maximal prefix of  $d_\beta^*(1)$  appearing as a suffix of the word. So  $S_0$  is the set of words containing no prefix of  $d_\beta^*(1)$  as a suffix,  $S_1$  is the set of words terminating with the first letter of  $d_\beta^*(1)$  but containing no larger prefix as a suffix, and so on. Note that, as we are considering all finite words, not just finite words of a fixed length, there will be infinitely many classes, each containing words of many different lengths.

Let  $k, n \in \mathbb{N}$  and  $w \in L_n(X_\beta)$ . Then all of the words in  $S_k$  are in the predecessor set of  $w$  if and only if  $w \preceq (\sigma^k(d_\beta^*(1)))_n$  (we are implicitly using Lemma 3.3.10 to rule out the possibility that a prefix of  $d_\beta^*(1)$  of length less than  $k$  could introduce some illegal word). For the rest of this proof, when a word  $w$  satisfies the condition  $w \preceq (\sigma^k(d_\beta^*(1)))_n$ , we will simply say that  $w$  *satisfies condition  $k$* . Since  $k$  may range from 0 to infinity, while  $n$  is fixed, every  $n$ -letter word in  $d_\beta^*(1)$  will appear as the upper bound in condition  $k$  for some  $k$ . Moreover, these predecessor sets are nested—if  $S_k$  is part of the predecessor set of a word  $w$ , then it is also part of the predecessor set of every lexicographically smaller word of the same length. So, given a word  $w$ , the predecessor set of  $w$  depends only on how many of  $\Phi_{d_\beta^*(1)}(n)$  different conditions on  $w$  are met. But since these conditions are nested, there are only  $\Phi_{d_\beta^*(1)}(n) + 1$  (nested) subsets of these conditions which may be simultaneously met. Moreover,  $w$  must satisfy at least *one* of them, because for any legal word  $w \in L_n(X_\beta)$ ,  $w \preceq (d_\beta^*(1))_n$ , so  $w$  satisfies condition 0. If this is the only one of the conditions satisfied, then the predecessor set of  $w$  only consists of words in  $S_0$  and  $S_i$  for any  $i$  such that  $(\sigma^i(d_\beta^*(1)))_n = (d_\beta^*(1))_n$  (that is, that condition  $i$  and condition 0 are the same condition). If  $w$  satisfies this condition and only

one other, the predecessor set of  $w$  contains  $S_0$  and any such  $S_i$  as before, along with any  $S_k$  such that  $(\sigma^k(d_\beta^*(1)))_n$  is equal to the lexicographically greatest  $n$ -letter word in  $d_\beta^*(1)$  besides  $(d_\beta^*(1))_n$ , and so on. In other words, words in  $S_k$  are part of the predecessor set of  $w$  when  $w$  satisfies condition  $k$ , but there are only  $\Phi_{d_\beta^*(1)}(n)$ -many subsets of  $\mathbb{N}$  which are realizable as the set of all  $k$  for which  $w$  satisfies condition  $k$ . Thus, for all  $n$ , there are  $\Phi_{d_\beta^*(1)}(n)$  predecessor sets of length  $n$  in  $X_\beta$ . So the predecessor set sequence of any  $\beta$ -shift is  $\{\Phi_{d_\beta^*(1)}(n)\}_{n \in \mathbb{N}}$ .  $\square$

This is enough to see that the predecessor and follower set sequences of  $\beta$ -shifts may exhibit vastly different limiting behavior:

**Example 3.3.13.** *Let  $d$  be any right-infinite sequence on  $\{0, 1\}$  which contains every word in  $L(X_{[2]})$  as a subword, that is, that  $\mathcal{O}(d) = \{\sigma^i(d) \mid i \in \mathbb{N}\}$  is dense in  $\{0, 1\}^{\mathbb{N}}$ . Let  $\tilde{d} = .2d$ , so the first digit of  $\tilde{d}$  is 2 and  $\sigma(\tilde{d}) = d$ . Then  $\tilde{d}$  satisfies the hypothesis of Lemma 3.3.5 and so there exists some  $2 < \beta \leq 3$  such that  $\tilde{d} = d_\beta^*(1)$ . Since  $\tilde{d}$  is certainly not eventually periodic,  $X_\beta$  is non-sofic, and so the follower set sequence of  $X_\beta$  is  $\{|F_{X_\beta}(n)|\} = \{n+1\}$  by Corollary 1. On the other hand, by Theorem 3.3.12, the predecessor set sequence of  $X_\beta$  is  $\{|P_{X_\beta}(n)|\} = \{\Phi_{\tilde{d}}(n)\} = \{2^n+1\}$ . This shows that the predecessor set sequence may grow exponentially in  $n$  even when the follower set sequence grows only linearly in  $n$ .*

We may also use Theorem 3.3.12 to address our primary question: what complexity sequences may be realized as follower, predecessor, or extender set sequences of shift spaces? This is mostly answered by Theorems 1.0.6 and 1.0.7.

*Proof of Theorem 1.0.6.* If  $d$  is such that for all  $i \in \mathbb{N}$ ,  $\sigma^i(d) \preceq d$ , and  $d$  does not end in an infinite string of 0's, then by Lemma 3.3.5 there exists  $\beta$  such that  $d = d_\beta^*(1)$ , and by Theorem 3.3.12,  $\{|P_{X_\beta}(n)|\} = \{\Phi_d(n)\}$ . If  $d$  is such that for all  $i \in \mathbb{N}$ ,  $\sigma^i(d) \preceq d$ , but  $d$  does end in an infinite string of 0's, define a new sequence  $\hat{d}$  such that each digit of  $\hat{d}$  is one greater than the corresponding digit of  $d$ . Then  $\hat{d}$  has the same complexity sequence as  $d$ , and also satisfies the property that for all  $i \in \mathbb{N}$ ,  $\sigma^i(\hat{d}) \preceq \hat{d}$ , and so by Lemma 3.3.5 there exists  $\beta$  such that  $\hat{d} = d_\beta^*(1)$ , and by Theorem 3.3.12,  $\{|P_{X_\beta}(n)|\} = \{\Phi_{\hat{d}}(n)\} = \{\Phi_d(n)\}$ .  $\square$

Even if the sequence  $d$  does not satisfy the property that for all  $i \in \mathbb{N}$ ,  $\sigma^i(d) \preceq d$ , we may achieve a predecessor set sequence very close to the complexity sequence of  $d$ :

*Proof of Theorem 1.0.7.* Suppose  $d$  does not satisfy the property that for all  $i \in \mathbb{N}$ ,  $\sigma^i(d) \preceq d$ . (We may assume that we have already added 1 to each digit, if necessary, so that  $d$  does not end in a string of 0's). Then create a new sequence  $\tilde{d}$  so that  $\sigma(\tilde{d}) = d$ , but the first letter of  $\tilde{d}$  is lexicographically greater than any symbol in  $d$ . Then  $\tilde{d}$  satisfies the hypotheses of Lemma 3.3.5 and so there exists  $\beta$  such that  $\tilde{d} = d_\beta^*(1)$ , and by Theorem 3.3.12,  $\{|P_{X_\beta}(n)|\} = \{\Phi_{\tilde{d}}(n)\}$ . For each length  $n$ ,  $\tilde{d}$  will have one more  $n$ -letter word than the number occurring in  $d$ . Specifically, the first  $n$  letters of  $\tilde{d}$  cannot occur in  $d$  because the first symbol of  $\tilde{d}$  is not in the alphabet of  $d$ . Moreover, any word occurring in  $\tilde{d}$  which does not see the first letter of  $\tilde{d}$  is necessarily a subword of  $d$ , as  $\sigma(\tilde{d}) = d$ . Therefore, for any  $n \in \mathbb{N}$ ,  $\{\Phi_{\tilde{d}}(n)\} = \{\Phi_d(n) + 1\}$ .  $\square$

Continuing, we explore the extender set sequences of non-sofic  $\beta$ -shifts. Recall that predecessor sets in one-sided shifts must be considered as sets of finite words, rather than left-infinite sequences. Now that we must discuss both predecessor and follower sets in the same proof, for Lemma 3.3.14 we will also consider follower sets to be sets of finite words, rather than right-infinite sequences, to avoid confusion. Thus, extender sets will be viewed as sets of pairs of finite words.

**Lemma 3.3.14.** *Let  $X_\beta$  be non-sofic. Then for any distinct words  $w, v \in L_n(X_\beta)$  with  $w \in L_n(d_\beta^*(1))$ —that is,  $w$  is a word actually appearing in the sequence  $d_\beta^*(1)$ — $w$  and  $v$  have distinct extender sets.*

*Proof.* If both words  $w$  and  $v$  appear in  $d_\beta^*(1)$ ,  $w$  and  $v$  will have different predecessor sets—there will exist some  $n$ -letter word in  $d_\beta^*(1)$  which one of the words is lexicographically less than or equal to but the other is not, namely, whichever of  $w$  and  $v$  is lexicographically lesser. Thus in this case,  $w$  and  $v$  have different extender sets. On the other hand, suppose  $w \in L_n(d_\beta^*(1))$  and  $v \notin L_n(d_\beta^*(1))$ . If  $v$  and  $w$  have different follower or predecessor sets, they trivially have different extender sets, so suppose  $P(w) = P(v)$  and  $F(w) = F(v)$ . (This implies that  $w$  and  $v$  end with the same maximal prefix of  $d_\beta^*(1)$ , and, by the above argument, that  $v \prec w$ . Note that obviously any digits on which  $v$  and  $w$  differ must occur before their shared suffix). Let  $u$  be the shortest finite word such that  $uw$  is a prefix of  $d_\beta^*(1)$ .

Choose a word  $z \in F(w)$  which is lexicographically strictly larger than  $(\sigma^{|uw|}(d_\beta^*(1)))_{|z|}$ . Such a  $z$  must exist: If there does not exist a legal word  $z \in F(w)$  which is lexicographically strictly larger than  $(\sigma^{|uw|}(d_\beta^*(1)))_{|z|}$ , then

either  $w$  is the first  $n$  letters of  $d_\beta^*(1)$ , in which case the requirement that  $v$  and  $w$  end in the same maximal prefix of  $d_\beta^*(1)$  forces  $v = w$ , or else  $(\sigma^{|uw|}(d_\beta^*(1)))_{|z|}$  is the lexicographically largest legal  $|z|$ -letter word in  $X_\beta$ , meaning that  $(\sigma^{|uw|}(d_\beta^*(1)))_{|z|} = (d_\beta^*(1))_{|z|}$ . But since this must be true for every possible value of  $|z|$ , we have that  $\sigma^{|uw|}(d_\beta^*(1)) = d_\beta^*(1)$ , and so  $X_\beta$  is sofic.

After choosing such a  $z$ , it is clear that  $(u, z)$  is not in the extender set of  $w$ . However,  $(u, z)$  is in the extender set of  $v$ :  $v$  is an intrinsically synchronizing word, and since  $P(w) = P(v)$  and  $F(w) = F(v)$ ,  $uv$  and  $vz$  are legal, so  $uvz$  is legal.  $\square$

Note that, in the sofic case, if only one of  $w$  and  $v$  is in  $L_n(d_\beta^*(1))$ ,  $w$  and  $v$  may still have the same extender set:

**Example 3.3.15.** *As we saw in Example 3.3.4, the one-sided golden mean shift is the  $\beta$ -shift for  $\beta = \varphi = \frac{1+\sqrt{5}}{2}$ , and in such a case, we have  $d_\varphi^*(1) = .1010101010\dots$ , a periodic sequence (so the golden mean shift is sofic). We also saw in Example 2.0.25 that because the golden mean shift is a nearest-neighbor shift of finite type, the extender set of a word in the golden mean shift is determined only by the first and last letters of that word. Thus the word 000 has the same extender set as the word 010, despite 010 appearing in  $d_\varphi^*(1)$ , showing that the result of Lemma 3.3.14 does not hold in the sofic case.*

We now use Lemma 3.3.14 to characterize the extender set sequences of non-sofic  $\beta$ -shifts.

**Theorem 3.3.16.** *Let  $X_\beta$  be a non-sofic  $\beta$ -shift. For  $n \in \mathbb{N}$ , let the classes  $S_k$ ,  $0 \leq k \leq n$ , be as defined in the proof of Theorem 3.3.8. Let  $\eta(w, k) :$*

$L_n(X_\beta) \times \{0, 1, \dots, n\} \rightarrow \{0, 1\}$  be defined by

$$\eta(w, k) = \begin{cases} 1 & \text{if } \exists v \in S_k, v \neq w \text{ s.t. } P_{X_\beta}(v) = P_{X_\beta}(w) \\ 0 & \text{otherwise.} \end{cases}$$

Then for any  $n \in \mathbb{N}$ ,  $|E_{X_\beta}(n)| = \Phi_{d_\beta^*(1)}(n) + \sum_{w \in L_n(d_\beta^*(1))} \left( \sum_{k=0}^n \eta(w, k) \right)$ .

*Proof.* Let  $X_\beta$  be a non-sofic  $\beta$ -shift and  $n \in \mathbb{N}$ . By Lemma 3.3.14, for any word  $w$  in  $L_n(d_\beta^*(1))$ , and any word  $v \in L_n(X_\beta), v \neq w$  implies  $E_{X_\beta}(w) \neq E_{X_\beta}(v)$ . Thus  $X_\beta$  has  $\Phi_{d_\beta^*(1)}(n)$  distinct extender sets corresponding to words in  $L_n(d_\beta^*(1))$ , and moreover, any words not in  $L_n(d_\beta^*(1))$  will have extender sets distinct from those of words in  $L_n(d_\beta^*(1))$  (but not necessarily from each other). Also, for any  $\beta$ -shift  $X_\beta$  (sofic or not), if  $w$  and  $v$  both fail to be in  $L_n(d_\beta^*(1))$ , then  $w$  and  $v$  are both intrinsically synchronizing, and so  $P(w) = P(v)$  and  $F(w) = F(v)$  is sufficient to prove  $E(w) = E(v)$ . Thus, the number of additional extender sets in  $X_\beta$  corresponding to words not in  $L_n(d_\beta^*(1))$  is the number of pairings of predecessor and follower sets achievable by those words. By Thm.3.3.12, each predecessor set is represented by some  $w \in L_n(d_\beta^*(1))$ . By Thm.3.3.8, the follower set of a word is determined by the class  $S_k, 0 \leq k \leq n$ , in which it lives. Hence the total number of possible pairings of predecessor and follower sets achievable by words not in  $L_n(d_\beta^*(1))$  may be expressed by

$$\sum_{w \in L_n(d_\beta^*(1))} \left( \sum_{k=0}^n \eta(w, k) \right). \quad \square$$

**Remark 3.3.17.** *The formula given in Theorem 3.3.16 is an upper bound on the extender set sequence for sofic  $\beta$ -shifts. The only difference between the two is that in the sofic case, some of the extender sets of words in  $L_n(d_\beta^*(1))$*

may be the same as extender sets of words not in  $L_n(d_\beta^*(1))$ , so the formula may have overcounted.

Though the formula in Theorem 3.3.16 is exact, the following bounds on the extender set sequence of a non-sofic  $\beta$ -shift are simpler to express:

**Corollary 3.** *For any non-sofic  $\beta$ -shift  $X_\beta$ ,  $n \in \mathbb{N}$ ,  $\Phi_{d_\beta^*(1)}(n) \leq |E_{X_\beta}(n)| \leq (n+1)\Phi_{d_\beta^*(1)}(n)$ .*

*Proof.* Clearly for any  $n \in \mathbb{N}$ ,  $|E_{X_\beta}(n)| \geq |P_{X_\beta}(n)| = \Phi_{d_\beta^*(1)}(n)$ , proving the first inequality. For the second inequality, we use the equation from Theorem 3.3.16:

$$|E_{X_\beta}(n)| = \Phi_{d_\beta^*(1)}(n) + \sum_{w \in L_n(d_\beta^*(1))} \left( \sum_{k=0}^n \eta(w, k) \right)$$

Recall that the class  $S_n$  contains only one word,  $(d_\beta^*(1))_n$ . Therefore, for any  $w \in L_n(d_\beta^*(1))$ ,  $\eta(w, n) = 0$ . Hence,

$$\begin{aligned} |E_{X_\beta}(n)| &= \Phi_{d_\beta^*(1)}(n) + \sum_{w \in L_n(d_\beta^*(1))} \left( \sum_{k=0}^{n-1} \eta(w, k) \right) \\ &\leq \Phi_{d_\beta^*(1)}(n) + \sum_{w \in L_n(d_\beta^*(1))} n \\ &= \Phi_{d_\beta^*(1)}(n) + (\Phi_{d_\beta^*(1)}(n)) \cdot n \\ &= (n+1)\Phi_{d_\beta^*(1)}(n). \end{aligned}$$

□

**Remark 3.3.18.** *Corollary 3 shows that for a non-sofic  $\beta$ -shift  $X_\beta$ ,  $|E_{X_\beta}(n)| \leq |F_{X_\beta}(n)| \cdot |P_{X_\beta}(n)|$  for any  $n \in \mathbb{N}$ .*



Finally, we discuss the application of our results to two-sided  $\beta$ -shifts. The following theorem asserts that for  $\beta$ -shifts, the natural extension will have exactly the same language as the one-sided  $\beta$ -shift  $X_\beta$ .

**Theorem 3.3.19.** *For any  $\beta > 1$ , let  $X_\beta$  be the one-sided  $\beta$ -shift and  $\widehat{X}_\beta$  be the two-sided  $\beta$ -shift formed by the natural extension of  $X_\beta$ . Then  $L(X_\beta) = L(\widehat{X}_\beta)$ .*

*Proof.* By definition of the natural extension,  $L(\widehat{X}_\beta) \subseteq L(X_\beta)$ . Let  $w \in L(X_\beta)$ . Then there exists a one-sided sequence  $z$  such that  $z \in X_\beta$  and  $w$  is a subword of  $z$ . But then for any  $k$ ,  $0^k z \in X_\beta$  as well. (If every shift of  $z$  is lexicographically less than or equal to  $d_\beta^*(1)$ , then surely the same is true of  $0^k z$ ). By taking limits of such sequences, we see that the sequence  $0^\infty z$  is an element of  $\widehat{X}_\beta$ , and so  $w \in L(\widehat{X}_\beta)$ , completing the proof.  $\square$

**Remark 3.3.20.** *The reader may check that Thm. 3.3.19 allows us to generalize any of the results of this section about one-sided  $\beta$ -shifts to two-sided  $\beta$ -shifts as well.*

We finish the section with two Corollaries which affirm our assertion from Remark 3.1.2 that we may usually ignore predecessor set sequences in favor of follower set sequences.

**Corollary 4.** *Let  $\{\Phi_d(n)\}$  be the complexity sequence of a right-infinite sequence  $d$  such that  $\sigma^i(d) \preceq d$  for all  $i \in \mathbb{N}$ . Then the sequence  $\{\Phi_d(n)\}$  is the follower set sequence of  $X_\beta$  for some  $\beta > 1$ .*

*Proof.* Let  $d$  be a right-infinite sequence with complexity  $\{\Phi_d(n)\}$  such that  $\sigma^i(d) \preceq d$  for all  $i \in \mathbb{N}$ . By Theorem 1.0.6, there exists a  $\beta$  such that  $\{\Phi_d(n)\}$  is the predecessor set sequence of  $X_\beta$ . Then take  $\widehat{X}_\beta$  to be the natural extension

of  $X_\beta$ . Using Theorem 3.3.19 it can be shown that the predecessor set sequence of  $\widehat{X}_\beta$  is also  $\{\Phi_d(n)\}$ . Define a new two-sided shift space  $-\widehat{X}_\beta$  as in Remark 3.1.2. Then the follower set sequence of  $-\widehat{X}_\beta$  is

$$\{|F_{-\widehat{X}_\beta}(n)|\} = \{|P_{\widehat{X}_\beta}(n)|\} = \{|P_{X_\beta}(n)|\} = \{\Phi_d(n)\}.$$

□

**Corollary 5.** *Let  $\{\Phi_d(n)\}$  be the complexity sequence of any right-infinite sequence  $d$ . Then the sequence  $\{\Phi_d(n) + 1\}$  is the follower set sequence of  $X_\beta$  for some  $\beta > 1$ .*

*Proof.* Let  $d$  be a right-infinite sequence with complexity  $\{\Phi_d(n)\}$ . By Theorem 1.0.7, there exists a  $\beta$  such that  $\{\Phi_d(n) + 1\}$  is the predecessor set sequence of  $X_\beta$ . Following the same argument as in Corollary 4, we may construct  $-\widehat{X}_\beta$  such that

$$\{|F_{-\widehat{X}_\beta}(n)|\} = \{|P_{X_\beta}(n)|\} = \{\Phi_d(n) + 1\}.$$

□

# Bibliography

- [1] F. Blanchard.  $\beta$ -expansions and symbolic dynamics. *Theoret. Comput. Sci.*, 65(2):131–141, 1989.
- [2] N. Pytheas Fogg. *Substitutions in dynamics, arithmetics and combinatorics*, volume 1794 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2002. Edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel.
- [3] T. French, N. Ormes, and R. Pavlov. Subshifts with Slowly Growing Numbers of Follower Sets. *AMS Contemporary Math Conference Proceedings*, to appear.
- [4] Thomas French. Characterizing follower and extender set sequences. *Dynamical Systems*, 2015.
- [5] Steve Kass and Kathleen Madden. A sufficient condition for non-soficness of higher-dimensional subshifts. *Proc. Amer. Math. Soc.*, 141(11):3803–3816, 2013.
- [6] Mark V. Lawson. *Finite automata*. Chapman & Hall/CRC, Boca Raton, FL, 2004.

- [7] Douglas Lind and Brian Marcus. *An introduction to symbolic dynamics and coding*. Cambridge University Press, Cambridge, 1995.
- [8] Marston Morse and Gustav A. Hedlund. Symbolic Dynamics. *Amer. J. Math.*, 60(4):815–866, 1938.
- [9] Nic Ormes and Ronnie Pavlov. Extender sets and multidimensional subshifts. *Ergodic Theory and Dynamical Systems*, FirstView:1–16, 4 2015.
- [10] W. Parry. On the  $\beta$ -expansions of real numbers. *Acta Math. Acad. Sci. Hungar.*, 11:401–416, 1960.
- [11] Ronnie Pavlov. A class of nonsofic multidimensional shift spaces. *Proc. Amer. Math. Soc.*, 141(3):987–996, 2013.
- [12] A. Rényi. Representations for real numbers and their ergodic properties. *Acta Math. Acad. Sci. Hungar.*, 8:477–493, 1957.