

Chapter 4

Some Applications of Algebra to Automatic Sequences



Jason Bell

Abstract We give an overview of the theory of rings satisfying a polynomial identity and use this to give a proof of a characterization due to Berstel and Reutenauer of automatic and regular sequences in terms of two properties, which we call the shuffle property and the power property. These properties show that if one views an automatic sequence f as a map on a free monoid on k -letters to a finite subset of a ring, then the values of f are closely related to values of f on related words obtained by permuting letters of the word. We use this characterization to give answers to three questions from Allouche and Shallit, two of which have not appeared in the literature. The final part of the chapter deals more closely with the shuffle property, and we view this as giving a generalization of regular sequences. We show that sequences with the shuffle property are closed under the process of taking sums and taking products; in addition we show that there is closure under a noncommutative product, which turns the collection of shuffled sequences into a noncommutative algebra. We show that this algebra is very large, in the sense that it contains a copy of a free associative algebra on countably many generators. We conclude by giving some open questions, which we hope will begin a more careful study of shuffled sequences.

4.1 Introduction

We recall that, given a finite set Δ , a sequence $f : \mathbb{N} \rightarrow \Delta$ is said to be k -automatic if the n^{th} term of this sequence is generated by a finite-state machine taking the base- k expansion of n as input, starting with the least significant digit. Automatic sequences appear in many different contexts (see, e.g., [14, 189]), and there are many examples of their application to algebra, such as their involvement in the characterization of algebraic power series over finite fields [14, Chapter 12]. The purpose of this chapter, however, is to look at applications of algebra, in particular the theory of

J. Bell (✉)

Department of Pure Mathematics, University of Waterloo, 200 University Ave. W., Waterloo, ON, Canada N2L 3G1

e-mail: jpbell@uwaterloo.ca

polynomial identities, in giving a characterization of automatic sequences due to Berstel and Reutenauer [77, Chapter 3] and to use this characterization to answer some open questions about automatic and regular sequences.

Another way of defining the k -automatic property comes from looking at the k -kernel of a sequence. The k -kernel of a sequence $f(n)$ is defined to be the collection of sequences of the form $f(k^i n + j)$ where $i \geq 0$ and $0 \leq j < k^i$. A sequence is k -automatic if and only if its k -kernel is finite. Using this definition of k -automatic sequences, Allouche and Shallit [14, 16] generalized the notion of k -automatic sequences, defining k -regular sequences.

Let K be a field. Given a sequence $f(n)$ taking values in K , we create a vector subspace of $K^{\mathbb{N}}$, $V(f(n); k)$, which is defined to be the subspace spanned by all sequences $f(k^i n + j)$, where $i \geq 0$ and $0 \leq j < k^i$.

Definition 4.1.1. A sequence is k -regular if $V(f(n); k)$ is a finite-dimensional K -vector space.

Since the k -kernel of a sequence $f(n)$ spans $V(f(n); k)$ as a K -vector space, we see that a k -automatic sequence is necessarily k -regular. We remark that one can adopt a slightly more general viewpoint by replacing the field K by a commutative ring and asking that submodules generated by elements of the kernel be finitely generated. Everything we do in this chapter holds in this more general setting, but for ease of exposition we restrict to the case where our sequences are K -valued with K a field.

In the first half of this chapter, we take a ring theoretic look at automatic sequences and regular sequences and prove a result due to Berstel and Reutenauer [77, Chapter 3], which is not as well known as it probably should be. Part of the reason for this is that their result is written in the context of noncommutative rational series, so we give a straightforward translation of these results into our setting. In fact, we shall give a quantitative version of their result, which we hope will be of future use. We then give some applications of this result.

Let \mathcal{A} be a finite alphabet. We let \mathcal{A}^* denote the free monoid on \mathcal{A} ; that is, \mathcal{A}^* is the collection of all words on \mathcal{A} . We let ε denote the empty word on \mathcal{A} .

We are interested in maps $f : \mathcal{A}^* \rightarrow K$, where \mathcal{A} is some finite alphabet and K is a field. In the case that $\mathcal{A} = \{1, 2, \dots, k\}$ for some positive integer k , we can regard a word in \mathcal{A}^* as an integer as follows. We associate 0 to the empty word ε , and for nontrivial words $a_m \cdots a_0 \in \mathcal{A}^*$ with $1 \leq a_i \leq k$, we associate a positive integer using the correspondence

$$a_m \cdots a_0 \mapsto a_m k^m + a_{m-1} k^{m-1} + a_{m-2} k^{m-2} + \cdots + a_0. \quad (4.1)$$

Observe that this gives a bijection between \mathcal{A}^* and the natural numbers and hence, in this case, we may think of f as being a sequence indexed by the nonnegative integers taking values in K . We note that it is more common to use the alphabet $\{0, 1, \dots, k-1\}$ and instead restrict to the regular sublanguage of the monoid $\{0, \dots, k-1\}$ consisting of words whose first letter is not zero when dealing with

k -automatic and k -regular sequences. The point of view we adopt is completely equivalent, but we find the framework used in this chapter easier to deal with.

Using this correspondence and the k -kernel definition of a k -automatic sequence, we see that we can think of a k -automatic sequence taking values in a field K as being a map $f : \{1, 2, \dots, k\}^* \rightarrow K$ with the property that the collection of maps $f^u(w) := f(wu)$ obtained by taking a word $u \in \{1, 2, \dots, k\}^*$ is finite. Similarly, if the vector space of set maps from $\{1, \dots, k\}^*$ to K generated by the maps of the form f^u is finite-dimensional, then f is k -regular. We shall give an overview of the work of Berstel and Reutenauer [77, Chapter 3], describing k -automatic and k -regular sequences, in terms of two additional properties, which for the purposes of this chapter we shall call the *shuffle property* and the *power property*. These properties are defined in Section 4.2 and Section 4.3, respectively. In Section 4.4 we present some results from the theory of rings satisfying a polynomial identity which will be necessary in giving the aforementioned characterization of k -regular sequences. In Section 4.5 we show that for a sequence, possessing these two properties is equivalent to being regular. If in addition to having these two properties, the sequence only takes on a finite number of values, the sequence is automatic. In Section 4.6 and Section 4.7, we use our characterization to answer some questions of Allouche and Shallit about whether certain sequences are k -regular and k -automatic. In Section 4.8 we develop the basic properties of sequences with the shuffle property and show that they have nice closure properties. In Section 4.9, we conclude with some open problems and some useful remarks.

4.2 The Shuffle Property

In this section we define the *shuffle property* and show that regular sequences necessarily possess this property. Let \mathcal{A} be a finite alphabet and let K be a field. Given a map $f : \mathcal{A}^* \rightarrow K$, we say that f has the *d -shuffle property* if for any words $w, w_1, \dots, w_d, w' \in \mathcal{A}^*$ we have

$$\sum_{\sigma \in S_d} \text{sgn}(\sigma) f(w w_{\sigma(1)} w_{\sigma(2)} \cdots w_{\sigma(d)} w') = 0. \quad (4.2)$$

We define the *d^{th} -shuffle function*

$$\text{Shuf}_d(f; w, w_1, \dots, w_d, w') := \sum_{\sigma \in S_d} \text{sgn}(\sigma) f(w w_{\sigma(1)} w_{\sigma(2)} \cdots w_{\sigma(d)} w'). \quad (4.3)$$

Example 4.2.1. Let $\mathcal{A} = \{0, 1\}$, let K be the field of rational numbers, and define $f(w)$ to be the number of ones in the word $w \in \mathcal{A}^*$. Then f has the 2-shuffle property.

Example 4.2.2. Let $\mathcal{A} = \{0, 1\}$ and define $f(w)$ to be the nonnegative integer whose binary expansion is equal to w . Then f has the 4-shuffle property, but does not have the 3-shuffle property.

We postpone the proof of the fact that f has the 4-shuffle property till after the proof of Proposition 4.2.8. To see that f does not have the 3-shuffle property, take $w_1 = 0, w_2 = 1, w_3 = 10$. Then $f(w_1w_2w_3) = [0110]_2 = 6$. Similarly, $f(w_1w_3w_2) = 5, f(w_2w_3w_1) = 12, f(w_2w_1w_3) = 10, f(w_3w_1w_2) = 9, f(w_3w_2w_1) = 10$. Thus

$$\sum_{\sigma \in S_3} \text{sgn}(\sigma) f(w_{\sigma(1)}w_{\sigma(2)}w_{\sigma(3)}) = 6 - 5 + 12 - 10 + 9 - 10 = 2 \neq 0.$$

The motivation for this shuffle property definition comes from the following theorem of Amitsur and Levitzki [20].

Theorem 4.2.3 (Amitsur-Levitzki [20]). *Let $\mathbf{A}_1, \dots, \mathbf{A}_{2m}$ be $m \times m$ matrices with entries in a commutative ring C . Then*

$$S_d(\mathbf{A}_1, \dots, \mathbf{A}_{2m}) := \sum_{\sigma \in S_{2m}} \text{sgn}(\sigma) \mathbf{A}_{\sigma(1)} \cdots \mathbf{A}_{\sigma(2m)} = 0. \tag{4.4}$$

Before we prove this result, we need a basic result about matrices.

Lemma 4.2.4. *Let C be a commutative \mathbb{Q} -algebra and let $\mathbf{Y} \in M_n(C)$. Suppose that \mathbf{Y} has the property that the trace of \mathbf{Y}^i is zero for $i = 1, 2, \dots, n$. Then $\mathbf{Y}^n = 0$.*

Proof. Let $y_{i,j}$ denote the (i, j) -entry of \mathbf{Y} . Then we may assume without loss of generality that C is generated by the $y_{i,j}$. Now let $R = \mathbb{Q}[x_{i,j} : 1 \leq i, j \leq n]$ and let $\mathbf{Z} \in M_n(R)$ be the matrix whose (i, j) -entry is $x_{i,j}$. We note that if we impose the constraints on the $x_{i,j}$ that give that the trace of \mathbf{Z}^i is equal to zero for $i = 1, 2, \dots, n$, then we obtain a homomorphic image R' of R and by hypothesis the map $x_{i,j} \mapsto y_{i,j}$ gives a surjective map \mathbb{Q} -algebra homomorphism from R' to C . Thus it is sufficient to show that $\mathbf{Z}^n = 0$ in $M_n(R')$. We note that there is a \mathbb{Q} -algebra S that contains R and is a finite R -module and that contains all the eigenvalues of the matrix \mathbf{Z} . Then the relations imposed by setting the trace of \mathbf{Z}^i to zero for $i = 1, 2, \dots, n$ give a homomorphic image S' of S that is a finite R' -module. Now \mathbf{Z} is triangularizable in $M_n(S')$ and so we may assume that \mathbf{Z} is upper triangular. Then if $\lambda_1, \dots, \lambda_n \in S'$ are the diagonal entries of \mathbf{Z} , then we have $\sum_{i=1}^n \lambda_i^j = 0$ for $j = 1, 2, \dots, n$. We now claim that $\lambda_1, \lambda_2, \dots, \lambda_n$ are nilpotent elements of S' . Once we establish this claim, we see that some power of \mathbf{Z} is strictly upper triangular and so \mathbf{Z} is nilpotent. Then the Cayley-Hamilton theorem gives that $\mathbf{Z}^n = 0$ in $M_n(S')$ and hence in $M_n(R')$, and so we get $\mathbf{Y}^n = 0$.

To establish the claim, we note that it is sufficient to show that if C is a finitely generated commutative \mathbb{Q} -algebra and $\lambda_1, \dots, \lambda_n \in C$ satisfy $\sum_{i=1}^n \lambda_i^j = 0$ for $j = 1, 2, \dots, n$ then $\lambda_1, \dots, \lambda_n$ are nilpotent. Let N denote the nil radical of C (the set of all nilpotent elements of C). Then we may replace C by C/N , and we see that

it is sufficient to show that if C is a finitely generated reduced (i.e., it has no nonzero nilpotent elements) commutative \mathbb{Q} -algebra and $\lambda_1, \dots, \lambda_n \in C$ satisfy $\sum_{i=1}^n \lambda_i^j = 0$ for $j = 1, 2, \dots, n$, then $\lambda_1, \dots, \lambda_n$ are all zero. Since a reduced commutative ring embeds in a direct product of integral domains, we then see we can use projection maps to reduce to the case of an integral domain. So we prove the more general fact that if C is a commutative \mathbb{Q} -algebra that is an integral domain and $\lambda_1, \dots, \lambda_n \in C$ are distinct and nonzero and $\beta_1, \dots, \beta_n \in C$ are nonzero then $\sum \beta_i \lambda_i^j \neq 0$ for some $j = 1, \dots, n$. We note that the original claim follows easily from this more general statement (but this requires characteristic zero). To see this general claim, we note that since we are in an integral domain, we get the result for $n = 1$. Now suppose that the result holds for $n < d$ and consider the case where $n = d$. Then if $\sum_{i=1}^d \beta_i \lambda_i^j = 0$ for $j = 1, 2, \dots, d$, we see that

$$\sum_{i=1}^d \beta_i \lambda_i^j \lambda_d - \sum_{i=1}^d \beta_i \lambda_i^{j+1} = 0$$

for $j = 1, 2, \dots, d - 1$. Simplifying, we see that

$$\sum_{i=1}^{d-1} \beta_i (\lambda_i - \lambda_d) \lambda_i^j = 0,$$

for $j = 1, \dots, d - 1$. But this contradicts the induction hypothesis, and we get the claim. This finishes the proof. \square

Proof (Proof of Theorem 4.2.3). We only prove the case when C is an integral domain, which is sufficient for the considerations of this chapter. Here we give an argument due to Rosset [512] (we note that Rosset does the more general case of a commutative ring). We note that one can immediately reduce to the case when C is a field of characteristic zero, since an integral domain C of positive characteristic is a homomorphic image of an integral domain C' of characteristic zero, and one can show that if the result holds for the field of fractions of C' then it holds for C . Now let $\mathbf{A}_1, \dots, \mathbf{A}_{2n}$ be $2n$ elements in $M_n(C)$ and let V be a $2n$ -dimensional C -vector space with basis $\mathbf{e}_1, \dots, \mathbf{e}_{2n}$. For each $i \geq 0$, we recall that there is a d -th exterior product, $\wedge^d V$ is formed by taking the vector space $V_d := \otimes_{i=1}^d V$ and then forming the quotient V_d/W_d where W_d is the subspace of V_d spanned by elements of the form $\mathbf{e}_{i_1} \otimes \dots \otimes \mathbf{e}_{i_d} - \text{sgn}(\sigma) \mathbf{e}_{\sigma(i_1)} \otimes \dots \otimes \mathbf{e}_{\sigma(i_d)}$, where $\sigma \in S_d$ and $1 \leq i_1, \dots, i_d \leq 2n$. Then we let $\mathbf{e}_{i_1} \wedge \dots \wedge \mathbf{e}_{i_d}$ denote the image of $\mathbf{e}_{i_1} \otimes \dots \otimes \mathbf{e}_{i_d}$ in this quotient. Then one can check that $\wedge^d V$ is a $\binom{2n}{d}$ -dimensional space with basis consisting of elements of the form $\mathbf{e}_{i_1} \wedge \dots \wedge \mathbf{e}_{i_d}$ with $i_1 < i_2 < \dots < i_d$.

Now let $E = \bigoplus_{d=0}^{2n} \wedge^d(V)$; this is called the exterior algebra on V . Notice that E is a ring with multiplication formed by taking the natural bilinear “wedge” map $\wedge^i(V) \times \wedge^j(V) \rightarrow \wedge^{i+j}(V)$ given by

$$((\mathbf{e}_{p_1} \wedge \dots \wedge \mathbf{e}_{p_i}), (\mathbf{e}_{q_1} \wedge \dots \wedge \mathbf{e}_{q_j})) \mapsto \mathbf{e}_{p_1} \wedge \dots \wedge \mathbf{e}_{p_i} \wedge \mathbf{e}_{q_1} \wedge \dots \wedge \mathbf{e}_{q_j}.$$

Then E is generated as a C -algebra by $\mathbf{e}_1, \dots, \mathbf{e}_{2n}$. Let E_e denote the subalgebra of E consisting of the direct sum of $\wedge^i V$ with i even. Then E_e is generated by elements of the form $\mathbf{e}_i \wedge \mathbf{e}_j$ as an algebra, and it is straightforward to check, using the relations given above, that these elements commute with one another and so E_e is a commutative ring of characteristic zero. Now consider $B := M_n(C) \otimes_C E \cong M_n(E)$. Let $\mathbf{X} = \mathbf{A}_1 \otimes \mathbf{e}_1 + \dots + \mathbf{A}_{2n} \otimes \mathbf{e}_{2n} \in B$. Then

$$\mathbf{Y} := \mathbf{X}^2 = \sum_{i < j} (\mathbf{A}_i \mathbf{A}_j - \mathbf{A}_j \mathbf{A}_i) \mathbf{e}_i \wedge \mathbf{e}_j \in M_n(E_e).$$

Notice that \mathbf{Y} has trace zero, as it is an E_e -linear combination of commutators. More generally,

$$\mathbf{Y}^i = \sum_{1 \leq j_1, \dots, j_{2i} \leq 2n} \mathbf{A}_{j_1} \cdots \mathbf{A}_{j_{2i}} \mathbf{e}_{j_1} \wedge \cdots \wedge \mathbf{e}_{j_{2i}},$$

which is equal to

$$\sum_{1 \leq j_1 < j_2 < \dots < j_{2i} \leq 2n} \mathbf{S}_{2i}(\mathbf{A}_{j_1}, \dots, \mathbf{A}_{j_{2i}}) \mathbf{e}_{j_1} \wedge \cdots \wedge \mathbf{e}_{j_{2i}}.$$

It is straightforward to check that $\mathbf{S}_{2i}(\mathbf{A}_{j_1}, \dots, \mathbf{A}_{j_{2i}})$ always has trace zero for $i \geq 1$ and so we see that the trace of \mathbf{Y}^i is zero for $i = 1, 2, \dots, n$. Thus \mathbf{Y} is an $n \times n$ matrix over the commutative \mathbb{Q} -algebra E_e , and it has the property that the trace of all of its powers is equal to zero. By Lemma 4.2.4, we have that $\mathbf{Y}^n = 0$. Thus $\mathbf{Y}^n = \mathbf{X}^{2n} = 0$. As before, we have $\mathbf{X}^{2n} = \mathbf{S}_{2n}(\mathbf{A}_1, \dots, \mathbf{A}_{2n}) \mathbf{e}_1 \wedge \cdots \wedge \mathbf{e}_{2n}$ and so we get the desired result. \square

We now show that the d -shuffle property implies all larger shuffle properties hold.

Proposition 4.2.5. *If f has the d -shuffle property, then f has the e -shuffle property for all $e \geq d$.*

Proof. By induction, it is sufficient to prove this when $e = d + 1$. Notice that

$$\begin{aligned} & \text{Shuf}_{d+1}(f; w, w_1, \dots, w_{d+1}, w') \\ &= \sum_{i=1}^{d+1} (-1)^{i-1} \text{Shuf}_d(f, ww_i, w_1, \dots, \widehat{w}_i, \dots, w_{d+1}, w'), \end{aligned}$$

where \widehat{w}_i means that w_i is omitted from the list. Hence if f has the d -shuffle property, it must also have the $(d + 1)$ -shuffle property. \square

We now introduce some notation. Given a finite alphabet \mathcal{A} , a field K , a word $w \in \mathcal{A}^*$ and a function $f : \mathcal{A} \rightarrow K$, we define two functions $f_w, f^w : \mathcal{A} \rightarrow \mathcal{A}$ by

$$f_w(u) := f(wu) \quad \text{and} \quad f^w(u) = f(uw). \tag{4.5}$$

We now give some definitions.

Definition 4.2.6. Given a finite alphabet \mathcal{A} , a field K , and a function $f : \mathcal{A}^* \rightarrow K$, we say that f is left (resp. right) \mathcal{A} -regular if the K -vector space spanned by the functions $\{f_w \mid w \in \mathcal{A}^*\}$ (resp. $\{f^w \mid w \in \mathcal{A}^*\}$) is finite-dimensional. If in addition to being left (resp. right) \mathcal{A} -regular, the range of f is a finite subset of K , we say that f is left (resp. right) \mathcal{A} -automatic.

Later, we will give Kleene's theorem, which states that being left \mathcal{A} -regular is equivalent to being right \mathcal{A} -regular. Thus we will omit the words left and right and just use the term \mathcal{A} -regular. In the case that $\mathcal{A} = \{1, 2, \dots, k\}$, we shall say that a right \mathcal{A} -regular function is k -regular and shall say that a right \mathcal{A} -automatic function is k -automatic.

Notice that this definition of k -regular coincides with the definition of k -regular given by Allouche and Shallit [14] and the definition of k -automatic coincides with the conventional definitions of the k -automatic property.

For convenience, we use the following notation. Given a finite alphabet \mathcal{A} , a field K , and a function $f : \mathcal{A}^* \rightarrow K$, we let $L(f)$ denote the K -vector space spanned by the functions $\{f_w \mid w \in \mathcal{A}^*\}$, and we let $R(f)$ denote the vector space spanned by $\{f^w \mid w \in \mathcal{A}^*\}$.

Proposition 4.2.7. *Let \mathcal{A} be a finite alphabet and let f be a left (resp. right) \mathcal{A} -regular function taking values in a field K . Let m denote the dimension of $L(f)$ (resp. the dimension of $R(f)$). Then there exist some $m \geq 1$, $m \times m$ matrices $\{\mathbf{A}_a \mid a \in \mathcal{A}\}$ with entries in K , and $\mathbf{v}, \mathbf{w} \in K^{d \times 1}$ such that*

$$f(x_1 \cdots x_i) = \mathbf{w}^T \mathbf{A}_{x_1} \cdots \mathbf{A}_{x_i} \mathbf{v}$$

for all words $x_1 \cdots x_i \in \mathcal{A}^*$.

Proof. Choose $\varepsilon = w_1, \dots, w_m \in \mathcal{A}^*$ such that f_{w_1}, \dots, f_{w_m} span the vector space $L(f)$. Given $x \in \mathcal{A}$, for each i pick $c_{i,j} \in K, j \leq m$, such that

$$f_{xw_i} = \sum_{j=1}^m c_{i,j} f_{w_j}.$$

Define the $m \times m$ matrix \mathbf{A}_x whose (i, j) -entry is given by $c_{i,j}$. Take \mathbf{v} to be the $m \times 1$ column vector whose i^{th} coordinate is $f(w_i)$. Notice that if $x_1, \dots, x_i \in \mathcal{A}$, then

$$\mathbf{e}_1^T \mathbf{A}_{x_1} \cdots \mathbf{A}_{x_i} \mathbf{v} = f(x_1 \cdots x_i).$$

In the case that f is right \mathcal{A} -regular, an analogous construction gives the same result. \square

Proposition 4.2.8. *Let \mathcal{A} be a finite alphabet, let K be a field, and let $f : \mathcal{A}^* \rightarrow K$ be a left (resp. right) \mathcal{A} -regular sequence. Then f has the d -shuffle property for $d = 2\dim(L(f))$ (resp. $d = 2\dim(R(f))$).*

Proof. Let m denote the dimension of $L(f)$. Since f is left \mathcal{A} -regular, there exist $m \times m$ matrices $\{\mathbf{A}_x \mid x \in \mathcal{A}\}$ with entries in K and some vector $\mathbf{v} \in K^{d \times 1}$ such that

$$f(x_1 \cdots x_m) = \mathbf{e}_1^T \mathbf{A}_{x_1} \cdots \mathbf{A}_{x_m} \mathbf{v}$$

for all words $x_1 \cdots x_m \in \mathcal{A}^*$. Let $d = 2m$. We claim that f has the d -shuffle property. To see this, let w_1, \dots, w_d, w, w' be words in \mathcal{A}^* . Let \mathcal{S} denote the monoid on $\{\mathbf{A}_x \mid x \in \mathcal{A}^*\}$. Then there exist matrices $\mathbf{U}_1, \dots, \mathbf{U}_d, \mathbf{U}, \mathbf{U}'$ in \mathcal{S} which correspond to w_1, \dots, w_d, w, w' , respectively, given by the correspondence $\varepsilon \mapsto \mathbf{I}_d$ and

$$x_1 \cdots x_m \in \mathcal{A}^* \mapsto \mathbf{A}_{x_1} \cdots \mathbf{A}_{x_m}.$$

Then

$$f(w w_{\sigma(1)} \cdots w_{\sigma(d)} w') = \mathbf{e}_1^T \mathbf{U} \mathbf{U}_{\sigma(1)} \cdots \mathbf{U}_{\sigma(d)} \mathbf{U}' \mathbf{v}.$$

Hence

$$\begin{aligned} \sum_{\sigma \in S_d} \text{sgn}(\sigma) f(w w_{\sigma(1)} \cdots w_{\sigma(d)} w') &= \sum_{\sigma \in S_d} \mathbf{e}_1^T \mathbf{U} \left(\sum_{\sigma \in S_d} \mathbf{U}_{\sigma(1)} \cdots \mathbf{U}_{\sigma(d)} \right) \mathbf{U}' \mathbf{v} \\ &= 0, \end{aligned}$$

where the last step follows from the Amitsur-Levitzki theorem. \square

We now prove that the function given in Example 4.2.2 has the 4-shuffle property. We note that if $f : \{0, 1\}^* \rightarrow \mathbb{Q}$ is the function which maps a word w to $[w]_2$, where $[w]_2$ is the natural number whose binary expansion is equal to w , then the \mathbb{Q} -vector space $R(f)$ is two-dimensional, spanned by f and the constant function g which sends every word to 1. To see this, notice that $f^w(u) = [uw]_2 = 2^{\text{length}(w)} [u]_2 + [w]_2$ and so

$$f^w = 2^{\text{length}(w)} f + [w]_2 g.$$

Clearly f and g are both in $R(f)$ and are linearly independent. Thus $\dim(R(f)) = 2$ and so f has the 4-shuffle property.

4.3 The Power Property

In this section we define the *power property*, which is the second ingredient of the characterization of automatic and regular sequences of Berstel and Reutenauer. Given a finite alphabet \mathcal{A} and a field K , we say that $f : \mathcal{A} \rightarrow K$ has the d -power property if for any word w_0 , there exists a polynomial $\Phi(t) \in K[t]$ of degree at most

d with constant coefficient 1 such that for any words w, w' we have

$$\Phi(t) \left(\sum_{i=0}^{\infty} f(ww_0^i w') \right) t^i,$$

is a polynomial in $K[t]$ of degree at most d . We shall say that f has the *power property* if f has the d -power property for some $d \geq 1$.

Lemma 4.3.1. *Let K be a field and let \mathbf{X} be a $d \times d$ matrix with entries in K . Then*

$$\sum_{i=0}^{\infty} \mathbf{X}^i t^i = \mathbf{Y}(t) \det(1 - t\mathbf{X})^{-1},$$

for some matrix \mathbf{Y} with entries in $K[t]$ of degree at most $d - 1$.

Proof. Let $\mathbf{Y}(t)$ denote the classical adjoint of $1 - t\mathbf{X}$. Then $\mathbf{Y}(t)$ is a matrix with entries given by polynomials in t of degree at most $d - 1$ and

$$\mathbf{Y}(t)(1 - t\mathbf{X}) = \det(1 - t\mathbf{X})\mathbf{I}_d.$$

Notice that

$$\mathbf{Y}(t)(1 - t\mathbf{X}) \sum_{i=0}^{\infty} \mathbf{X}^i t^i = \mathbf{Y}(t)$$

and so

$$\sum_{i=0}^{\infty} \mathbf{X}^i t^i = \mathbf{Y}(t) \det(1 - t\mathbf{X})^{-1}.$$

□

Proposition 4.3.2. *Let f be a \mathcal{A} -regular function taking values in a field K . Then f has the m -power property, where m is the dimension of $L(f)$.*

Proof. It suffices to show that for any word w_0 and w, w' that

$$\sum_{i=1}^{\infty} f(ww_0^i w') t^i$$

is a rational function in t whose numerator and denominator have degrees that are at most m (with denominator independent of w, w' and depending only upon w_0 and having constant coefficient 1). By Proposition 4.2.7, there exists some m and $m \times m$ matrices $\mathbf{U}, \mathbf{U}_0, \mathbf{U}'$ such that $f(ww_0^i w') = \mathbf{e}_1^T \mathbf{U} \mathbf{U}_0^i \mathbf{U}' \mathbf{v}$. We then have

$$\begin{aligned}
 & \sum_{i=0}^{\infty} f(w w_0^i w') t^i \\
 &= \sum_{i=0}^{\infty} \mathbf{e}_1^T \mathbf{U} \left(\mathbf{U}_0^i \right) \mathbf{U}' \mathbf{v} \\
 &= \mathbf{e}_1^T \mathbf{U} \left(\sum_{i=0}^{\infty} \mathbf{U}_0^i t^i \right) \mathbf{U}' \mathbf{v} \\
 &= \mathbf{e}_1^T \mathbf{U} \mathbf{Y}(t) \mathbf{U}' \mathbf{v} \cdot \det(1 - \mathbf{U}_0 t)^{-1},
 \end{aligned}$$

where $\mathbf{Y}(t)$ is the classical adjoint of $1 - \mathbf{U}_0 t$. This expression is easily seen to be a rational function of the form

$$P(t)\Phi(t)^{-1},$$

and the numerator and denominator have degree at most m and $\Phi(0) = 1$ and Φ depends only upon \mathbf{U}_0 and hence only on w_0 . □

4.4 Shirshov’s Height Theorem

We now introduce an important result in combinatorial ring theory: Shirshov’s theorem. We have seen that \mathcal{A} -regular functions have both the shuffle and power properties. In fact, it is the case that the shuffle and power properties characterize regular sequences. To deduce this we need a famous combinatorial result due to Shirshov. We first take a detour and survey the beautiful field of polynomial identity algebras.

Definition 4.4.1. Let K be a field and let B be a K -algebra. We say that B is a *polynomial identity ring* if there exists a nonzero, noncommutative polynomial $p(x_1, \dots, x_d)$ with coefficients in K , such that $p(b_1, \dots, b_d) = 0$ for all $b_1, \dots, b_d \in B$. In this case the polynomial p is called a *polynomial identity* for B . The total degree of the polynomial identity p of B of least positive degree is called the *PI degree* of B .

Example 4.4.2. Any commutative algebra B is a polynomial identity ring since it satisfies the identity $x_1 x_2 - x_2 x_1 = 0$.

Example 4.4.3 (Wagner). The ring of 2×2 matrices over a field K satisfies the identity $[x_1, [x_2, x_3]^2] = 0$, where $[a, b] = ab - ba$.

Proof. Notice that if \mathbf{X} is a 2×2 matrix then by the Cayley-Hamilton theorem

$$\mathbf{X}^2 - \text{Tr}(\mathbf{X})\mathbf{X} + \det(\mathbf{X})\mathbf{I}_2 = 0.$$

Hence if \mathbf{X} has trace 0, then its square is a scalar matrix. In particular the square of a commutator must commute with every 2×2 matrix. \square

An important fact is that if a ring B satisfies a nontrivial polynomial identity, then it in fact satisfies a homogeneous multilinear identity (i.e., each monomial occurring in the identity has degree precisely one in each variable).

Proposition 4.4.4. *Let K be a field and let B be a K -algebra satisfying a polynomial identity of degree at most d . Then B satisfies a multilinear homogeneous identity of degree at most d .*

Proof. Let $m(f)$ be the maximum degree of a variable appearing in f . Among all nonzero polynomial identities, we pick one with the property that $m(f)$ is minimal. Let us call this minimum m . Among all such identities with $m(f) = m$, we pick one with the property that the number of variables of degree m is minimal. Let $f(x_1, \dots, x_d)$ be such a minimal polynomial identity for the ring B . By permuting the variables, if necessary, we may assume that m is the degree of x_1 in f . Consider the identity $g(x_1, y_1, x_2, \dots, x_d) := f(x_1 + y_1, \dots, x_d) - f(x_1, \dots, x_d) - f(y_1, \dots, x_d) \in K\{x_1, y_1, x_2, \dots, x_d\}$. We note that this is an identity for B . Then it is straightforward to see that this transforms any monomial of degree m in x_1 to a monomial of total degree m in x_1 and y_1 and no terms of degree m in just x_1 or just y_1 . That means that either $m(g) < m$ or $m(g) = m$ but the number of variables of degree m in g is strictly less than that of f . By minimality of f we have that $g = 0$. But this occurs only if $m = m(f) = 1$. So having $m = 1$ says that every monomial appears with degree at most 1. Now pick a monomial occurring in f with nonzero coefficient of smallest degree, say $r \leq d$. By relabeling indices, we may assume that the monomial is $x_1 \cdots x_r$. Then consider $f(x_1, \dots, x_r, 0, \dots, 0)$. This is nonzero and must be homogeneous by minimality of r .

Notice this process yields an algorithm to convert a polynomial identity into a homogeneous multilinear identity. One begins with an identity, and if it is of degree strictly greater than one in some variable, say x_1 , then we add a new variable y_1 , and we look at $f(x_1 + y_1, \dots) - f(x_1, \dots) - f(y_1, \dots)$. This creates an additional variable, but the total degree does not increase. If we keep repeating this process, the argument above shows that it must terminate and so we obtain an identity in some set of variables in which each variable occurs with degree at most 1. Then we pick a monomial of minimal length and set all variables not occurring in this monomial equal to zero to get a homogeneous multilinear identity. Notice that the total degree never increases at any step, so we see the total degree of the homogeneous multilinear identity we ultimately produce is at most that of the original identity. \square

Remark 4.4.5. We note that in the case that the polynomial $p(x_1, \dots, x_d)$ is multilinear and homogeneous, to check that a K -algebra B satisfies this identity, it is sufficient to check that $p(b_1, \dots, b_d) = 0$ for $(b_1, \dots, b_d) \in Y^d$ where Y is a K -spanning set for B .

Proof. To see this, we remark that if $a_1, \dots, a_d \in B$ then we can write each a_i as a K -linear combination of elements of Y . Then using multilinearity, we can write

$p(a_1, \dots, a_d)$ as a K -linear combination of elements of the form $p(b_1, \dots, b_d)$ with $(b_1, \dots, b_d) \in Y^d$. Thus if the latter all vanish, then p is necessarily an identity for B . \square

We will make use of this fact for algebras B of the form $B := K\{x_1, \dots, x_k\}/I$, where I is a two-sided ideal of the free algebra on noncommuting variables x_1, \dots, x_k . In this case, we can take the images of the words in x_1, \dots, x_k to be our spanning set, and so to check that a homogeneous multilinear identity holds for an algebra B , it suffices to check that it vanishes when evaluated at elements from the semigroup generated by a finite set of generators.

Arguably, the most important types of identities studied in the theory of polynomial identities are the *standard identities*, which we already encountered, albeit in a different form, when addressing the shuffle property. The n^{th} standard identity is defined as follows:

$$S_n(x_1, \dots, x_n) = \sum_{\sigma \in S_n} \text{sgn}(\sigma)x_{\sigma(1)} \cdots x_{\sigma(n)}. \tag{4.6}$$

The Amitsur-Levitzki theorem shows that the ring of $n \times n$ matrices over a commutative ring satisfies the $2n^{\text{th}}$ standard identity. In fact, every finitely generated algebra satisfying a polynomial identity must satisfy one of these standard identities.

Theorem 4.4.6 (Braun). *Let K be a field and let B be a finitely generated K -algebra satisfying an identity. Then B satisfies the identity S_n for some $n \geq 1$.*

Proof. See Braun [105] or Amitsur and Small [21, Corollary 1.2.8A]. \square

We recall that if we are given a free monoid \mathcal{A}^* with $\mathcal{A} = \{x_1, \dots, x_k\}$, we can put a *pure lexicographic order* \preceq on \mathcal{A}^* by declaring that

$$x_1 < x_2 < \cdots < x_k.$$

In this order we declare that if w is a proper initial factor of w' then $w < w'$. So, for example, $x_1 \preceq x_1x_3$ and $x_2x_3x_4 < x_3x_1$. We note that this lexicographic order extends to right infinite words on \mathcal{A} . If we alter the order somewhat and use the pure lexicographic order to order words of the same length and for words of different length declare that the longer word is bigger, then this new order is called the *degree lexicographic order*.

Using intricate combinatorial techniques, Shirshov proved the following beautiful result.

Theorem 4.4.7 (Shirshov). *Let $\mathcal{A} = \{x_1, \dots, x_k\}$ be a finite alphabet, and let m be a positive integer. If w is a right-infinite word over the alphabet \mathcal{A} , then either there is some nontrivial word $w_0 \in \mathcal{A}^*$ such that w_0^d is a factor of w for every $d \geq 1$ or w contains a finite factor of the form $w_1w_2 \cdots w_m$ where $w_1 \succ w_2 \succ \cdots \succ w_m$ are nontrivial words in \mathcal{A}^* with no w_i equal to a prefix of w_j for $i \neq j$ and \succ is the pure lexicographic order induced by $x_1 \succ x_2 \succ \cdots \succ x_k$.*

We postpone the proof of Shirshov's theorem until we have developed a few basic combinatorial tools.

To prove Theorem 4.4.7, we require two basic combinatorial tools. The first is König's infinity lemma; the second is a theorem of Furstenberg.

Theorem 4.4.8 (König's infinity lemma). *Let \mathcal{A} be a finite alphabet, and let \mathcal{S} be an infinite subset of \mathcal{A}^* that is closed under the process of taking factors. Then there exists $w \in \mathcal{A}^{\mathbb{N}}$ such that every finite factor of w is in \mathcal{S} .*

Proof. We define $w = a_1a_2 \cdots$ as follows. Since \mathcal{S} is infinite, there is some $a_1 \in \mathcal{A}$ such that a_1 is the first letter of infinitely many elements of \mathcal{S} ; now suppose that for every $i \geq 1$ we have defined a word $a_1a_2 \cdots a_i$ with the property that it is a prefix of infinitely many elements of \mathcal{S} . Then since \mathcal{A} is finite, there is some $a_{i+1} \in \mathcal{A}$ such that $a_1a_2 \cdots a_i a_{i+1}$ is a prefix of infinitely many elements of \mathcal{S} . Continuing in this way, we obtain an infinite word w with the desired property. \square

The next result is Furstenberg's theorem, which is part of a more general result in Ergodic theory. We give an algebraic proof in the case in which we are interested. We recall that a right-infinite word is *uniformly recurrent* if each factor u has the property that there is some natural number $N = N(u)$ such that whenever u occurs as a factor, its next occurrence is at most N positions later in our right-infinite word. As far as we know, this rather simple proof (which requires the axiom of choice) does not appear in the literature.

Theorem 4.4.9 (Furstenberg's theorem). *Let \mathcal{A} be a finite alphabet and let $w \in \mathcal{A}^{\mathbb{N}}$. Then there is a uniformly recurrent word $u \in \mathcal{A}^{\mathbb{N}}$ such that every finite factor of u is a factor of w .*

Proof. Let K be a field. Write $\mathcal{A} = \{x_1, \dots, x_k\}$, and consider the algebra $B = K\{x_1, \dots, x_k\}/I$, where I is the ideal generated by all monomials that do not occur as a factor of w . Then B is infinite-dimensional as a K -vector space since w is infinite and the images of the factors of w are linearly independent in B . Now let \mathcal{S} denote the collection of ideals J of $K\{x_1, \dots, x_k\}$ that are generated by monomials, contain I , and have the property that B/J is infinite-dimensional. Then \mathcal{S} is nonempty since I is in \mathcal{S} . We note that \mathcal{S} is closed under unions of chains, for if L is the union of a chain in \mathcal{S} , then L is certainly generated by monomials and contains I ; if $K\{x_1, \dots, x_k\}/L$ is finite-dimensional, then there is some N such that L contains all monomials of length $\geq N$. In particular, L is finitely generated as it generated by all monomials of length exactly N along with the finite set of monomials of length $< N$ that are in L . Since L is the union of a chain in \mathcal{S} , and L is finitely generated, there is some element in our chain that must be equal to L . Thus we can pick a maximal element L of \mathcal{S} by Zorn's lemma. Now since L is generated by monomials, $K\{x_1, \dots, x_k\}/L$ is spanned by the images of all monomials over x_1, \dots, x_k that are not in the ideal L . Since $K\{x_1, \dots, x_k\}/L$ is infinite-dimensional and the collection of words that are not in L is closed under taking factors, we see by König's infinity lemma that there is a right-infinite word u with the property that all factors of u are not in L . In particular, they are not in I , and so all factors of u are factors of w .

We claim that u is uniformly recurrent. If not, there is some finite factor u_0 such that there are arbitrarily long factors of u that avoid u_0 . Then let L' be the ideal generated by u_0 and L . Then since there are arbitrarily long words in u that avoid u_0 , we see that these words have nonzero image in the ring $K\{x_1, \dots, x_k\}/L'$, and so $K\{x_1, \dots, x_k\}/L'$ is infinite-dimensional. But this contradicts maximality of L in \mathcal{S} . The result follows. \square

Proof (Proof of Theorem 4.4.7). We follow the proof of Pirillo [484]. By Furstenberg’s theorem, there is some uniformly recurrent right-infinite word v such that every factor of v is a factor of w . Now if v is eventually periodic, then $v = v_0w_0^\omega$ for some v_0, w_0 with w_0 nontrivial, and we get the claim. If v is not eventually periodic, then v must have at least m distinct factors of length $m - 1$ (see [14, Theorem 10.2.6]). Let w_1, \dots, w_m be these distinct factors and suppose that $w_1 \succ w_2 \succ \dots \succ w_m$. Since v is uniformly recurrent, there is a factor v_0 of v that can be written in the form $v_0 = v_1 \dots v_m$ where w_i is a prefix of v_i . Then we see that $v_{\sigma(1)} \dots v_{\sigma(m)} \prec v_1 \dots v_m$, where the inequality is strict whenever σ is not the identity. \square

Shirshov’s theorem has as an immediate application the following result about algebras satisfying a polynomial identity.

Corollary 4.4.10 (Shirshov). *Let K be a field and let B be a finitely generated K -algebra with generators x_1, \dots, x_k that satisfies a polynomial identity. Then either B is finite-dimensional as a K -vector space or there is some word $w \in \{x_1, \dots, x_d\}^*$ such that the images of $1, w, w^2, \dots$ in the algebra B are linearly independent over K .*

Proof. We put the degree lexicographic order, \prec , on the monomials in $\{x_1, \dots, x_k\}^*$ induced by the $x_1 \succ x_2 \succ \dots \succ x_k$. We let \prec_p be the corresponding pure lexicographic order. Let I be the two-sided ideal of $K\{x_1, \dots, x_k\}$ generated by the collection of monomials $w \in \{x_1, \dots, x_k\}^*$ with the property that the image of w in B can be expressed as a K -linear combination of the image of monomials that are strictly smaller than w in the degree lexicographic order. Since B satisfies a polynomial identity, we know that it satisfies a homogeneous multilinear identity by Proposition 4.4.4. We may write this identity as

$$f(x_1, \dots, x_m) = x_1 \dots x_m + \sum_{\sigma \in S_m \setminus \text{id}} c_\sigma x_{\sigma(1)} \dots x_{\sigma(m)}.$$

It then follows that if $w_1 \succ_p w_2 \succ_p w_3 \succ_p \dots \succ_p w_m$ with no w_i equal to a prefix of w_j for $i \neq j$ then the identity f shows that $w_1 w_2 \dots w_m$ can be written as a K -linear combination of strictly smaller words in the degree lexicographic order.

Now if $K\{x_1, \dots, x_k\}/I$ is finite-dimensional, then B is finite-dimensional since by construction the images in B of words over $\{x_1, \dots, x_k\}^*$ that are not in I span B as a K -vector space. So if B is infinite-dimensional over K , then $K\{x_1, \dots, x_k\}/I$ must be infinite-dimensional. Then there are arbitrarily long words on $\{x_1, \dots, x_k\}^*$ that are not in the ideal I , and so by König’s infinite lemma, we see that there is a

right-infinite word w whose factors all lie outside of the ideal I . But by the remarks above, we see that w cannot contain any factor of the form $w_1 w_2 \cdots w_m$ with $w_1 \succ_p w_2 \succ_p w_3 \succ_p \cdots \succ_p w_m$ with no w_i equal to a prefix of w_j for $i \neq j$. Thus, by Shirshov's theorem, there is a nontrivial word w_0 such that for every $d \geq 1$, w_0^d is a factor of w . This means that w_0, w_0^2, \dots are not in I . By definition of I this means that their images in B are linearly independent in B . \square

In fact, there is an even stronger version of Shirshov's theorem for rings satisfying a polynomial identity, which we give now.

Theorem 4.4.11 (Strong version of Shirshov's theorem). *Let C be a commutative ring, and let B be a finitely generated C -algebra with generators x_1, \dots, x_k and which satisfies a polynomial identity of degree d . Then every element in B is a C -linear combination of elements of the form*

$$w_1^{i_1} \cdots w_m^{i_m},$$

where $m \leq d^5 k^d$ and w_1, \dots, w_m are words of length less than d .

Proof. See Theorem 1.2.2 of Amitsur and Small [21]. \square

We have the following unexpected application of Shirshov's theorem to functions with a shuffle property.

Corollary 4.4.12. *Let K be a field and let $f : \mathcal{A} \rightarrow K$ have the d -shuffle property. Then the vector space $L(f)$ (resp. $R(f)$) is spanned by elements of the form $f_{w_1^{j_1} \dots w_m^{j_m}}$ (resp. $f^{w_1^{i_1} \dots w_m^{i_m}}$) with $m \leq d^5 |\mathcal{A}|^d$ and w_1, \dots, w_m words of length at most d .*

Proof. Write $\mathcal{A} = \{x_1, x_2, \dots, x_k\}$. Let $B = K\{x_1, \dots, x_k\}$ be the free K -algebra on k -variables, i.e., the ring of "noncommutative polynomials" in k variables with coefficients in K .

We define an ideal $I \subseteq K\{x_1, \dots, x_k\}$ as follows. Given words w_1, \dots, w_m in x_1, \dots, x_k , we declare that

$$c_1 w_1 + \cdots + c_m w_m \in I$$

if

$$c_1 f_{uw_1} + \cdots + c_m f_{uw_m} \quad \text{is identically 0 for all } u \in \mathcal{A}^*.$$

We note that all such relations form an ideal. Notice that since f has the d -shuffle property, for any d words w_1, \dots, w_d , we have

$$S_d(w_1, \dots, w_d) = \sum_{\sigma \in S_d} \text{sgn}(\sigma) w_{\sigma(1)} \cdots w_{\sigma(d)} \in I.$$

Since the function

$$S_d(x_1, \dots, x_d)$$

is multilinear, we see that $\mathfrak{S}_d(b_1, \dots, b_d) \in I$ for all b_1, \dots, b_d in B . Consequently, B/I satisfies a polynomial identity. It follows from Theorem 4.4.11 that every element of B/I is a K -linear combination of the images of elements of the form $w_1^{j_1} \cdots w_m^{j_m}$ with $m \leq d^5 |\mathcal{A}|^d$, $j_1, \dots, j_m \geq 0$ and w_1, \dots, w_m words of length at most d . It follows that any word $w \in \{x_1, \dots, x_k\}^*$ there is some $b \in I$ such that $w - b$ can be written as a linear combination of words of the form $w_1^{j_1} \cdots w_m^{j_m}$ with $m \leq d^5 |\mathcal{A}|^d$ and w_1, \dots, w_m of length at most d . By definition of I we then have

$$f_w \in \text{Span}_K \{f_{w_1^{j_1} \dots w_m^{j_m}} \mid m \leq d^5 |\mathcal{A}|^d, \text{length}(w_i) \leq d \text{ for } i \leq m\}.$$

We thus obtain the desired result. A similar argument works for the vector space $R(f)$. \square

4.5 Characterization of \mathcal{A} -Regular Sequences

In this section, we are finally able to give the general version of Kleene’s theorem along with the structure result of Berstel and Reutenauer.

Lemma 4.5.1. *Let K be a field, let d and e be positive integers, and let $f : \mathcal{A}^* \rightarrow K$ be a function with the e -power property and let m be at most $d^5 |\mathcal{A}|^d$ and let w_1, \dots, w_m be words of length at most d . Then there exists $e \geq 0$ such that the K -vector space spanned by $\{f_{w_1^{i_1} \dots w_m^{i_m}} \mid i_1, \dots, i_m \geq 0\}$ (resp. $f_{w_1^{i_1} \dots w_m^{i_m}} \mid i_1, \dots, i_m \geq 0\}$) is spanned by*

$$\{f_{w_1^{i_1} \dots w_m^{i_m}} \mid 0 \leq i_1, \dots, i_m \leq e\}$$

(resp. $\{f_{w_1^{i_1} \dots w_m^{i_m}} \mid 0 \leq i_1, \dots, i_m \leq e\}$).

Proof. By the e -power property, there exist polynomials Φ_1, \dots, Φ_m of degree at most e with constant coefficient 1 such that for every word u we have

$$\Phi_i(t) \sum_{s=0}^{\infty} f(w_1^{i_1} \cdots w_j^s \cdots w_m^{i_m} u) t^{i_1} \cdots t^{i_m}$$

is a polynomial in $K[t]$ of degree at most e . Let e be the maximum of the degrees of P, Φ_1, \dots, Φ_m . Now suppose that it is not the case that $\{f_{w_1^{i_1} \dots w_m^{i_m}} \mid 0 \leq i_1, \dots, i_m \leq e\}$ spans $\{f_{w_1^{i_1} \dots w_m^{i_m}} \mid i_1, \dots, i_m \geq 0\}$. Then there exist j_1, \dots, j_m with $j_i > e$ for some i such that $f_{w_1^{j_1} \dots w_m^{j_m}}$ is not in the span of \mathcal{S} . It is no loss of generality to assume that (j_1, \dots, j_m) has the property that if (j'_1, \dots, j'_m) satisfies:

- $j'_k \leq j_k$ for $1 \leq k \leq m$; and
- $(j'_1, \dots, j'_m) \neq (j_1, \dots, j_m)$,

then

$$f_{w_1^{j_1} \dots w_m^{j_m}} \in \{f_{w_1^{i_1} \dots w_m^{i_m}} \mid 0 \leq i_1, \dots, i_m \leq e\}.$$

By assumption $j_i > e$ and by the e -power property, we have

$$\sum_{k=0}^e c_k f_{w_1^{j_1} \dots w_i^{j_i-k} \dots w_m^{j_m}} = 0,$$

where c_k is the coefficient of x^k in Φ_i . Hence

$$f_{w_1^{j_1} \dots w_m^{j_m}} \in \text{Span}_K \{f_{w_1^{j_1} \dots w_i^{j_i-k} \dots w_m^{j_m}} \mid 1 \leq k \leq j_i\},$$

since Φ_i has constant coefficient 1. By minimality, we deduce that $f_{w_1^{j_1} \dots w_m^{j_m}}$ is in the span of $\{f_{w_1^{i_1} \dots w_m^{i_m}} \mid 0 \leq i_1, \dots, i_m \leq e\}$, a contradiction. The result follows. \square

Theorem 4.5.2 (Berstel-Reutenauer). *Let K be a field and let $f : \mathcal{A}^* \rightarrow K$. Then the following are equivalent:*

1. f is right k -regular;
2. f is left k -regular;
3. f has the d -shuffle and the d -power property for some d ;
4. f has the d -shuffle and the d -power property for all sufficiently large d .

Proof. From Propositions 4.2.5, 4.2.8, and 4.3.2, we have that if f is either right or left k -regular, then f has the d -shuffle and the d -power property for all sufficiently large d . Hence (1) and (2) both imply (4). Clearly (4) implies (3). Thus, it is sufficient to show that if f has the d -shuffle and the d -power property for some d , then f is both left and right regular. Suppose that f has the d -shuffle and the d -power property. We claim that the K -vector space $L(f)$ is finite-dimensional. To see this, notice that by Corollary 4.4.12, the vector space $L(f)$ is spanned by elements of the form $f_{w_1^{i_1} \dots w_m^{i_m}}$ with $m \leq d^5 |\mathcal{A}|^d$ and w_1, \dots, w_m having length at most d . Since f has the d -power property, by Lemma 4.5.1 for any words w_1, \dots, w_m of length at most d with $m \leq d^5 |\mathcal{A}|^d$, the K -vector space spanned by $\{f_{w_1^{i_1} \dots w_m^{i_m}} \mid i_1, \dots, i_m \geq 0\}$ is spanned by

$$\{f_{w_1^{i_1} \dots w_m^{i_m}} \mid 0 \leq i_1, \dots, i_m \leq d\}.$$

It now follows that $L(f)$ is in the K -span of the set

$$\bigcup_{m \leq d^5 |\mathcal{A}|^d} \bigcup_{\substack{\text{length}(w_j) \leq d, \\ 1 \leq j \leq m}} \{f_{w_1^{i_1} \dots w_m^{i_m}} \mid i_1, \dots, i_m \leq d\}.$$

Thus $L(f)$ is finite-dimensional and so f is left \mathcal{A} -regular. A similar argument shows that f is right \mathcal{A} -regular and hence (3) implies both (1) and (2). \square

As a result of the equivalence between left and right regularity, we drop the words left and right and talk only of \mathcal{A} -regular functions from now on.

Corollary 4.5.3. *Let $f : \mathcal{A}^* \rightarrow A$. Then the following are equivalent:*

1. f is k -automatic;
2. f has the d -shuffle and the d -power property for some d and has a finite range;
3. f has the d -shuffle and the d -power property for all sufficiently large d and has a finite range.

We make the following remark that follows from the proof of Theorem 4.5.2. This gives a bound on the dimension of the vector space in terms of the quantities d and e for which one has the d -shuffle and e -power property, which is undoubtedly far from optimal, but has the advantage of being explicit in terms of d and e .

Remark 4.5.4. If $|\mathcal{A}| \geq 2$ and $f : \mathcal{A}^* \rightarrow K$ has the d -shuffle property and the e -power property, then $L(f)$ and $R(f)$ are both at most N -dimensional where

$$N = (e + 1)^{d^5 |\mathcal{A}|^d} |\mathcal{A}|^{2d^6 |\mathcal{A}|^d}.$$

Proof. The proof of Theorem 4.5.2 shows that $L(f)$ is in the K -span of the set

$$\bigcup_{m \leq d^5 |\mathcal{A}|^d} \bigcup_{\substack{\text{length}(w_j) \leq d, \\ 1 \leq j \leq m}} \{f_{w_1^{i_1} \dots w_m^{i_m}} \mid i_1, \dots, i_m \leq e\}.$$

Since there are at most $|\mathcal{A}|^{d+1}$ words of length at most d , we see that the number of m -tuples of words of length at most d with $m \leq d^5 |\mathcal{A}|^d$ is at most

$$\sum_{m=1}^{d^5 |\mathcal{A}|^d} |\mathcal{A}|^{(d+1)m} \leq |\mathcal{A}|^{2d^6 |\mathcal{A}|^d}.$$

Now for each such m -tuple, we pick up a space of dimension at most $(e + 1)^m$ in our spanning set, and so the dimension of $L(f)$ is at most $(e + 1)^{d^5 |\mathcal{A}|^d} |\mathcal{A}|^{2d^6 |\mathcal{A}|^d}$. A similar argument works for $R(f)$. \square

4.6 Sandwich Functions

In this section we introduce a special type of function that is produced from a K -valued function on a free monoid, where K is a field; these functions will be called, for reasons that will soon become apparent, *sandwich functions*. We will then characterize all automatic sandwich functions.

Let $\mathcal{A} = \{x_1, \dots, x_k\}$ be a finite alphabet. We put a pure lexicographic order \preceq on \mathcal{A}^* by declaring that

$$x_1 < x_2 < \dots < x_k.$$

We note that this lexicographic order extends to right infinite words over the alphabet \mathcal{A} .

Let w_1 and w_2 be two (possibly right-infinite) words on \mathcal{A} with $w_1 \preceq w_2$. In dealing with right-infinite words, it will be convenient to use w^ω to denote the right-infinite word $www\omega\dots$. We define $f : \mathcal{A} \rightarrow \{0, 1\}$ by $f(w) = 1$ if $w_1 \preceq w \preceq w_2$ and $f(w) = 0$ otherwise. We call f a *sandwich function* since the words which get mapped to 1 are sandwiched between w_1 and w_2 .

Example 4.6.1. Let $\mathcal{A} = \{x_1, \dots, x_k\}$. Then the constant function 1 and the constant function 0 are both sandwich functions.

Proof. To get the constant function 1, take $w_1 = \varepsilon$ and take $w_2 = x_k^\omega$. To get the constant function 0, take $w_1 = w_2 = x_k^\omega$. \square

Example 4.6.2. Let $\mathcal{A} = \{x_1, \dots, x_k\}$. Then the function f which sends words beginning with x_1 to 1 and all other words to 0 is a sandwich function.

Proof. Take $w_1 = x_1$ and take $w_2 = x_1 x_k^\omega$. Then the sandwich function given by these words is f . \square

We now characterize \mathcal{A} -automatic sandwich functions. To do this we first prove some basic results. As notation for the following lemma, we introduce the function χ which inputs a statement and outputs 1 if the statement is true and 0 if the statement is false.

Lemma 4.6.3. *Let w be a (possibly right-infinite) word on a finite alphabet \mathcal{A} which is either finite or eventually periodic. Then the functions $f(u) := \chi(u \preceq w)$ and $g(u) = \chi(u \succeq w)$ are both \mathcal{A} -automatic.*

Proof. If w is finite, notice that if v is a word whose length is greater than the length of w , then

$$f_v(u) = f(vu) = \chi(vu < w) = \chi(v \preceq w).$$

Hence the vector space $L(f)$ is contained in the space spanned by the constant function 1 and the functions $\{f_v \mid \text{length}(v) \leq \text{length}(w)\}$. Thus $L(f)$ is finite-dimensional. A similar argument shows that $L(g)$ is finite-dimensional. If w is eventually periodic, then we can write $w = w_1 w_2^\omega$. Notice that if v is not an initial factor of w , then

$$f_v(u) = f(vu) = \chi(vu \preceq w) = \chi(v \preceq w),$$

which is a constant function. If v is an initial factor of w of length at least $\text{length}(w_1) + 2\text{length}(w_2)$, then $f_v = f_{v'}$, where v' is an initial factor of v obtained by removing the last $\text{length}(w_2)$ letters from v . Hence $L(f)$ is again contained in a finite-dimensional vector space and hence must be finite-dimensional. A similar argument works for $L(g)$. \square

We need a lemma about monotonic subsequences.

Lemma 4.6.4. *Let w be a right-infinite word on a finite alphabet \mathcal{A} which is not eventually periodic. Then for any $d \geq 1$, there exist finite words u_1, \dots, u_d such that:*

- $\text{length}(u_1) < \text{length}(u_2) < \dots < \text{length}(u_d)$;
- the sequence u_1, \dots, u_d is monotonic with respect to $<$;
- for $2 \leq i \leq d$, there exists a word $u'_i \neq u_i$ of the same length as u_i which does not have u_{i-1} as an initial factor and has the property that u_{i-1}, u'_i, u_i is monotonic;
- $u_1 u_2 \dots u_d$ is a factor of w .

Proof. For $i \geq 0$ define w_i to be the right-infinite word obtained by removing the first i letters from w . Since w is not eventually periodic, the words w_0, w_1, \dots are all distinct. It follows that they are totally ordered by $<$. It follows that there is a monotonic (with respect to $<$) subsequence w_{i_1}, w_{i_2}, \dots of w_0, w_1, \dots . Without loss of generality $w_{i_1} < w_{i_2} < \dots$. Let $v_j = w_{i_j}$ for $j \geq 1$. Pick $k_1 = 1$. Since $v_1 < v_2$ and v_2 is not eventually periodic, there exists some number m_1 such that the first m_1 letters of v_1 differ from the first m_1 letters of v_2 and the first m_1 letters of v_2 differ with the first m_1 letters of v_3 . Choose $j_1 \geq m_1$ such that if we remove the first j_1 letters of v_1 we obtain a word v_{k_2} with $k_2 \geq 3$. Define u_1 to be the first j_1 letters of v_1 . Now $v_{k_2} < v_{k_2+1}$ and hence there is some m_2 such that the first m_2 letters of v_{k_2} differ from the first m_2 letters of v_{k_2+1} and the first m_2 letters of v_{k_2+1} differ from the first m_2 letters of v_{k_2+2} . As before, we choose $j_2 > \max(j_1, m_2)$ such that if we remove the first j_2 letters of v_{k_2} we obtain some word v_{k_3} with $k_3 \geq k_2 + 2$. We define u_2 to be the first j_2 letters of v_{k_2} and u'_2 to be the first j_2 letters of v_2 . Notice that $u_1 < u'_2 < u_2$ and $u'_2 \neq u_2$, and it cannot have u_1 as an initial factor of u_1 . Continuing in this manner, we see that we can write

$$w = uu_1u_2u_3 \dots$$

with u some initial factor and words $u_1, u_2, \dots, u_d, u'_1, \dots, u'_d$ satisfying the conditions in the statement of the lemma. \square

Theorem 4.6.5. *Let w_1 and w_2 be (possibly right-infinite) words on a finite alphabet. Then the sandwich function f corresponding to w_1 and w_2 is \mathcal{A} -automatic if and only if w_1 and w_2 are both either finite or ultimately periodic.*

Proof. Suppose that w_1 is neither finite nor ultimately periodic. Write $w_1 = w'_1 w''_1$ where w'_1 is a finite word which is not an initial factor of w_2 and w''_1 is a right-infinite

word. Then w'_1 is not eventually periodic and hence by Lemma 4.6.4, for any d we can find words u, u_1, u_2, \dots, u_d such that $w_1 = w'_1 u u_1 \cdots u_d u'$ for some right-infinite word u' ; u_1, \dots, u_d is monotonic with respect to $<$;

$$\text{length}(u_1) < \cdots < \text{length}(u_d);$$

and for $i \geq 2$ there exist words u'_i with u_{i-1}, u'_i, u_i monotonic with $u'_i \neq u_i$, $\text{length}(u'_i) = \text{length}(u_i)$ and u_{i-1} not an initial factor of u'_i . We have two cases:

Case I $u_1 < u_2 < \cdots < u_d$.

In this case take $v_d = u'_d$ and for $1 \leq i \leq d - 1$ take $v_i = u_i$. Consider

$$\text{Shuf}_d(f; w'_1 u, v_1, \dots, v_{d-1}, v_d, \epsilon).$$

Notice that since $w'_1 u$ is lexicographically less than the initial factor of w_2 of the same length and since $v_1 < v_2 < \cdots < v_d$, we have that

$$w_1 < w'_1 u v_{\sigma(1)} \cdots v_{\sigma(d)} < w_2$$

unless σ is the identity. When σ is the identity, we have $w_1 \not\leq w'_1 u v_1 \cdots v_d$ and hence

$$\sum_{\sigma \in S_d} \text{sgn}(\sigma) f(w'_1 u v_{\sigma(1)} \cdots v_{\sigma(d)}) \equiv 1 \pmod{2}.$$

Thus f cannot have the d -shuffle property. Since d is arbitrary we conclude that f is not \mathcal{A} -automatic.

Case II: $u_1 > u_2 > \cdots > u_d$.

In this case take $v_d = u'_d$ and for $1 \leq i \leq d - 1$ take $v_i = u_i$. In this case we have

$$f(w'_1 u v_{\sigma(1)} \cdots v_{\sigma(d)}) = 1 \quad \text{if and only if } \sigma \neq \text{id}.$$

Thus the result again holds in this case.

In either case, the d -shuffle property fails to hold for any d , and so our function cannot be \mathcal{A} -regular. A similar argument shows that f is not \mathcal{A} -regular if w_2 is right-infinite and not eventually periodic.

Next consider what happens if w_1 and w_2 are finite or eventually periodic. Then by the lemma

$$f(w) = \chi(w > w_1) \chi(w < w_2)$$

and hence f is the coordinate-wise product of two \mathcal{A} -automatic sequences. It follows that f is automatic. □

4.7 Applications

4.7.1 The Logarithm and Automaticity

We now give an application of our description of automatic sandwich functions to answer a question of Allouche and Shallit. We note that this question had been answered via a different method by Yossi [425] in 2008.

Proposition 4.7.1. *Let $\alpha \in \mathbb{R}$. Then the sequence given by $f(0) = 1$ and*

$$f(n) = \lfloor \log_k n + \alpha \rfloor \quad \text{for } n \geq 1$$

is k -regular if and only if k^α is rational.

Proof. We may assume that $\alpha \in [0, 1)$. It is easy to verify that the function $g(0) = 0$ and $g(n) = \lfloor \log_k n \rfloor$ for $n \geq 1$ is k -regular. Let us consider the function $f(n) - g(n)$. This function takes values in $\{0, 1\}$ and is 1 at an integer $n \geq 1$ if and only if there is some integer m such that

$$\log_k n + \alpha \geq m > \log_k n.$$

Equivalently, we must have

$$k^\alpha n \geq k^m \geq n.$$

Write

$$k^{-\alpha} = \sum_{i=1}^{\infty} a_i/k^i$$

with $a_i \in \{1, 2, 3, \dots, k\}$. Let $w_1 = \varepsilon$ and let w_2 be the right-infinite word $a_1 a_2 a_3 \dots$. Let w be the word in $\{1, 2, \dots, k\}$ which corresponds to n using the correspondence described in equation (4.1). Then if we regard $h := f - g$ as a function on $\{1, 2, \dots, k\}^*$, then $h(w) = 1$ if and only if $w \leq w_2$, where $<$ is the lexicographic order induced by taking $1 < 2 < \dots < k$; equivalently, $h(w) = 1$ if $w_1 \leq w \leq w_2$ and is 0 otherwise. Hence h is a sandwich function. Notice that w_2 is a right-infinite word which is eventually periodic if and only if k^α is rational. Thus h is k -regular if and only if k^α is rational. Since $f = g + h$ and g is k -regular, we see that f is k -regular if and only if h is k -regular. The result now follows. \square

Allouche and Shallit [14] ask whether the sequence $\lfloor \log_2 n + \frac{1}{2} \rfloor$ is 2-automatic; since $\sqrt{2}$ is irrational, we deduce that it is not.

4.7.2 The 2-Adic Behavior of the Logarithm

Allouche and Shallit [14, Section 16.7, Q. 4] ask whether the sequence

$$f(n) = \min_{i \geq n+1} i - v_2(i)$$

is 2-regular, where $v_2(i)$ is the 2-adic valuation; that is, $v_2(i)$ satisfies $2^{v_2(i)}|i$ but $2^{v_2(i)+1} \nmid i$. We show that it is not. Here we use the fact that right regularity and left regularity are equivalent properties to answer a question which is difficult to handle using the traditional definition of regularity in terms of right regularity, but which is easily handled if one uses the notion of left regularity.

Proposition 4.7.2. *Let*

$$f(n) = \min_{i \geq n+1} i - v_2(i).$$

Then $f(n)$ is not 2-regular.

Proof. Let $\mathcal{A} = \{0, 1\}$. Let $w_i = 1^i 0 \in \mathcal{A}^*$. Consider the subspace of $L(f)$ generated by $\{f_{w_i} \mid i \geq 0\}$. Suppose that f is \mathcal{A} -regular. Then since $L(f)$ is finite-dimensional, there exists some $m \geq 3$ such that this subspace is spanned by $\{f_{w_i} \mid i \leq m\}$. Let $d = 2^{2^m}$. By assumption, there exist integers c_0, \dots, c_m such that

$$f_{w_d} = \sum_{i=0}^m c_i f_{w_i}.$$

In particular, we have

$$\begin{aligned} f(w_d w_j) &= f_{w_d}(w_j) \\ &= \sum_{i=0}^m c_i f_{w_i}(w_j) \\ &= \sum_{i=0}^m c_i f(w_i w_j) \end{aligned} \tag{4.7}$$

for all $j \geq 0$. Observe that for $[w_i w_j]_2 = 2^{i+j+2} - 2 - 2^j$. Hence

$$f(w_i w_j) = \begin{cases} 2^{i+j+2} - 2^j - j & \text{if } 2^j \geq i + 2; \\ 2^{i+j+2} - i - j - 2 & \text{if } 2^j \leq i + 2. \end{cases}$$

In particular, if $m \leq j \leq 2^m$, we have $f(w_d w_j) = 2^{d+j+2} - d - j - 2$ and $f(w_i w_j) = 2^{i+j+2} - 2^j - j$ for $1 \leq i \leq m$. Using equation (4.7), we see that for $m \leq j \leq 2^m$ we have

$$\begin{aligned} 2^{d+j+2} - d - j - 2 &= f(w_d w_j) \\ &= \sum_{i=0}^m c_i f(w_i w_j) \\ &= \sum_{i=0}^m c_i (2^{i+j+2} - 2^j - j). \end{aligned}$$

Simplifying, we deduce

$$2^j \left(2^{d+2} - \sum_{i=0}^m c_i (2^{i+2} - 1) \right) + j \left(-1 + \sum_{i=0}^m c_i \right) = d + 2,$$

for $m \leq j \leq 2^m$. Let

$$A = 2^{d+2} - \sum_{i=0}^m c_i (2^{i+2} - 1)$$

and let

$$B = -1 + \sum_{i=0}^m c_i.$$

Then we have

$$2^j A + jB = d + 2,$$

for $m \leq j \leq 2^m$. Consider the function $G(x) = 2^x A + xB - (d + 2)$. We have $G(m) = G(m + 1) = \dots = G(2^m) = 0$ and hence by Rolle's theorem $G'(x) = 2^x A \log 2 + B$ must have at least $2^m - m \geq 2$ real zeros. This implies that $A = B = 0$ since the function $G'(x)$ is monotonic. Then the fact that $G(m) = 0$ along with $A = B = 0$ now implies that $d + 2 = 0$, which is a contradiction. It follows that $f(n)$ is not a 2-regular function. \square

4.7.3 Nim Sums and Nim Products

In this subsection we consider the 2-regularity of sequences constructed using nim sums and nim products. Given nonnegative integers n and m , we define the *nim sum*, $n \oplus m$, of n and m as follows. We write $n = [a_d \cdots a_0]_2$ with $a_i \in \{0, 1\}$ and $m = [b_d \cdots b_0]_2$ with $b_i \in \{0, 1\}$, where we may pad the binary expansion of either a or b with zeros at the beginning to ensure that they have the same length. We then define

$$n \oplus m := [(a_d + b_d \bmod 2) \cdots (a_0 + b_0 \bmod 2)]_2.$$

For example, if $m = 12$ and $n = 21$, then

$$m \oplus n = [01100]_2 \oplus [10101]_2 = [11001]_2 = 25.$$

The *nim product*, \otimes , is defined as follows:

$$2^{2^a} \otimes 2^{2^b} = \begin{cases} 2^{2^a} \cdot 2^{2^b} & \text{if } a \neq b; \\ 3 \cdot 2^{2^a-1} & \text{if } a = b. \end{cases} \quad (4.8)$$

The product is then defined for all pairs of natural numbers using associativity and distributivity. For example,

$$\begin{aligned} 8 \otimes 3 &= (2^{2^1} \otimes 2^{2^0}) \otimes (2^{2^0} \oplus 1) \\ &= 2^{2^1} \otimes (2^{2^0} \otimes 2^{2^0}) \oplus (2^{2^1} \otimes 2^{2^0}) \\ &= 2^{2^1} \otimes 3 \oplus 8 \\ &= 2^{2^1} \otimes (2^{2^0} \oplus 1) \oplus 8 \\ &= 8 \oplus 4 \oplus 8 \\ &= 4. \end{aligned}$$

The nim sum and nim products have the following properties, which can be found in Conway [162, Chap. 6]:

- $2^{2^a} \otimes x = 2^{2^a} x$ for $0 \leq x < 2^{2^a}$;
- the set of nonnegative numbers less than 2^{2^a} is a field under \oplus and \otimes .

Allouche and Shallit [14, Section 16.7, Q. 5,6] ask the following questions:

- Is the nim sum of two 2-regular sequences a 2-regular sequence?
- Is the sequence $\{n \otimes n\}$ a 2-regular sequence?

We show the answer to these questions is “no.” In fact, Allouche and Shallit ask questions involving 2-dimensional arrays of numbers, but negative answers to the

above questions imply negative answers to their questions about arrays. This time, we use the power property and show that it fails to hold.

Lemma 4.7.3. *Let i be a nonnegative integer. Then $2^i \otimes 2^i = 3 \cdot 2^{i-1}$ if and only if i is a power of 2.*

Proof. If i is a power of 2, then $2^i \otimes 2^i = 3 \cdot 2^{i-1}$ by the definition of the nim product. Next suppose that i is not a power of 2. Then there is some a such that $i = 2^a + j$ with $0 < j < 2^a$. Then

$$\begin{aligned} 2^i \otimes 2^i &= (2^{2^a} \otimes 2^j) \otimes (2^{2^a} \otimes 2^j) \\ &= (2^{2^a} \otimes 2^{2^a}) \otimes (2^j \otimes 2^j) \\ &= (2^{2^a} \oplus 2^{2^a-1}) \otimes (2^j \otimes 2^j) \\ &= (2^{2^a} \otimes (2^j \otimes 2^j)) \oplus (2^{2^a-1} \otimes (2^j \otimes 2^j)) \\ &= (2^{2^a} \cdot (2^j \otimes 2^j)) \oplus (2^{2^a-1} \otimes (2^j \otimes 2^j)), \end{aligned}$$

where we are using the two facts from Conway mentioned above to obtain these equalities.

Observe that if $2^i \otimes 2^i = 3 \cdot 2^{i-1}$, then the binary expansion of $2^i \otimes 2^i$ cannot have any 1's appearing in the $2^a + j - 1$ least significant digits. In particular it must have 0's appearing in the 2^a least significant digits. Since $2^{2^a} \otimes (2^j \otimes 2^j)$ has this property, $2^{2^a-1} \otimes (2^j \otimes 2^j)$ must also have this property if $2^i \otimes 2^i = 3 \cdot 2^{i-1}$, as the nim sum of these two numbers is $2^i \otimes 2^i$. But since $\{0, 1, \dots, 2^{2^a} - 1\}$ is a field under \otimes and \oplus , we see that $2^{2^a-1} \otimes (2^j \otimes 2^j)$ is a nonzero number less than 2^{2^a} . We conclude that $2^i \otimes 2^i \neq 3 \cdot 2^{i-1}$. \square

Proposition 4.7.4. *The sequence $\{m \otimes m\}$ is not 2-regular.*

Proof. To do this, we show that the sequence does not have the power property. It is sufficient to show that the power series

$$F(x) = \sum_{i=0}^{\infty} (2^i \otimes 2^i) x^i$$

is not rational. Suppose that $F(x)$ is rational. Then

$$G(x) = F(x) - \frac{3}{2}(1-2x)^{-1} = \sum_{i=0}^{\infty} (2^i \otimes 2^i - 3 \cdot 2^{i-1}) x^i$$

must also be rational. Notice that by Lemma 4.7.3 the coefficient of x^n is zero in $G(x)$ if and only if n is a power of 2. It follows from the Skolem-Mahler-Lech theorem (see [282]) that $G(x)$ is not rational. We conclude that F is not rational and so $\{n \otimes n\}$ is not 2-regular. \square

Proposition 4.7.5. *There exist 2-regular sequences $f(n)$ and $g(n)$ such that $f(n) \oplus g(n)$ is not 2-regular.*

Proof. Let $f(n)$ be defined by

$$f(n) = \begin{cases} \frac{4^m-1}{3} & \text{if } n = 2^m \\ 0 & \text{otherwise.} \end{cases}$$

Then it is easy to check that $f(n)$ is 2-shuffled and has the power property. Hence f is 2-regular. Let

$$g(n) = \begin{cases} m & \text{if } n = 2^m \\ 0 & \text{otherwise.} \end{cases}$$

Then $g(n)$ is 2-regular. Observe that $f(2^m) \oplus g(2^m) = f(2^m) - g(2^m)$ if and only if the binary expansion of m has 0's in all the even positions (beginning the count from the least significant digit). Suppose that $f(n) \oplus g(n)$ is 2-regular. Then the power property gives that the sequence

$$f(2^m) \oplus g(2^m)$$

must satisfy a linear recurrence. If this is the case, then by the Skolem-Mahler-Lech theorem (see [282]), the set of m such that $f(2^m) \oplus g(2^m) = \frac{4^m-1}{3} - m$ must contain an infinite arithmetic progression. But

$$\#\{m \leq 4^N \mid f(2^m) \oplus g(2^m) = f(2^m) - g(2^m)\} \leq 2^N.$$

In other words, the density of the set of m such that $f(2^m) \oplus g(2^m) = \frac{4^m-1}{3} - m$ is 0. But this is a contradiction, since by the Skolem-Mahler-Lech theorem, the density must be positive. It follows that $f \oplus g$ is not 2-regular. \square

We note that if $f(n)$ and $g(n)$ are k -automatic, then both $f(n) \otimes g(n)$ and $f(n) \oplus g(n)$ are k -automatic; this follows easily from the fact that k -automatic sequences assume only finitely many different values.

4.8 Shuffled Sequences

In this section we develop the basic properties of sequences possessing the shuffle property. We will find it more useful to work with abelian groups rather than fields in obtaining some of our closure properties, so we shall adopt this point of view in this section.

Definition 4.8.1. Let \mathcal{A} be a finite alphabet and let A be an abelian group. We say that $f : \mathcal{A}^* \rightarrow A$ is an \mathcal{A} -shuffled sequence if f has the d -shuffle property for some

positive integer d . In the case that $\mathcal{A} = \{1, 2, \dots, k\}$, we say that f is k -shuffled. We point out that being k -shuffled is not the same as having the k -shuffle property; when we use the word *shuffled*, the k is making reference to the underlying alphabet and when we use the word *shuffle*, the k is making reference to the shuffle identity satisfied by a function.

In the case that no confusion will arise, we will drop the alphabet and refer to a sequence as being a *shuffled sequence*. We note that the shuffle property, although originally defined in the case where our abelian group is a field (with group law given by addition), makes sense in this more general setting.

We have the following containments:

$$\begin{aligned} \mathcal{A}\text{-automatic sequences} &\subseteq \mathcal{A}\text{-regular sequences} \\ &\subseteq \mathcal{A}\text{-shuffled sequences.} \end{aligned}$$

It is well known [14, Theorem 16.3.1] that a k -regular sequence $f : \mathbb{N} \rightarrow \mathbb{Z}$ has polynomially bounded growth in the sense that there is some $d > 0$ such that $|f(n)| \leq n^d$ for all sufficiently large n . The following example shows that no such growth restriction on shuffled sequences holds.

Example 4.8.2. A k -shuffled sequence can have arbitrarily rapid growth.

Proof. Let a_1, a_2, \dots be a sequence of integers and let $f : \{1, 2\}^* \rightarrow \mathbb{C}$ be defined by

$$f(w) = \begin{cases} 0 & \text{if 2 appears in } w \\ a_d & \text{if } w = 1^d. \end{cases}$$

Then $f(w)$ has the 2-shuffle property and hence is a shuffled sequence. Since the a_d are arbitrary, we see that shuffled sequences can have arbitrarily rapid growth. \square

We now give some closure properties for shuffled sequences.

Proposition 4.8.3. *Let A and B be two abelian groups, and let \mathcal{A} be a finite alphabet. If $f : \mathcal{A}^* \rightarrow A$ and $g : \mathcal{A}^* \rightarrow B$ are shuffled sequences, then so are $(f \oplus g) : \mathcal{A}^* \rightarrow A \oplus B$ and $f \otimes g : \mathcal{A}^* \rightarrow A \otimes_{\mathbb{Z}} B$. (Here $(f \oplus g)(w) = f(w) \oplus g(w)$ and $f \otimes g(w) = f(w) \otimes g(w)$.)*

Proof. Create the ideals I_1 and I_2 in $\mathbb{Z}\{x_1, \dots, x_k\}$ as follows. Let I_1 be the set of elements of the form $c_1w_1 + \dots + c_dw_d$ such that

$$c_1f(ww_1w') + \dots + c_df(ww_dw') = 0$$

for all $w, w' \in \mathcal{A}^*$. Similarly, define I_2 to be the set of elements of the form $c_1w_1 + \dots + c_dw_d$ with

$$c_1g(ww_1w') + \dots + c_dg(ww_dw') = 0$$

for all $w, w' \in \mathcal{A}^*$. Then I_1 and I_2 are ideals and $R_1 := \mathbb{Z}\{x_1, \dots, x_k\}/I_1$ and $R_2 := \mathbb{Z}\{x_1, \dots, x_k\}/I_2$ both satisfy the identity S_d and hence satisfy polynomial identities. It follows from Regev's theorem [498], that $R_1 \otimes_{\mathbb{Z}} R_2$ also satisfies a polynomial identity. Since it is finitely generated as a \mathbb{Z} -algebra, it satisfies the standard identity S_m for some $m \geq 0$. Now suppose that the image of $\sum c_{i,j} w_i \otimes w_j$ is zero in $R_1 \otimes_{\mathbb{Z}} R_2$. Then by construction, we have

$$\sum c_{i,j} f(w w_i w') \otimes g(w w_j w') = 0$$

for all words w, w' . Let $R \subseteq R_1 \otimes_{\mathbb{Z}} R_2$ be the subalgebra generated by the images of $x_1 \otimes x_1, \dots, x_k \otimes x_k$. Then R must also satisfy the identity S_m , since it is a subring of $R_1 \otimes R_2$. Consequently, the sequence $f \otimes g$ must have the m -shuffle property. \square

Remark 4.8.4. Let \mathcal{A} be a finite alphabet and let A and B be abelian groups. If $f : \mathcal{A}^* \rightarrow A$ is a shuffled sequence and $\phi : A \rightarrow B$ is a homomorphism of abelian groups, then $\phi \circ f : \mathcal{A}^* \rightarrow B$ is a shuffled sequence.

Corollary 4.8.5. *Let \mathcal{A} be a finite group and let A be an abelian group. If $f, g : \mathcal{A} \rightarrow A$ are shuffled, then $(f + g)$ is shuffled. If, in addition, A is a ring then $f \cdot g$ is shuffled.*

Proof. For the first part, use Proposition 4.8.3 and Remark 4.8.4, taking the homomorphism $\phi : A \oplus A \rightarrow A$ given by $\phi(a, a') = a + a'$. For the product, again use Proposition 4.8.3 and Remark 4.8.4, this time taking the homomorphism $\phi : A \otimes A \rightarrow A$ given by $\phi(a \otimes a') = aa'$. \square

Sometimes it is easier to verify that a sequence is shuffled by showing that an identity holds other than the standard identity of the shuffle property holds. We make this more precise in the next theorem.

Theorem 4.8.6. *Let n be a nonnegative integer. Suppose that there exist integers $\{c_\sigma \mid \sigma \in S_n\}$, at least one of which is equal to 1, such that*

$$\sum_{\sigma \in S_n} c_\sigma f(w w_{\sigma(1)} \cdots w_{\sigma(n)} w') = 0.$$

for all words $w, w_1, \dots, w_n, w' \in \mathcal{A}^*$. Then f is shuffled.

Proof. Write $\mathcal{A} = \{x_1, \dots, x_k\}$. We define an ideal I in the algebra $\mathbb{Z}\{x_1, \dots, x_k\}$ as follows. We declare that $\sum a_w w \in I$ if

$$\sum a_w f(w' w w'') = 0$$

for all $w', w'' \in \mathcal{A}^*$. Notice that $R = \mathbb{Z}\{x_1, \dots, x_k\}/I$ satisfies a polynomial identity, as it satisfies the multilinear, homogeneous polynomial identity $\sum_{\sigma \in S_n} c_\sigma t_{\sigma(1)} \cdots t_{\sigma(n)}$. Since R is finitely generated, it satisfies a standard identity S_m for some m (cf. Theorem 4.4.6 and see, also, Braun [105]). Consequently,

$$\sum_{\sigma \in S_m} \text{sgn}(\sigma) w_{\sigma(1)} \cdots w_{\sigma(m)} \in I$$

for all $(w_1, \dots, w_m) \in (\mathcal{A}^*)^m$. Hence

$$\sum_{\sigma \in S_m} \text{sgn}(\sigma) f(w w_{\sigma(1)} \cdots w_{\sigma(m)} w') = 0$$

for all $w, w_1, \dots, w_m, w' \in \mathcal{A}^*$. \square

Let \mathcal{A} be a finite alphabet and let R be a ring. The set of maps $f : \mathcal{A}^* \rightarrow R$ has a multiplication defined as follows: Given $x_1, \dots, x_d \in \mathcal{A}$, we define

$$f \star g(x_1 \cdots x_d) := \sum_{i=0}^d f(x_1 \cdots x_i) g(x_{i+1} \cdots x_d), \quad (4.9)$$

where $f(x_1 \cdots x_i)$ is taken to mean $f(\varepsilon)$ when $i = 0$ and we take $g(x_{i+1} \cdots x_d)$ to be $g(\varepsilon)$ when $i = d$. This product, along with ordinary sum, turns the set of maps $f : \mathcal{A}^* \rightarrow R$ into an associative ring [77]. We show that the shuffled sequences form a subalgebra.

Proposition 4.8.7. *If $f : \mathcal{A}^* \rightarrow A$ and $g : \mathcal{A}^* \rightarrow A$ are shuffled, then $f \star g$ is shuffled.*

Proof. We may pick d such that f and g both have the d -shuffled property. We claim that if $w_1, \dots, w_d, u_1, \dots, u_d, w, w' \in \mathcal{A}^*$, then

$$\sum_{\sigma \in S_d} \sum_{\tau \in S_d} \text{sgn}(\sigma) \text{sgn}(\tau) (f \star g)(w w_{\sigma(1)} \cdots w_{\sigma(d)} u_{\tau(1)} \cdots u_{\tau(d)} w') = 0.$$

To see this, let

$$v(\sigma, \tau) = w w_{\sigma(1)} \cdots w_{\sigma(d)} u_{\tau(1)} \cdots u_{\tau(d)} w'.$$

Then

$$\sum_{\sigma, \tau \in S_d} \text{sgn}(\sigma) \text{sgn}(\tau) (f \star g)(v(\sigma, \tau)) = \sum_{\sigma, \tau \in S_d} \sum_{\substack{v_1 v_2 = \\ v(\sigma, \tau)}} \text{sgn}(\sigma) \text{sgn}(\tau) f(v_1) g(v_2).$$

If $v_1 v_2 = v(\sigma, \tau)$ then either $v_1 = v_1(\sigma, \tau)$ contains $w w_{\sigma(1)} \cdots w_{\sigma(d)}$ as an initial factor, or $v_2 = v_2(\sigma, \tau)$ contains $u_{\tau(1)} \cdots u_{\tau(d)} w'$ as a terminal factor (or both). Notice that if $v_1 = w w_{\sigma(1)} \cdots w_{\sigma(d)} v'_1$, then v'_1 and v_2 depend only on τ . Hence

$$\sum_{\sigma \in S_d} \sum_{\tau \in S_d} \text{sgn}(\sigma) \text{sgn}(\tau) f(v_1) g(v_2)$$

$$\begin{aligned}
 &= \sum_{\tau \in S_d} \sum_{\sigma \in S_d} \operatorname{sgn}(\sigma) \operatorname{sgn}(\tau) f(w w_{\sigma(1)} \cdots w_{\sigma(d)} v'_1) g(v_2) \\
 &= \sum_{\tau \in S_d} \operatorname{sgn}(\tau) g(v_2) \left(\sum_{\sigma \in S_d} \operatorname{sgn}(\sigma) f(w w_{\sigma(1)} \cdots w_{\sigma(d)} v'_1) \right) \\
 &= 0,
 \end{aligned}$$

since f has the d -shuffled property. Similarly, if v_2 contains the terminal factor $u_{\tau(1)} \cdots u_{\tau(d)} w'$, then the fact that g has the d -shuffled property guarantees that

$$\sum_{\sigma \in S_d} \sum_{\tau \in S_d} f(v_1) g(v_2) = 0.$$

Observe that $S_d \times S_d$ embeds in S_{2d} by taking an ordered pair (σ, τ) and letting σ act on $\{1, 2, \dots, d\}$ and letting τ act on $\{d + 1, \dots, 2d\}$. Given $\mu \in S_{2d}$, we define $c_\mu = \operatorname{sgn}(\sigma) \operatorname{sgn}(\tau)$ if μ corresponds to (σ, τ) under this inclusion and we take $c_\mu = 0$ otherwise. Then from Theorem 4.8.6, we see that $f \star g$ is shuffled. \square

We let \mathcal{S}_k denote the collection of k -shuffled sequences taking values in \mathbb{C} . Then we have just shown that $(\mathcal{S}_k, +, \star)$ is a \mathbb{C} -algebra. The following proposition shows that in some sense this algebra is very large.

Proposition 4.8.8. *The algebra $(\mathcal{S}_k, +, \star)$ contains a copy of the free \mathbb{C} -algebra on infinitely many generators.*

Proof. Since the free algebra on two generators contains a copy of the free algebra on infinitely many generators, it is sufficient to do show we contain a copy of the free algebra on two generators. Let $f_0 : \{1, 2, \dots, k\}^* \rightarrow \mathbb{C}$ be defined to be 1 on words of the form 12^i and be defined to be 0 on all other words. Let $f_1 : \{1, 2, \dots, k\}^* \rightarrow \mathbb{C}$ be defined to be i on the word 12^i and to be 0 on words not of the form 12^i . Then it is easy to check that f_0 and f_1 are k -shuffled sequences. We claim that f_0 and f_1 generate a free algebra. Suppose that

$$G := \sum_{k=1}^d \sum_{(i_1, \dots, i_k) \in \{0,1\}^k} \alpha_{i_1, \dots, i_k} f_{i_1} \star f_{i_2} \star \cdots \star f_{i_k} = 0$$

for some d with $\alpha_{i_1, \dots, i_d} \neq 0$ for some $(i_1, \dots, i_d) \in \{0, 1\}^d$. Notice that if w is a word of the form $12^{a_1} 12^{a_2} \cdots 12^{a_d}$ then $f_{i_1} \star \cdots \star f_{i_k}(w) = 0$ for $k < d$. Hence

$$\begin{aligned}
 &G(12^{a_1} 12^{a_2} \cdots 12^{a_d}) \\
 &= \sum_{(i_1, \dots, i_d) \in \{0,1\}^d} \alpha_{i_1, \dots, i_d} f_{i_1} \star f_{i_2} \star \cdots \star f_{i_d}(12^{a_1} 12^{a_2} \cdots 12^{a_d})
 \end{aligned}$$

$$= \sum_{(i_1, \dots, i_d) \in \{0,1\}^d} \alpha_{i_1, \dots, i_d} a_1^{i_1} \cdots a_d^{i_d}.$$

by assumption G is identically 0 and hence the polynomial

$$H(x_1, \dots, x_d) := \sum_{(i_1, \dots, i_d) \in \{0,1\}^d} \alpha_{i_1, \dots, i_d} x_1^{i_1} \cdots x_d^{i_d}$$

vanishes on all points $(x_1, \dots, x_d) \in \mathbb{N}^d$. But this implies that H is the zero polynomial, which contradicts the fact that $\alpha_{i_1, \dots, i_d} \neq 0$ for some $(i_1, \dots, i_d) \in \{0, 1\}^d$. We conclude that the algebra generated by f_0 and f_1 is free. This completes the proof. \square

4.9 Open Problems and Concluding Remarks

One of the fundamental theorems from the theory of automatic sequences is Cobham's theorem. Two integers p and q are multiplicatively independent if $p^a \neq q^b$ for $(a, b) \neq (0, 0)$. Cobham's theorem [14, Chapter 11] states that if a sequence is p -automatic and q -automatic and p and q are multiplicatively independent, then the sequence is eventually periodic. Given a sequence $f(n)$ taking values in an abelian group, by the correspondence described in item 4.1, it makes sense to talk about the sequence being a k -shuffled sequence.

Question 4.9.1. Suppose a \mathbb{Z} -valued sequence $f(n)$ is both p -shuffled and q -shuffled for two multiplicatively independent integers p and q . What can be said about $f(n)$? For instance, does $f(n)$ satisfy a linear recurrence?

Another question comes from looking at closure properties. In Section 4.8 we showed that shuffled sequences are closed under ordinary products and under the \star product. We now ask if shuffled sequences are closed under Cauchy products.

Question 4.9.2. Given integer-valued sequences $f(n)$ and $g(n)$ that are k -shuffled, is the Cauchy product of $f(n)$ and $g(n)$ also k -shuffled?

In Section 4.8 we showed that the set \mathcal{S}_k of k -shuffled sequences taking values in \mathbb{C} forms a \mathbb{C} -algebra under the \star product.

Question 4.9.3. Can one find nice generating sets for the \mathbb{C} -algebra $(\mathcal{S}_k, +, \star)$?

Question 4.9.4. Can one extend the notion of shuffled sequences to nonconstant length substitutions? See, for example, Shallit [542] and Allouche, Scheicher, and Tichy [13].

Two final questions we pose come from the nim sum and nim product. In Section 4.7 we studied sequences defined using the nim sum and the nim product. In each example, we showed that the power property failed to hold. We did not show, however, that the shuffle property fails to hold.

Question 4.9.5. Let \oplus and \otimes denote, respectively, the nim sum and the nim product. Is the sequence $\{m \otimes m\}$ a 2-shuffled sequence? If $f(n)$ and $g(n)$ are 2-shuffled sequences, is $f(n) \oplus g(n)$ also a 2-shuffled sequence?

Question 4.9.6. Does $\{2^n \oplus 3^n\}$ satisfy a linear recurrence? We note that this is related to Mahler's study of Z -numbers [406].

We note that throughout we have been looking at sequences taking values in a field or occasionally in an abelian group. In fact, we can work more generally by using a commutative ring C and look at sequences taking values in a C -module. Allouche and Shallit [16] define the more general notion of a (C, k) -regular sequence. We note that since Shirshov's height theorem is over a more general base ring C , all the results in this paper which relate to k -regular sequences have analogues in this more general context of (C, k) -regular sequences.

Acknowledgements I thank Jean-Paul Allouche and Jeffrey Shallit for many helpful comments. I also thank Jean-Paul Allouche for raising Question 4.9.4.