Pumping Lemmata for Recognizable Weighted Languages over ARTINIAN Semirings

Andreas Maletti
Universität Leipzig
Faculty of Mathematics and Computer Science
PO Box 100 920, 04009 Leipzig, Germany
maletti@informatik.uni-leipzig.de

Nils Oskar Nuernbergk
nils.nuernbergk@gmail.com

Pumping lemmata are the main tool to prove that a certain language does not belong to a class of languages like the recognizable languages or the context-free languages. Essentially two pumping lemmata exist for the recognizable weighted languages: the classical one for the BOOLEAN semiring (i.e., the unweighted case), which can be generalized to zero-sum free semirings, and the one for fields. A joint generalization of these two pumping lemmata is provided that applies to all ARTINIAN semirings, over which all finitely generated semimodules have a finite bound on the length of chains of strictly increasing subsemimodules. Since ARTINIAN rings are exactly those that satisfy the Descending Chain Condition, the ARTINIAN semirings include all fields and naturally also all finite semirings (like the BOOLEAN semiring). The new pumping lemma thus covers most previously known pumping lemmata for recognizable weighted languages.

1 Introduction

The class of recognizable languages [28] is certainly the best-studied and one of the most useful classes of languages. It has excellent closure properties, and all standard decision problems for it are decidable. Applications of the recognizable languages are too numerous to list, but include pattern matching [2, Chapter 10], lexical analysis [1], input validation [25], network protocols [18], and DNA sequence analysis [26]. Pumping lemmata are statements of the form that given a suitably long word in the language, we can always identify a subword that can be iterated (or pumped) at will without leaving the language. Such statements exist for many language classes including the recognizable [28] and context-free languages [6], and they allow a relatively straightforward proof that a given language does not belong to the class (e.g., is not recognizable).

In several applications [3, 7, 15], the purely qualitative yes/no-decision of languages is completely insufficient. This led to the introduction of weighted languages [24] (see [21] for an excellent survey), in which each word is assigned a weight from a semiring [12, 11]. The classical recognizable languages are reobtained by considering the support of the recognizable weighted languages over the BOOLEAN semiring $(\{0,1\}, \max, \min, 0, 1)$. The theory of recognizable weighted languages is also very well developed and several textbooks [22, 16, 9] provide excellent introductions.

Determining whether a given weighted language is recognizable is often even more difficult than in the unweighted case, and we again mostly rely on pumping lemmata [13, 20] to prove that a given weighted language is not recognizable. However, the coverage situation is very unsatisfactory. The classical pumping lemma for unweighted languages can be lifted to all zero-sum free semirings [12, 11] (i.e., semirings in which $a+b=0$ implies $a=0=b$) by means of a semiring homomorphism from such a semiring into the BOOLEAN semiring [27] and a construction that avoids zero-divisors [14]. On the other hand, the pumping lemmata of [13, 20] require the semiring to be a field, which necessarily is not

Zs. Gazdag, Sz. Iván, G. Kovásznai (Eds.): 16th International Conference on Automata and Formal Languages (AFL 2023)
EPTCS 386, 2023, pp. 155–169, doi:10.4204/EPTCS.386.13

zero-sum free. Despite their similarities, the two recalled pumping lemmata thus apply to completely disjoint sets of semirings, which do not even cover all semirings (e.g., the finite ring $\mathbb{Z}_4$ is not zero-sum free and not a field). Indeed it is well-known [10] how to handle finite semirings like $\mathbb{Z}_4$ (by encoding the weights into the states), so that the classical unweighted pumping lemma becomes applicable. Similarly, it is known how to handle semirings like $\mathbb{Z}$ that embed into a field, but there are also infinite semirings that are not zero-sum free and not (embeddable into) a field like the ring $\mathbb{Q}[x]/(x^2)$ of rational linear polynomials. The ring $\mathbb{Q}[x]/(x^2)$ cannot embed into a field since it has zero-divisors (e.g., $x \cdot x = 0$), but it fulfills the requirements for our pumping lemma. Hence there are semirings for which we currently have no available pumping lemma, as well as different semirings that permit essentially the same pumping lemma for their recognizable weighted languages but with totally different justifications.

Let us recall the statement of these pumping lemmata. Let $L\colon \Sigma^* \to S$ be a recognizable weighted language, which assigns to each word $w \in \Sigma^*$ a weight $L(w) \in S$ in the semiring S. The support of L is the set $\operatorname{supp} L = \{w \in \Sigma^* \mid L(w) \neq 0\}$ of nonzero-weighted words in L. The pumping lemma states that given a sufficiently long word $w \in \operatorname{supp} L$, there exists a decomposition $w = uxv$ such that $ux^kv \in \operatorname{supp} L$ for infinitely many $k \in \mathbb{N}$. In other words, ux^kv is also nonzero-weighted in L for infinitely many $k \in \mathbb{N}$, where $ux^kv = ux \cdots xv$ with k repetitions of x.

In this contribution we will establish such a pumping lemma for a class of semirings that includes all fields and all finite semirings. Thus, we directly cover both the pumping lemmata of [13, 20] as well as the classical pumping lemma [19, Lemma 2]. We achieve this by following the general approach of [20] while trying to avoid the vector space structure utilized there. This requires some minor adjustments and, in particular, a replacement for the dimension, for which we use the length of a semimodule. A semimodule has finite length if there is a finite bound on the length of strictly increasing chains of subsemimodules. This notion also allows us to define the ARTINIAN semirings that we consider. A semiring is ARTINIAN if each finitely generated semimodule has finite length. The ARTINIAN semirings include all fields and all finite semirings, but not all zero-sum free semirings. However, the mentioned approach for zero-sum free semirings (applying the homomorphism into the BOOLEAN semiring and avoiding zero-divisors) naturally also works with our pumping lemma.

We first show that any endomorphism of a semimodule over an ARTINIAN semiring is surjective if and only if it is injective, which is a generalization of a well-known statement for vector spaces. Following the approach of [20], we introduce pseudoregular endomorphisms using 2 of the 5 characterizing properties utilized in [20, Proposition 1]. Fortunately, these are the two main properties needed for the proof of our pumping lemma, and the remaining 3 properties rely on infrastructure that is not generally available in our semimodules (instead of the vector spaces used in [20]). The argument that a sufficiently long composition of endomorphisms needs to contain a pseudoregular endomorphism can be taken over mostly unchanged from [13], which then almost directly yields our main pumping lemma. Finally, we also briefly consider pumping lemmata for infinite alphabets.

2 Preliminaries

We denote the non-negative integers by $\mathbb{N}$ and the positive integers by $\mathbb{N}_+ = \mathbb{N} \setminus \{0\}$. Moreover, we let $\mathbb{Q}^{\geqslant 0} = \{q \in \mathbb{Q} \mid q \geqslant 0\}$ be the set of non-negative rational numbers. For every alphabet Σ we denote the free monoid over Σ by Σ^*, i.e., Σ^* is the set of all finite words with letters in Σ. We write ε for the empty word (the neutral element of the free monoid). Additionally, we let $\Sigma^+ = \Sigma^* \setminus \{\varepsilon\}$. For all sets A, B, and C and all maps $f\colon A \to B$ and $g\colon B \to C$, we let $\mathrm{id}_A = \{(a,a) \mid a \in A\}$ and $(gf)\colon A \to C$ be the map such that $(gf)(a) = g(f(a))$ for every $a \in A$. Finally, if $A = B$, then we let $f^0 = \mathrm{id}_A$ and $f^{k+1} = ff^k$

for every $k \in \mathbb{N}$.

A *(commutative) semiring* [12, 11] is an algebraic structure $(S,+,\cdot,0,1)$, in which S is a set, called *carrier*, $(S,+,0)$ and $(S,\cdot,1)$ are commutative monoids, called *additive* and *multiplicative monoid* respectively, and

$$r\cdot(s+t) = (r\cdot s)+(r\cdot t), \qquad \text{(distributivity)}$$
$$0\cdot r = 0 \qquad \text{(absorption of 0)}$$

for all $r,s,t \in S$. We will refer to the semiring $(S,+,\cdot,0,1)$ simply by its carrier set S and denote multiplication by juxtaposition as usual. For the rest of the contribution, let S be a commutative semiring.

A (commutative) *ring* is simply a semiring in which every element has an additive inverse, and a (commutative) *semifield* is similarly a semiring in which every element $s \in S \setminus \{0\}$ has a multiplicative inverse. As usual, a (commutative) *field* is a ring that is also a semifield. The BOOLEAN semifield is $\mathbb{B} = (\{0,1\},\max,\min,0,1)$.

An S-semimodule [12, 11] is a tuple $(M,\oplus,0_M,\odot)$ consisting of a commutative monoid $(M,\oplus,0_M)$ and a mapping $\odot\colon S\times M \to M$ such that

$$(r\cdot s)\odot u = r\odot(s\odot u), \qquad \text{(associativity)}$$
$$r\odot(u\oplus v) = (r\odot u)\oplus(r\odot v), \qquad \text{(left distributivity)}$$
$$(r+s)\odot u = (r\odot u)\oplus(s\odot u), \qquad \text{(right distributivity)}$$
$$0\odot u = 0_M \qquad \text{(absorption of 0)}$$

for all semiring elements $r,s \in S$, also called *scalars*, and semimodule elements $u,v \in M$. As before, we write just M for the semimodule $(M,\oplus,0_M,\odot)$, and due to the compatibility axioms presented above, we can safely stop distinguishing the semimodule addition $\oplus$ and semiring addition $+$, writing just $+$ for both, as well as mixed multiplication $\odot$ and semiring multiplication $\cdot$, writing $\cdot$ for both, and the additive neutral element 0_M of the semimodule and its corresponding element 0 of the semiring, writing 0 for both. Finally, we let $su = s\cdot u$ for all $s \in S$ and $u \in M$. It is clear that the semiring S itself forms a semimodule, semimodules over rings are simply modules, and semimodules over fields are vector spaces. A *subsemimodule* of M is a subset $N \subseteq M$ such that $0 \in N$, $m+n \in N$ for all $m,n \in N$, and $r\cdot n \in N$ for all $r \in S$ and $n \in N$. In other words, a subsemimodule is a subset that forms a semimodule itself with respect to the operations of M suitably restricted to N. We write $N \preceq M$ if N is a subsemimodule of M. For every subset $V \subseteq M$ we write $\langle V\rangle$ for the *span* of V (i.e., the smallest subsemimodule of M that contains V) and say that $\langle V\rangle$ is *generated* by V.

Let M and N be two semimodules and $\varphi\colon M \to N$ a mapping. Then φ is *linear* (or a *semimodule homomorphism*) if

$$s\cdot\varphi(u) = \varphi(s\cdot u) \qquad \text{and} \qquad \varphi(u+v) = \varphi(u)+\varphi(v)$$

for all $s \in S$ and $u,v \in M$. Note that $\varphi(0) = 0$ if φ is linear by the former condition. If φ is bijective and linear, then we call φ an *isomorphism* and say that M and N are *isomorphic*, which we write as $M \cong N$. We let $\ker\varphi = \{m \in M \mid \varphi(m) = 0\}$ be the *kernel* of φ and $\operatorname{im}\varphi = \{\varphi(m) \mid m \in M\}$ be the *image* of φ in N, which is always a subsemimodule of N provided that φ is linear. The first isomorphism theorem [4, p. 162, Corollary 5.16] states that $M/\ker\varphi \cong \operatorname{im}\varphi$ for every ring S and linear map φ. Here, $M/\ker\varphi$ is the set of equivalence classes $M/\sim$ with the equivalence relation $\sim$ given by $m \sim n$ if $m-n \in \ker\varphi$ and addition and scalar multiplication defined by $[m]+[n] = [m+n]$ and $s[m] = [sm]$ (where $[m]$ denotes the equivalence class of m). Thus, over a ring S, the linear map φ is injective if and only if $\ker\varphi = \{0\}$.

Moreover, we let

$$\mathrm{Hom}(M,N) = \{\varphi\colon M \to N \mid \varphi \text{ is linear}\}, \quad \mathrm{End}(M) = \mathrm{Hom}(M,M), \quad \text{and} \quad M^{\vee} = \mathrm{Hom}(M,S),$$

which form semimodules with pointwise addition and scalar multiplication. The semimodule $\mathrm{End}(M)$ contains the *endomorphisms* of M, and $M^{\vee}$ is called the *dual semimodule* of M.

Let Q be an arbitrary set. Then

$$S^Q = \{f\colon Q \to S \mid \ker f \text{ is co-finite}\}$$

forms a semimodule with pointwise addition and scalar multiplication that we call the *free semimodule over* Q (unique up to isomorphism as usual). This is justified by the fact [11, p. 194] that for any semimodule M every mapping $\varphi\colon Q \to M$ uniquely extends to a linear map $\widetilde{\varphi}\colon S^Q \to M$ such that $\widetilde{\varphi}(\iota_q) = \varphi(q)$, where $\iota_q \in S^Q$ is the mapping given for every $p \in Q$ by

$$\iota_q(p) = \begin{cases} 1 & \text{if } p = q \\ 0 & \text{otherwise.} \end{cases}$$

In particular, if $\mathrm{im}\,\varphi$ generates M, then $\widetilde{\varphi}$ is surjective. If S is a field, then every semimodule (i.e., vector space) is free, but the same is not true for arbitrary semirings S. If Q is finite, then we say that S^Q is of *rank* $n = |Q|$ and will often identify S^Q with the semimodule S^n.

The spaces $\mathrm{Hom}(M,N)$, $\mathrm{End}(M)$, and $M^{\vee}$ are particularly easy to describe when M and N are free of finite rank [11, p. 195]. These are exactly the matrix spaces

$$\mathrm{Hom}(S^Q,S^P) \cong S^{P\times Q}, \qquad \mathrm{End}(S^Q) \cong S^{Q\times Q}, \qquad \text{and} \qquad (S^Q)^{\vee} \cong S^{\{1\}\times Q} \cong S^Q.$$

Note also that S^Q itself can be identified with the matrix space $S^{Q\times\{1\}}$. Matrix multiplication (i.e., composition of linear maps) is then defined as follows: for every $M \in S^{P\times Q}$ and $N^{Q\times R}$, the matrix $M \cdot N \in S^{P\times R}$ is given for all $p \in P$ and $r \in R$ by

$$(M \cdot N)_{pr} = \sum_{q\in Q} M_{pq} \cdot N_{qr}.$$

We will usually state theorems in terms of linear maps instead of matrices due to their greater generality (non-free semimodules do not generally permit descriptions by matrices) and clarity of presentation.

Let Σ be an alphabet. A *weighted language* over Σ is a function $L\colon \Sigma^* \to S$. Given $w \in \Sigma^*$ and a weighted language $L\colon \Sigma^* \to S$, we occasionally write L_w instead of $L(w)$. The *support* of L is the set $\mathrm{supp}\,L = \{w \in \Sigma^* \mid L_w \neq 0\}$.

A *linear representation* [10] of a weighted language $L\colon \Sigma^* \to S$ is a tuple $(Q, \mathrm{in}, \mathrm{out}, \mu)$, where Q is a finite set of *states*, $\mathrm{in} \in (S^Q)^{\vee}$ is an *input vector*, $\mathrm{out} \in S^Q$ is an *output vector*, and $\mu\colon \Sigma^* \to \mathrm{End}(S^Q)$ is a monoid homomorphism (where the monoid structure on $\mathrm{End}(S^Q)$ is given by composition of maps), such that for every $w \in \Sigma^*$

$$L_w = \mathrm{in} \cdot \mu(w) \cdot \mathrm{out}.$$

If a weighted language L admits a linear representation, then we call L *recognizable*. This definition of recognizability is equivalent to other common definitions given in terms of weighted automata [21].

3 Semimodules of Finite Length

We recall that the dimension of a finite dimensional vector space V provides an upper bound on the number of proper inclusions in any chain of subspaces of V; i.e., if $V_0 \preceq \cdots \preceq V_r$ is a chain of subspaces of V and $r > \dim V$, then there is at least one $0 \leqslant i < r$ such that $V_i = V_{i+1}$.

In this spirit, we define the *length* $\ell(M) \in \mathbb{N} \cup \{\infty\}$ of a semimodule M to be the (possibly infinite) least upper bound on the number of proper inclusions in any chain of subsemimodules of M; i.e.,

$$\ell(M) = \sup\{r \mid M_0 \prec \cdots \prec M_r \text{ is a chain of strictly increasing subsemimodules of } M\}.$$

Clearly, $\dim V = \ell(V)$ for every finite dimensional vector space V. However, the length is distinct from the rank of a free module even if S is a ring. For example, $\mathbb{Z}$ has rank 1 as a $\mathbb{Z}$-module, but $\ell(\mathbb{Z}) = \infty$ since

$$\langle k^m \rangle \prec \langle k^{m-1} \rangle \prec \cdots \prec \langle k \rangle$$

is a chain of strictly increasing submodules of $\mathbb{Z}$ for every $k \in \mathbb{Z} \setminus \{0, 1, -1\}$ and $m \geqslant 2$.

Definition 3.1. We say that a semimodule M has *finite length* if $\ell(M) \in \mathbb{N}$; i.e., $\ell(M)$ is finite. ∎

Let us provide some examples of semimodules that have finite length.

Example 3.2.

(i) Finite dimensional vector spaces over fields have finite length.

(ii) Finite semimodules have finite length.

(iii) We consider the commutative monoid $M = \mathbb{Q}^{\geqslant 0} \cup \{\infty\}$ with $u + \infty = \infty$ for all $u \in M$ and addition defined as in $\mathbb{Q}$ otherwise. Then M is a semimodule over $\mathbb{Q}^{\geqslant 0}$ via

$$m \odot u = \begin{cases} 0 & \text{if } m = 0 \\ \infty & \text{if } m \neq 0 \text{ and } u = \infty \\ m \cdot u & \text{otherwise.} \end{cases}$$

We can easily see that the only subsemimodules of M are $\{0\}$, $\{0, \infty\}$, $\mathbb{Q}^{\geqslant 0}$ and M itself. By considering the inclusions among these subsemimodules, we obtain $\ell(M) = 2$. Notably, this is an example of an infinite semimodule that has finite length, but cannot be embedded into a module over a ring. The embedding fails since ∞ is additively absorptive (i.e., $u + \infty = \infty$ for all $u \in M$, which yields that ∞ cannot be inverted). ∎

Let M be a semimodule that has finite length. Next we show that the image $\operatorname{im}\varphi$ of a linear map $\varphi \colon M \to N$ necessarily has finite length as well.

Lemma 3.3. *Let M and N be semimodules and $\varphi \colon M \to N$ be a linear map. Then $\ell(\operatorname{im}\varphi) \leqslant \ell(M)$.*

Proof. If $\ell(M) = \infty$, then the statement holds automatically. Therefore, suppose that $\ell(M) \in \mathbb{N}$ is finite. We recall that the preimage $\varphi^{-1}(L)$ of a subsemimodule $L \preceq N$ is a subsemimodule of M. To see this, let $u, v \in \varphi^{-1}(L)$. Then $\varphi(u+v) = \varphi(u) + \varphi(v) \in L$, and thus $u + v \in \varphi^{-1}(L)$. Similarly, for every $s \in S$ we have $\varphi(su) = s\varphi(u) \in L$, and thus $su \in \varphi^{-1}(L)$. Thus, any chain $N_0 \preceq \cdots \preceq N_r$ of subsemimodules of $\operatorname{im}\varphi$ induces a chain $\varphi^{-1}(N_0) \preceq \cdots \preceq \varphi^{-1}(N_r)$ of subsemimodules of M. Next, we establish that $\varphi^{-1}(N_i) \prec \varphi^{-1}(N_{i+1})$ for every $0 \leqslant i < r$ such that $N_i \prec N_{i+1}$. To this end, let $u \in N_{i+1} \setminus N_i$ and select $v \in \varphi^{-1}(\{u\})$, which is possible because $N_{i+1} \preceq \operatorname{im}\varphi$. Obviously, $v \notin \varphi^{-1}(N_i)$, which proves that $v \in \varphi^{-1}(N_{i+1}) \setminus \varphi^{-1}(N_i)$ and thus $\varphi^{-1}(N_i) \prec \varphi^{-1}(N_{i+1})$. Hence, $\ell(\operatorname{im}\varphi) \leqslant \ell(M)$ follows immediately from the definition. □

The preceding lemma already suggests that semimodules of finite length share nice properties with finite dimensional vector spaces. In order to harness these, it would be very desirable for the class of finite length semimodules to have good closure properties. However, it is not even closed under direct sums, as the following example demonstrates.

Example 3.4. Consider the semifield $S = \mathbb{Q}^{\max} = (\mathbb{Q}^{\geqslant 0}, \max, \cdot, 0, 1)$. As usual, $\mathbb{Q}^{\max}$ is a semimodule over itself, and the presence of multiplicative inverses immediately yields that $\ell(\mathbb{Q}^{\max}) = 1$ because its only subsemimodules are $\{0\}$ and $\mathbb{Q}^{\max}$: if $H \preceq \mathbb{Q}^{\max}$ and $H \neq \{0\}$, there is an $h \in H$ with $h \neq 0$, so $s = s \cdot h^{-1} \cdot h \in \mathbb{Q}^{\max}$ for all $s \in \mathbb{Q}^{\max}$; whereby $H = \mathbb{Q}^{\max}$ (indeed, this argument works for any semifield).

Now we consider the direct sum $M = \mathbb{Q}^{\max} \oplus \mathbb{Q}^{\max}$ of two copies of $\mathbb{Q}^{\max}$, which consists of pairs of rational numbers with the maximum applied coordinate-wise. Clearly, M is also a $\mathbb{Q}^{\max}$-semimodule via a coordinate-wise product. However, M does not have finite length over $\mathbb{Q}^{\max}$ by the following lemma. ■

Lemma 3.5. *The $\mathbb{Q}^{\max}$-semimodule $\mathbb{Q}^{\max} \oplus \mathbb{Q}^{\max}$ has length $\ell(\mathbb{Q}^{\max} \oplus \mathbb{Q}^{\max}) = \infty$.*

Proof. Let $M = \mathbb{Q}^{\max} \oplus \mathbb{Q}^{\max}$. First, we define the function $q \colon M \to \mathbb{Q}$ such that $q\big(\langle a,b\rangle\big) = \frac{a}{b}$ for every $a, b \in \mathbb{Q}^{\max}$. Obviously,

$$q\big(s\langle a,b\rangle\big) = q\big(\langle sa,sb\rangle\big) = \frac{sa}{sb} = \frac{a}{b} = q\big(\langle a,b\rangle\big) \tag{1}$$

for all $\langle a,b\rangle \in M$ and $s \in \mathbb{Q}^{\max}$. Additionally, for all $\langle a,b\rangle, \langle c,d\rangle \in M$ we have

$$q\Big(\max\big(\langle a,b\rangle, \langle c,d\rangle\big)\Big) \leqslant \max\Big(q\big(\langle a,b\rangle\big), q\big(\langle c,d\rangle\big)\Big) \tag{2}$$

because

$$\frac{a}{\max(b,d)} \leqslant \frac{a}{b} = q\big(\langle a,b\rangle\big) \qquad \text{and} \qquad \frac{c}{\max(b,d)} \leqslant \frac{c}{d} = q\big(\langle c,d\rangle\big),$$

which yield

$$q\Big(\max\big(\langle a,b\rangle, \langle c,d\rangle\big)\Big) = \frac{\max(a,c)}{\max(b,d)} = \max\Big(\frac{a}{\max(b,d)}, \frac{c}{\max(b,d)}\Big) \leqslant \max\Big(q\big(\langle a,b\rangle\big), q\big(\langle c,d\rangle\big)\Big).$$

For every $i \in \mathbb{N}$ let $u_i = \langle i, 1\rangle$ and $M_i = \langle\{u_0, \dots, u_i\}\rangle$ be the subsemimodule generated by $\{u_0, \dots, u_i\}$. Due to the properties (1) and (2) of q, we have $q(u) \leqslant q(u_i)$ for every $u \in M_i$. This immediately yields $M_i \prec M_{i+1}$ for every $i \in \mathbb{N}$ and thus $M_0 \prec \cdots \prec M_i \prec \cdots$ is an infinite chain of strictly increasing subsemimodules.[1] □

Fortunately, for rings S the situation does not look nearly as bleak and the expected equalities for length hold, as expressed in the next theorem.

Theorem 3.6. *Suppose that S is a ring.*

(i) Let M and N be modules such that $N \preceq M$. If N and M/N both have finite length, then M has finite length and $\ell(M) = \ell(N) + \ell(M/N)$.

[1] In fact, this is an example of a more general pathology of semimodules. Finite length semimodules are NOETHERIAN since they satisfy the Ascending Chain Condition (i.e., every ascending chain of subsemimodules terminates). This proof demonstrates that NOETHERIAN semimodules, unlike NOETHERIAN modules over rings, are not closed under direct sums. The same is true for the Descending Chain Condition, which can be seen by setting $v_i = (1, i)$ and $N_i = \langle\{v_0, \dots, v_i\}\rangle$ for all $i \in \mathbb{N}$. Then the chain $N_0 \succ \cdots \succ N_i \succ \cdots$ does not terminate by the same argument as above.

(ii) If M and N are finite-length modules, then $\ell(M \oplus N) = \ell(M) + \ell(N)$.

(iii) If S has finite length, then every module generated by $n \in \mathbb{N}$ elements has length at most $n \cdot \ell(S)$.

Proof.

(i) The proof idea for the inequality $\ell(M) \leqslant \ell(N) + \ell(M/N)$ draws from a proof of the analogous fact for NOETHERIAN rings [17, 10f, Proposition 3.3]. For the sake of a contradiction, assume that $\ell(M) \geqslant r$, where $r = \ell(N) + \ell(M/N) + 1$. Then there exists a chain $L_0 \prec \cdots \prec L_r$ of strictly increasing submodules of M. In the corresponding chain

$$\frac{N + L_0}{N} \preceq \cdots \preceq \frac{N + L_r}{N}$$

of $r+1$ submodules of M/N, at most $\ell(M/N)$ inclusions are proper, so $\ell(N)+1$ inclusions are not. Similarly, in the chain $L_0 \cap N \preceq \cdots \preceq L_r \cap N$ of submodules of N, at most $\ell(N)$ inclusions are proper, so $\ell(M/N)+1$ inclusions are not. By the pigeonhole principle, there exists $0 \leqslant i < r$ such that

$$L_i \cap N = L_{i+1} \cap N \quad \text{and} \quad N + L_i = N + L_{i+1}.$$

We note that the latter result relies on the fact that $N \leqslant H \leqslant K$ and $H/N = K/N$ together imply $H = K$ (since if $k \in K$ and $[h] = [k]$ for some $h \in H$, then $k - h \in N \preceq H$, so $k = h + (k-h) \in H$).

Now, let $u \in L_{i+1}$ be arbitrary. By the second equation above, we have $u = n + v$ for some $n \in N$ and $v \in L_i$. This yields $n = u - v \in L_{i+1} \cap N = L_i \cap N$ and thus $u \in L_i$. Therefore, $L_{i+1} = L_i$, which is the desired contradiction.

Thus, we have shown that $\ell(M) \leqslant \ell(N) + \ell(M/N)$. It remains to show the converse inequality. Note that any submodule of M/N has the form L/N for some $L \preceq M$ such that $N \preceq L$ since N is the preimage of $[0] \in L/N$. This claim was already shown in a more general setting in the proof of Lemma 3.3. Therefore, let $N_0 \prec \cdots \prec N_n$ with $n = \ell(N)$ and $L_0/N \prec \cdots \prec L_m/N$ with $m = \ell(M/N)$ be chains of strictly increasing submodules of N and M/N, respectively, which exist by the definition of the respective length. These chains can be concatenated to obtain a chain

$$N_0 \prec \cdots \prec N_n \preceq L_0 \prec \cdots \prec L_m$$

of submodules of M. Any proper inclusion in the original chains must also be a proper inclusion in the concatenated chain. Thus, $\ell(M) \geqslant n + m = \ell(N) + \ell(M/N)$.

(ii) Let us consider $N_0 = \{(0,n) \mid n \in N\}$. Then $(M \oplus N)/N_0 \cong M$ and $N_0 \cong N$, which yields the claim by Statement (i).

(iii) Let M be a module generated by n elements. Hence M is a linear image of the free module S^n, which by iteration of Statement (ii) satisfies $\ell(S^n) = n \cdot \ell(S)$. Thus, the claim follows directly from Lemma 3.3. □

Hence every finite-length ring has the property that all its finitely generated modules also have finite length. Naturally, there are other semirings that enjoy this property. Trivially, every finitely generated semimodule over a finite semiring (such as the BOOLEAN semifield $\mathbb{B}$) is also finite and therefore of finite length. The following definition establishes the property just discussed, which is fulfilled in all rings and all finite semirings.

Definition 3.7. We say that S is ARTINIAN if every finitely generated semimodule has finite length. ■

As demonstrated in the proof of Theorem 3.6(iii), in order to establish that S is ARTINIAN it suffices to show that free semimodules of finite rank have finite length. Our naming ARTINIAN is a slight abuse of traditional notions since the term is usually used to characterize those modules that satisfy the Descending Chain Condition (i.e., every descending chain of submodules terminates). However, in rings these two notions coincide. Any ring that satisfies the Descending Chain Condition (DCC) also satisfies the Ascending Chain Condition (ACC) [5, p. 90, Theorem 8.5], and any module that satisfies both DCC and ACC has finite length [5, p. 77, Propositions 6.7 and 6.8]. By Theorem 3.6(iii), all finitely generated modules over a finite-length ring also have finite length. The converse implication is trivial. In general, for semirings this equivalence need not hold (see footnote to Lemma 3.5), but since the DCC is nowhere as important for semirings as it is for rings, the authors believe that our use of terminology is harmless.

ARTINIAN semirings retain a very convenient property of endomorphisms of vector spaces, which will be crucial for our approach.

Theorem 3.8. *Suppose that S is* ARTINIAN, *and let M be a finite-length semimodule and $\alpha \in \operatorname{End}(M)$. Then α is surjective if and only if α is injective.*

Proof. The proof simply combines the well-known facts that surjective endomorphisms of NOETHERIAN modules are injective, and injective endomorphisms of modules that satisfy the Descending Chain Condition (i.e., ARTINIAN in the traditional sense) are surjective. These two facts are established here for our semimodules.

We start with necessity. Suppose that α is surjective. For every endomorphism $\varphi \in \operatorname{End}(M)$, we let

$$\operatorname{Ker}\varphi = \{(u,v) \in M \oplus M \mid \varphi(u) = \varphi(v)\}.$$

Then $\operatorname{Ker}\varphi$ is a subsemimodule of $M \oplus M$ by the linearity of φ. Let $r = \ell(M \oplus M)$ and consider the chain

$$\{0\} \prec \operatorname{Ker}\alpha^0 \preceq \operatorname{Ker}\alpha^1 \preceq \cdots \preceq \operatorname{Ker}\alpha^r.$$

The first strictness is justified by $\operatorname{Ker}\alpha^0 = \operatorname{Ker}\mathrm{id}_M = \{(u,u) \mid u \in M\} \succ \{0\}$. Thus, by the finite length r, there exists some $0 \leqslant i < r$ such that $\operatorname{Ker}\alpha^i = \operatorname{Ker}\alpha^{i+1}$.

To prove injectivity, let $u, v \in M$ such that $\alpha(u) = \alpha(v)$. Recall that compositions of surjective functions are surjective. By the surjectivity of α and α^i, there exist $x, y \in M$ such that $\alpha^i(x) = u$ and $\alpha^i(y) = v$. Consequently, $\alpha^{i+1}(x) = \alpha^{i+1}(y)$ and thus $(x,y) \in \operatorname{Ker}\alpha^{i+1} = \operatorname{Ker}\alpha^i$ by our choice of i. However, $(x,y) \in \operatorname{Ker}\alpha^i$ directly yields $u = \alpha^i(x) = \alpha^i(y) = v$. Hence, α is injective.

We continue with sufficiency, so let α be injective. We show for all $j \in \mathbb{N}$ that the condition $u \notin \operatorname{im}\alpha^j$ implies $\alpha(u) \notin \operatorname{im}\alpha^{j+1}$. For the sake of a contradiction, suppose that $j \in \mathbb{N}$ and $u \in M \setminus \operatorname{im}\alpha^j$ are such that $\alpha(u) \in \operatorname{im}\alpha^{j+1}$. Clearly, there exists $v \in M$ such that $\alpha(u) = \alpha^{j+1}(v) = (\alpha\alpha^j)(v) = \alpha\big(\alpha^j(v)\big)$. Next we utilize the injectivity of α to conclude $u = \alpha^j(v)$, which yields $u \in \operatorname{im}\alpha^j$ and our desired contradiction. Thus, $\alpha(u) \notin \operatorname{im}\alpha^{j+1}$.

Suppose that α is not surjective. Then there exists $u \in M$ such that $u \notin \operatorname{im}\alpha$. A straightforward induction utilizing the statement proved in the previous paragraph can now be used to show that $\alpha^j(u) \notin \operatorname{im}\alpha^{j+1}$ for all $j \in \mathbb{N}$. However, this yields that the chain

$$M = \operatorname{im}\alpha^0 \succeq \operatorname{im}\alpha^1 \succeq \cdots \succeq \operatorname{im}\alpha^j \succeq \cdots$$

has infinitely many proper inclusions, which contradicts that M has finite length. Therefore, α must be surjective. We note that for sufficiency we only used that M has finite length (not that S is actually ARTINIAN). □

4 Pseudoregular Endomorphisms

At this point we have established sufficient background for our main notion, pseudoregular endomorphisms, that will be successfully utilized in our pumping lemmata. The special properties that define them are established in the next lemma.

Lemma 4.1 (see [20, Proposition 1]). *Let M be a semimodule and $\alpha \in \mathrm{End}(M)$. The following are equivalent:*

(i) $\operatorname{im}\alpha = \operatorname{im}\alpha^2$.

(ii) There exist $\gamma, \beta \in \mathrm{End}(M)$ such that $\alpha = \gamma\beta$ and $\operatorname{im}\beta = \operatorname{im}(\beta\gamma\beta)$.

Proof.

- We start with the implication (i) → (ii). To this end, we select $\gamma = \mathrm{id}_M$ and $\beta = \alpha$ and observe that

$$\alpha = \mathrm{id}_M\alpha = \gamma\beta \qquad \text{and} \qquad \operatorname{im}\beta = \operatorname{im}\alpha = \operatorname{im}\alpha^2 = \operatorname{im}(\alpha\mathrm{id}_M\alpha) = \operatorname{im}(\beta\gamma\beta).$$

- For the converse implication (ii) → (i), let $\gamma, \beta \in \mathrm{End}(M)$ such that $\alpha = \gamma\beta$ and $\operatorname{im}\beta = \operatorname{im}(\beta\gamma\beta)$. Then

$$\operatorname{im}\alpha^2 = \operatorname{im}(\gamma\beta\gamma\beta) = \gamma\big(\operatorname{im}(\beta\gamma\beta)\big) = \gamma(\operatorname{im}\beta) = \operatorname{im}(\gamma\beta) = \operatorname{im}\alpha. \qquad \square$$

Definition 4.2. Let M be a semimodule. An endomorphism $\alpha \in \mathrm{End}(M)$ satisfying the conditions of Lemma 4.1 is called *pseudoregular*. ■

REUTENAUER [20, Proposition 1] provides further characterizations of pseudoregular endomorphisms that hold for a field S. It is worthwhile to consider the following consequence. Let α be a nonzero pseudoregular endomorphism of a finite dimensional vector space V. Then there exists $k \leqslant \dim V$ and a basis $\mathscr{B}$ of V such that the matrix representation of α with respect to $\mathscr{B}$ is a block matrix

$$\begin{pmatrix} A & 0_{(n-k)\times k} \\ 0_{k\times(n-k)} & 0_{(n-k)\times(n-k)} \end{pmatrix},$$

where A is an invertible $k \times k$-matrix and $0_{m\times n}$ is the $m \times n$-zero matrix for every $m, n \in \mathbb{N}_+$.

Using Theorem 3.8 we can adapt another characterization mentioned in [20, Proposition 1] to ARTINIAN semirings.

Lemma 4.3. *Suppose that S is* ARTINIAN*, and let M be a semimodule that has finite length. Then $\alpha \in \mathrm{End}(M)$ is pseudoregular if and only if $\alpha_*\colon \operatorname{im}\alpha \to \operatorname{im}\alpha$, which is defined for every $u \in \operatorname{im}\alpha$ by $\alpha_*(u) = \alpha(u)$, is an isomorphism. If S is a ring, then this is equivalent to* $\operatorname{im}\alpha \cap \ker\alpha = \{0\}$.

Proof. Clearly, $\operatorname{im}\alpha = \operatorname{im}\alpha^2$ is equivalent to surjectivity of α_*, so the result follows from Theorem 3.8. If S is a ring, then $\operatorname{im}\alpha \cap \ker\alpha = \{0\}$ is equivalent to injectivity of α_*, and thereby surjectivity. □

Next we show a generalization of [13, Theorem 2.2]. The general proof idea is largely unchanged, but the lack of vector space structure requires some adjustments in the details. The same theorem can be shown for vector spaces in a much more straightforward manner using linear recurrences (see [20, Lemma 1]), but as this proof relies on the existence of characteristic polynomials of endomorphisms, it cannot be directly adapted to more general semirings.

Theorem 4.4 (see [13, Theorem 2.2]). *Let M be a semimodule such that its dual $M^\vee$ has finite length. Moreover, let $\alpha \in \mathrm{End}(M)$ be pseudoregular, and let $f \in M^\vee = \mathrm{Hom}(M,S)$ and $v \in M$. We consider the sequence $(s_k)_{k\in\mathbb{N}}$ of elements of S given for every $k \in \mathbb{N}$ by*

$$s_k = f\big(\alpha^k(v)\big).$$

If $s_1 \neq 0$, then $s_k \neq 0$ for infinitely many $k \in \mathbb{N}$. More precisely, at most $\ell(M^\vee)$ values of s_k vanish in a row.

Proof. We prove this statement in three steps.

(i) As before, we define $\alpha_*\colon\ \mathrm{im}\,\alpha \to \mathrm{im}\,\alpha$ for every $u \in \mathrm{im}\,\alpha$ by $\alpha_*(u) = \alpha(u)$. Since α_* is surjective, we can find a right inverse $\alpha^*\colon\ \mathrm{im}\,\alpha \to \mathrm{im}\,\alpha$ such that $\alpha_*\alpha^* = \mathrm{id}_{(\mathrm{im}\,\alpha)}$.[2] Next, we define ρ to be the map that sends each element $g\colon M \to S$ of $M^\vee$ to its restriction $g|_{\mathrm{im}\,\alpha}$ to $\mathrm{im}\,\alpha$; i.e.,

$$\rho\colon M^\vee \to (\mathrm{im}\,\alpha)^\vee \qquad \text{with} \qquad \rho(g) = g|_{\mathrm{im}\,\alpha}$$

for all $g \in M^\vee = \mathrm{Hom}(M,S)$. Clearly, ρ is linear, so $\mathrm{im}\,\rho$ has finite length by Lemma 3.3. Fix some $n_0 \in \mathbb{N}_+$ and let $f_i = \rho(f\alpha^{n_0+i})$ for every $i \in \mathbb{N}$. Then

$$f_i = \rho(f\alpha^{n_0+i}) = \rho(f\alpha^{n_0+i})\mathrm{id}_M = \rho(f\alpha^{n_0+i})\alpha_*\alpha^* = \rho(f\alpha^{n_0+i+1})\alpha^* = f_{i+1}\alpha^*$$

for every $i \in \mathbb{N}$.

(ii) Let $r = \ell(\mathrm{im}\,\rho) + 1$ and $M_i = \langle\{f_r, \ldots, f_{r-i}\}\rangle$ be the subsemimodule of $\mathrm{im}\,\rho$ that is generated by $\{f_r, \ldots, f_{r-i}\}$ for every $0 \leqslant i \leqslant r$. We consider the chain $M_0 \preceq M_1 \preceq \ldots \preceq M_r$. Since $r > \ell(\mathrm{im}\,\rho)$, at least one of these inclusions is not proper. Let $0 < i \leqslant r$. If $M_{i-1} = M_i$, then $M_i = M_{i+1}$, which we prove as follows. Since $M_i = M_{i-1}$, there exist coefficients $\lambda_0, \ldots, \lambda_r \in S$ such that

$$f_{r-i} = \sum_{j=0}^{i-1} \lambda_j f_{r-j}$$

and thus

$$f_{r-(i+1)} = f_{r-i-1} = f_{r-i}\alpha^* = \Big(\sum_{j=0}^{i-1} \lambda_j f_{r-j}\Big)\alpha^* = \sum_{j=0}^{i-1} \lambda_j f_{r-j-1} = \sum_{j=1}^{i} \lambda_{j-1} f_{r-j}.$$

Therefore, $f_{r-(i+1)} \in \langle\{f_{r-1}, \ldots, f_{r-i}\}\rangle \preceq M_i$, so we have $M_{i+1} = M_i$ by the construction of M_i. A straightforward induction then proves that $M_r = M_{r-1}$. Hence, there are coefficients $\mu_1, \ldots, \mu_r \in S$ such that

$$f_0 = \sum_{j=1}^{r} \mu_j f_j. \tag{†}$$

(iii) Finally, let $s_1 \neq 0$. Assume by way of contradiction that there are only finitely many $k \in \mathbb{N}$ such that $s_k \neq 0$. Then there is some $n \in \mathbb{N}$ such that $s_n \neq 0$ and $s_k = 0$ for all $k > n$. In particular, $s_{n+1} = \cdots = s_{n+r} = 0$. Set $n_0 = n - 1$ and define f_i as above. Then

$$s_n = f_0\big(\alpha(v)\big) = \Big(\sum_{j=1}^{r} \mu_j f_j\Big)\big(\alpha(v)\big) = \sum_{j=1}^{r} \mu_j f_j\big(\alpha(v)\big) = \sum_{j=1}^{r} \mu_j s_{n+j} = 0$$

[2]In the most general setting, finding a right inverse of a surjective function requires the Axiom of Choice. In all cases of interest to us, this is not necessary. If S is ARTINIAN, then α_* is bijective, so there is a unique both-sided linear inverse. If $\mathrm{im}\,\alpha$ is free of finite rank, then it suffices to choose finitely many preimages for the free generators of $\mathrm{im}\,\alpha$.

by (†), which contradicts the choice of n. Therefore, there must be infinitely many $k \in \mathbb{N}$ such that $s_k \neq 0$. In particular, we have shown that at most $r-1 = \ell(\operatorname{im}\rho) \leqslant \ell(M^\vee)$ values of s_k can vanish in a row. □

We note that the previous proof relies crucially on the commutativity of S, since $M^\vee$ need not be a semimodule in the non-commutative case. Semimodules of finite length allow us to determine that an endomorphism is pseudoregular simply by looking at its factorizations. We will later use a statement of this kind for the proof of our pumping lemma. However, one similar proposition can already be adapted directly from the theory of vector spaces without any further work.

Lemma 4.5 (see [13, Proposition 2.1]). *Let M be a finite-length semimodule and $\alpha \in \operatorname{End}(M)$. Then $\alpha^{\ell(M)}$ is pseudoregular.*

Proof. Consider the chain

$$M = \operatorname{im}\alpha^0 \succeq \operatorname{im}\alpha^1 \succeq \operatorname{im}\alpha^2 \succeq \ldots \succeq \operatorname{im}\alpha^{\ell(M)} \succeq 0.$$

By definition at least one of these inclusions is not proper. Let $0 < i \leqslant \ell(M)$. If $\operatorname{im}\alpha^{i-1} = \operatorname{im}\alpha^i$, then indeed also $\operatorname{im}\alpha^i = \operatorname{im}\alpha^{i+1}$, so by another straightforward induction we also obtain $\operatorname{im}\alpha^{\ell(M)} = \operatorname{im}\alpha^{2\ell(M)}$, which yields that $\alpha^{\ell(M)}$ is pseudoregular. If only the last inclusion is improper (i.e., $\operatorname{im}\alpha^{\ell(M)} = \{0\}$), then $\alpha^{\ell(M)}$ is the zero morphism and thereby trivially pseudoregular as well. This concludes all cases and in each case $\alpha^{\ell(M)}$ is pseudoregular. □

5 Pumping Lemmata

In this final section, we combine our derived results to provide a pumping lemma for recognizable weighted languages. In general, pumping lemmata are used to prove that a (weighted) language is not recognizable. For illustration, we recall the classical pumping lemma for recognizable languages, which is the main tool to prove that a given language is not recognizable [28].

Theorem 5.1 (see [19, Lemma 2]). *Let L be a recognizable language. Then there exists $n \in \mathbb{N}$ such that for every $w \in L$ with $|w| \geqslant n$ there is a factorization $w = uxv$ with $x \neq \varepsilon$ such that $ux^kv \in L$ for all $k \in \mathbb{N}$.*

Next, we show a similar result for recognizable weighted languages, which was originally proven for fields in [13, Theorem 5], although we adapted our proof using the ideas of [20, Theorem 2]. These ideas directly yield the basic approach using our Theorem 4.4. Given a linear representation $(Q, \mathrm{in}, \mathrm{out}, \mu)$ of a weighted language $L \colon \Sigma^* \to S$ such that (i) S^Q has finite length, (ii) $\mu(x)$ is pseudoregular for some $x \in \Sigma^*$, and (iii) $uxv \in \operatorname{supp} L$ for some $u, v \in \Sigma^*$, then for infinitely many $k \in \mathbb{N}$,

$$L(ux^kv) = \mathrm{in} \cdot \mu(ux^kv) \cdot \mathrm{out} \neq 0.$$

We use that $S^Q \cong (S^Q)^\vee$; i.e., that S^Q and $(S^Q)^\vee$ are isomorphic, yielding finite length for $(S^Q)^\vee$. Additionally, we note that we do not conclude that $L(ux^kv) \neq 0$ for all $k \in \mathbb{N}$ (as in Theorem 5.1), but rather the inequality only holds for infinitely many $k \in \mathbb{N}$. However, to make this approach applicable to any recognizable weighted language, we still need to identify suitable conditions that enforce that a given word $w \in \Sigma^*$ contains a nontrivial subword $x \in \Sigma^*$ with pseudoregular image $\mu(x)$. A simple combinatorial argument following [20] shows that if w is long enough, then there always exists a factorization $w = uxy$ with $x \neq \varepsilon$ such that $\mu(x)$ is pseudoregular.

Definition 5.2 (see [20])**.** Let Σ be a finite alphabet, $w \in \Sigma^*$, and $n \in \mathbb{N}$. We recursively define when w is a *quasipower of order n*.

(i) If $n = 0$ and $w \neq \varepsilon$, then w is a quasipower of order 0.

(ii) If $n > 0$ and $w = uvu$ for some $u, v \in \Sigma^*$ such that u is a quasipower of order $n-1$, then w is a quasipower of order n. ■

Next, we recall that given any order $r \in \mathbb{N}$ we can identify a bound N_r such that words whose length is at least N_r necessarily contain a quasipower of order r. Indeed the constant N_r can be recursively defined for every $r \in \mathbb{N}$ by

$$N_0 = 1 \qquad \text{and} \qquad N_{r+1} = N_r \cdot (1 + |\Sigma|^{N_r}).$$

Lemma 5.3 (see [23, IV. 5] as cited in [20, Lemma 2])**.** *Let Σ be a finite alphabet and $r \in \mathbb{N}$. There exists an integer $N_r \in \mathbb{N}$ such that every word $w \in \Sigma^*$ with $|w| \geqslant N_r$ contains a subword that is a quasipower of order r.*

Next, still following [20], we show that quasipowers of suitably large order are sufficient to establish the existence of a subword x such that $\mu(x)$ is pseudoregular.

Lemma 5.4 (see [13] as cited in [20, Theorem 1])**.** *Let Σ be a finite alphabet, M a semimodule that has finite length, and $\mu \colon \Sigma^* \to \mathrm{End}(M)$ a monoid homomorphism. Every word $w \in \Sigma^*$ that is a quasipower of order $r = \ell(M) + 1$ contains a subword $x \neq \varepsilon$ such that $\mu(x)$ is pseudoregular.*

Proof. Let $w \in \Sigma^*$ be a quasipower of order r, and let $u_r = w$. There are words $u_0, \dots, u_{r-1}, v_1, \dots, v_n \in \Sigma^+$ such that $u_i = u_{i-1} v_i u_{i-1}$ for all $1 \leqslant i \leqslant r$. Thus,

$$\operatorname{im} \mu(u_i) = \operatorname{im} \mu(u_{i-1} v_i u_{i-1}) \preceq \operatorname{im} \mu(u_{i-1}),$$

so we obtain the chain

$$\operatorname{im} \mu(u_r) \preceq \operatorname{im} \mu(u_{r-1}) \preceq \ldots \preceq \operatorname{im} \mu(u_0)$$

of $r+1$ subsemimodules of M. Therefore, $\operatorname{im} \mu(u_i) = \operatorname{im} \mu(u_{i-1})$ for some $1 \leqslant i \leqslant r$, which yields

$$\operatorname{im} \mu(u_{i-1}) = \operatorname{im} \mu(u_i) = \operatorname{im} \mu(u_{i-1} v_i u_{i-1}) = \operatorname{im}\big(\mu(u_{i-1}) \mu(v_i) \mu(u_{i-1})\big).$$

By Lemma 4.1(ii) we obtain that $\mu(v_i)\mu(u_{i-1}) = \mu(v_i u_{i-1})$ is pseudoregular. Hence, we set $x = v_i u_{i-1}$ to complete the proof. □

Our pumping lemma now follows directly. The next main theorem still contains the technical restriction that the semimodule S^Q has finite length, where Q is the set of states of a linear representation for a given recognizable weighted language. A slightly more direct statement is expressed in the corollary that follows the next theorem.

Theorem 5.5 (see [13, Theorem 5] as cited in [20, Theorem 2])**.** *Let Σ be a finite alphabet. Moreover, let $(Q, \mathrm{in}, \mathrm{out}, \mu)$ be a linear representation for the weighted language $L \colon \Sigma^* \to S$. If S^Q has finite length, then there exists an integer $N \in \mathbb{N}$ such that for every $w \in \operatorname{supp} L$ with $|w| \geqslant N$ there exists a factorization $w = uxv$ with $x \neq \varepsilon$ such that*

$$\{ux^k v \mid k \in \mathbb{N}\} \cap \operatorname{supp} L$$

is infinite.

Proof. Let $r = \ell(M) + 1$ and $N = N_r$ as in Lemma 5.3. Since $|w| \geqslant N$, the word w contains a quasipower of order r by Lemma 5.3, and by Lemma 5.4 there exists a factorization $w = uxv$ such that $x \neq \varepsilon$ and $\mu(x)$ is pseudoregular. Moreover, $\mathrm{in} \cdot \mu(u) \in (S^Q)^\vee$ and $\mu(v) \cdot \mathrm{out} \in S^Q$. By assumption we have

$$L_w = \mathrm{in} \cdot \mu(u)\mu(x)\mu(v) \cdot \mathrm{out} \neq 0.$$

Since $S^Q \cong (S^Q)^\vee$ and S^Q has finite length, we can apply Theorem 4.4 to obtain that for infinitely many $k \in \mathbb{N}$,

$$\big(\mathrm{in} \cdot \mu(u)\big) \cdot \mu(x)^k \cdot \big(\mu(v) \cdot \mathrm{out}\big) \neq 0.$$

Since $\mu(x)^k = \mu(x^k)$, this completes the proof. □

By extending this theorem from its original statement for fields to more general semirings, we have identified a unified framework for the classical pumping lemma by RABIN and SCOTT [19] (for the BOOLEAN semifield) and the pumping lemma for recognizable weighted languages over fields by JACOB [13]. In practice, it is useful to be able to reason about recognizability without knowing the number of states a potential linear representation might have, which makes the requirement that S^Q has finite length troublesome. This can be remedied by requiring our semiring S to be ARTINIAN, which of course still subsumes all the cases covered by the already mentioned pumping lemmata.

Corollary (of Theorem 5.5)**.** *Let Σ be a finite alphabet, S be an* ARTINIAN *semiring, and L be a recognizable weighted language $L\colon \Sigma^* \to S$. Then there exists an integer $N \in \mathbb{N}$ such that for every $w \in \operatorname{supp} L$ with $|w| \geqslant N$ there exists a factorization $w = uxv$ with $x \neq \varepsilon$ such that*

$$\{ux^k v \mid k \in \mathbb{N}\} \cap \operatorname{supp} L$$

is infinite.

Example 5.6. Directly generalizing a classical example of a non-regular language, there is no recognizable weighted language L over an ARTINIAN semiring such that $\operatorname{supp} L = \{a^n b^n \mid n \in \mathbb{N}\}$. Suppose that there is an ARTINIAN semiring S and a recognizable weighted language $L\colon \{a,b\}^* \to S$ such that $\operatorname{supp} L = \{a^n b^n \mid n \in \mathbb{N}\}$. By the Corollary of Theorem 5.5 there exists $N \in \mathbb{N}$ such that $w = a^N b^N$ admits a decomposition $w = uxv$ with $x \neq \varepsilon$ such that $\{ux^k v \mid k \in \mathbb{N}\} \cap \operatorname{supp} L$ is infinite. Obviously this is a contradiction since no suitable subword $x \neq \varepsilon$ (consider the cases $x = a^m$, $x = b^m$, and $x = a^m b^n$) exists. We note that such a recognizable weighted language L over a *non-commutative* semiring exists. ■

If we drop the assumption that the alphabet Σ is finite, then we obtain a notion of recognizable weighted languages that is useful when applying the same pumping techniques to weighted tree languages (see, for example, [8, Theorem 9.2]). One result that would be an ideal candidate for extension to semimodules is recalled next. Its extension would immediately yield pumping lemmata of various forms (e.g. [20, Theorem 4]).

Theorem 5.7 (see [20, Theorem 3])**.** *Let Σ be a (not necessarily finite) alphabet and V a vector space of finite nonzero dimension. There is an integer N such that for each homomorphism $\mu\colon \Sigma^* \to \mathrm{End}(V)$, every word $w \in \Sigma^*$ with $|w| \geqslant N$ contains a subword $x \neq \varepsilon$ such that $\mu(x)$ is pseudoregular.*

Unfortunately, the proof of this theorem uses the relationship of nonvanishing elements of exterior powers of V to their components' linear independence. This cannot be easily extended even to (non-integral) rings. We conclude this section by stating two weak pumping lemmata for recognizable weighted languages over infinite alphabets.

Theorem 5.8. *Let Σ be a (possibly infinite) alphabet and $L\colon \Sigma^* \to S$ be a recognizable weighted language with linear representation $(Q, \mathrm{in}, \mathrm{out}, \mu)$ such that S^Q has finite length $N = \ell(S^Q)$. If there exists $w \in \operatorname{supp} L$ with $w = ab^N c$, then the set $\{ab^k c \mid k \in \mathbb{N}\} \cap \operatorname{supp} L$ is infinite.*

Proof. By Lemma 4.5, $\mu(b^N) = \mu(b)^N$ is pseudoregular. Then the claim follows exactly as in Theorem 5.5. $\square$

Theorem 5.9. *Let the semiring S be finite, Σ a (possibly infinite) alphabet, and $L\colon \Sigma^* \to S$ be a recognizable weighted language with linear representation $(Q, \mathrm{in}, \mathrm{out}, \mu)$. There is an integer N such that for every $w \in \operatorname{supp} L$ with $|w| \geqslant N$ there exists a factorization $w = uxv$ with $x \neq \varepsilon$ such that*

$$\{ux^k v \mid k \in \mathbb{N}\} \cap \operatorname{supp} L$$

is infinite.

Proof. Since $\mathrm{End}(S^Q)$ is finite, we can reduce to the case of finite alphabets. To this end, we define the relation $\sim = \operatorname{Ker} \mu$ on Σ (where $\operatorname{Ker}\mu$ is defined as in Theorem 3.8). Clearly, $\sim$ is an equivalence relation. From each of the finite number of equivalence classes $[m]$ we choose a representative r_m. Now, we let $\Gamma = \{r_m \mid m \in \Sigma\}$ and extend the mapping $m \mapsto r_m$ to the unique monoid homomorphism $\psi\colon \Sigma^* \to \Gamma^*$. By definition of $\sim$, it is obvious that $\mu(w) = \mu(\psi(w))$ for all $w \in \Sigma^*$.

Since $|\mathrm{End}(S^Q)| \leqslant |S|^{|Q|^2}$ (consider matrices), we have $|\Gamma| \leqslant |S|^{|Q|^2}$. Let N be as in Theorem 5.5. For every $w \in \Sigma^*$ with $|w| \geqslant N$. there exists a factorization $w = uxv$ with $x \neq \varepsilon$ and infinite

$$\{\psi(ux^k v) \mid k \in \mathbb{N}\} \cap \operatorname{supp} L.$$

By definition of ψ, it is clear that this implies the infiniteness of the set

$$\{ux^k v \mid k \in \mathbb{N}\} \cap \operatorname{supp} L.$$

$\square$

References

[1] Alfred V. Aho, Ravi Sethi & Jeffrey D. Ullman (1985): *Compilers: Principles, Techniques, and Tools*. Addison-Wesley.

[2] Alfred V. Aho & Jeffrey D. Ullman (1992): *Foundations of Computer Science*. W.H. Freeman.

[3] Jürgen Albert & Jarkko Kari (2009): *Digital Image Compression*. In Handbook of Weighted Automata [9], chapter 11, pp. 453–479, doi:10.1007/978-3-642-01492-5_11.

[4] Paolo Aluffi (2009): *Algebra: Chapter 0*. Graduate Studies in Mathematics 104, American Mathematical Society, doi:10.1090/gsm/104/01.

[5] Michael F. Atiyah & Ian G. MacDonald (1994): *Introduction To Commutative Algebra*. Addison-Wesley Series in Mathematics, Avalon Publishing.

[6] Jean-Michel Autebert, Jean Berstel & Luc Boasson (1997): *Context-free Languages and Pushdown Automata*. In Grzegorz Rozenberg & Arto Salomaa, editors: Handbook of Formal Languages, chapter 3, 1, Springer, pp. 111–174, doi:10.1007/978-3-642-59136-5_3.

[7] Christel Baier, Marcus Größer & Frank Ciesinski (2009): *Model Checking Linear-Time Properties of Probabilistic Systems*. In Handbook of Weighted Automata [9], chapter 13, pp. 519–570, doi:10.1007/978-3-642-01492-5_13.

[8] Jean Berstel & Christophe Reutenauer (1982): *Recognizable Formal Power Series on Trees*. Theoretical Computer Science 18(2), pp. 115–148, doi:10.1016/0304-3975(82)90019-6.

[9] Manfred Droste, Werner Kuich & Heiko Vogler (2009): *Handbook of Weighted Automata*. Monographs in Theoretical Computer Science. An EATCS Series, Springer, doi:10.1007/978-3-642-01492-5.

[10] Manfred Droste & Dietrich Kuske (2021): *Weighted Automata*. In Jean-Eric Pin, editor: Handbook of Automata Theory, chapter 4, 1, EMS Press, pp. 113–150, doi:10.4171/automata-1/4.

[11] Jonathan S. Golan (1999): *Semirings and their Applications*. Kluwer Academic, Dordrecht, doi:10.1007/978-94-015-9333-5.

[12] Udo Hebisch & Hanns J. Weinert (1998): *Semirings—Algebraic Theory and Applications in Computer Science*. Series in Algebra 5, World Scientific, doi:10.1142/3903.

[13] Gerard Jacob (1980): *Un théorème de factorisation des produits d'endomorphismes de K^n*. Journal of Algebra 63(2), pp. 389–412, doi:10.1016/0021-8693(80)90080-0.

[14] Daniel Kirsten (2011): *The Support of a Recognizable Series over a Zero-sum Free, Commutative Semiring is Recognizable*. Acta Cybernetica 20(2), pp. 211–221, doi:10.14232/actacyb.20.2.2011.1.

[15] Kevin Knight & Jonathan May (2009): *Applications of Weighted Automata in Natural Language Processing*. In Handbook of Weighted Automata [9], chapter 14, pp. 571–596, doi:10.1007/978-3-642-01492-5_14.

[16] Werner Kuich & Arto Salomaa (1986): *Semirings, Automata, Languages*. Monographs in Theoretical Computer Science. An EATCS Series, Springer, doi:10.1007/978-3-642-69959-7.

[17] James Milne (2020): *A Primer of Commutative Algebra*. https://www.jmilne.org/math/xnotes/CA.pdf.

[18] Antonio Munoz, Sakir Sezer, Dwayne Burns & Gareth Douglas (2011): *An Approach for Unifying Rule Based Deep Packet Inspection*. In: Proc. IEEE Int. Conf. on Communications, IEEE, pp. 1–5, doi:10.1109/icc.2011.5963095.

[19] Micheal O. Rabin & Dana Scott (1959): *Finite Automata and Their Decision Problems*. IBM Journal of Research and Development 3(2), pp. 114–125, doi:10.1147/rd.32.0114.

[20] Christophe Reutenauer (1980): *An Ogden-like Iteration Lemma for Rational Power Series*. Acta Informatica 13(2), pp. 189–197, doi:10.1007/bf00263993.

[21] Jacques Sakarovitch (2009): *Rational and Recognisable Power Series*. In Handbook of Weighted Automata [9], chapter 4, pp. 105–174, doi:10.1007/978-3-642-01492-5_4.

[22] Arto Salomaa & Matti Soittola (1978): *Automata-Theoretic Aspects of Formal Power Series*. Texts and Monographs in Computer Science, Springer, doi:10.1007/978-1-4612-6264-0.

[23] Marcel-Paul Schützenberger (1961): *On a Special Class of Recurrent Events*. The Annals of Mathematical Statistics 32(4), pp. 1201–1213, doi:10.1214/aoms/1177704860.

[24] Marcel-Paul Schützenberger (1961): *On the Definition of a Family of Automata*. Information and Control 4(2–3), pp. 245–270, doi:10.1016/S0019-9958(61)80020-X.

[25] Mark P. J. van der Loo & Edwin de Jonge (2021): *Data Validation Infrastructure for R*. Journal of Statistical Software 97(10), pp. 1–31, doi:10.18637/jss.v097.i10.

[26] Gunnar von Heijne (1987): *Sequence Analysis in Molecular Biology*. Academic Press, doi:10.1016/B978-0-12-725130-1.X5001-2.

[27] Huaxiong Wang (1997): *On Characters of Semirings*. Houston Journal of Mathematics 23(3), pp. 391–405.

[28] Sheng Yu (1997): *Regular Languages*. In Grzegorz Rozenberg & Arto Salomaa, editors: Handbook of Formal Languages, chapter 2, 1, Springer, pp. 41–110, doi:10.1007/978-3-642-59136-5_2.