

# FACTORIZING THROUGH MONOMIAL REPRESENTATIONS: ARITHMETIC CHARACTERIZATIONS AND AMBIGUITY OF WEIGHTED AUTOMATA

ANTONI PUCH AND DANIEL SMERTNIG

**ABSTRACT.** We characterize group representations that factor through monomial representations, respectively, block-triangular representations with monomial diagonal blocks, by arithmetic properties. Similar results are obtained for semigroup representations by invertible transformations. The characterizations use results on unit equations from Diophantine number theory (by Evertse, van der Poorten, and Schlickewei in characteristic zero, and by Derksen and Masser in positive characteristic).

Specialized to finitely generated groups in characteristic zero, one of our main theorems recovers an improvement of a very recent similar characterization by Corvaja, Demeio, Rapinchuk, Ren, and Zannier that was motivated by the study of the bounded generation (BG) property. In positive characteristic, we get a characterization of linear BG groups, recovering a theorem of Abért, Lubotzky, and Pyber from 2003.

Our motivation comes from weighted finite automata (WFA) over a field. For invertible WFA we show that  $M$ -ambiguity, finite ambiguity, and polynomial ambiguity are characterized by arithmetic properties. We discover a full correspondence between arithmetic properties and a complexity hierarchy for WFA based on ambiguity. In the invertible case, this is a far-reaching generalization of a recent result by Bell and the second author, characterizing unambiguous WFA, that resolved a 1979 conjecture of Reutenauer. As a consequence, using the computability of the (linear) Zariski closure of a finitely generated matrix semigroup, the  $M$ -ambiguity, finite ambiguity, and polynomial ambiguity properties are algorithmically decidable for invertible WFA.

## CONTENTS

1. Introduction	2
2. Notions and Preliminaries	12
3. Statements of Main Results	15
4. Guiding Examples and Special Cases	17
5. Unit Equations: Linear Equations over Finitely Generated Groups	21
6. Locally Bézivin implies Weakly Epimonomial	25
7. Characterization of Locally Bézivin Groups	32
8. Characterization of Locally Finitely Generated Spectrum	42
9. Application to Automata	46
10. Decidability	52
11. Conclusion and Open Questions	56
References	57

---

2020 *Mathematics Subject Classification.* Primary 20H20; Secondary 11D61, 20G15, 20M35, 68Q70.

*Key words and phrases.* ambiguity of automata, monomial representations of groups, unit equations, rational series, weighted finite automata.

## 1. INTRODUCTION

Let  $f: \mathbb{Z}_{\geq 0} \rightarrow \mathbb{Q}$  be a linear recurrence sequence (LRS). It is well-known that this is equivalent to the formal power series  $F = \sum_{n=0}^{\infty} f(n)x^n$  representing a *rational* function (without a pole at the origin). Within the class of LRSs, several *structural* properties of the function  $F$  are reflected in *arithmetic* properties of the sequence  $f$  of coefficients, as demonstrated in the following instances.

- If each pole of  $F$  is at a root of unity, then the partial fraction decomposition immediately shows that the asymptotic growth of  $f$ , as measured by the logarithmic Weil height (meaning  $h(a/b) = \log \max\{|a|, |b|\}$  if  $\gcd(a, b) = 1$ ), is in  $O(\log(n))$ . Hence, the sequence  $f$  satisfies a *growth restriction*.
- If  $F$  has only simple poles, then the coefficients of  $f$  are linear combinations of exponentials, and hence are contained in a set of the form

$$M \cdot \Gamma_0 := \left\{ \sum_{i=1}^m g_i : m \leq M, g_i \in \Gamma \right\}$$

with  $\Gamma \leq \mathbb{Q}^\times$  a finitely generated subgroup and  $M \in \mathbb{Z}_{\geq 0}$  (a *multiplicative restriction*).

- If  $F$  has only simple poles *and* they are all at roots of unity, then  $f$  takes only finitely many values (a *finiteness restriction*).

It is much harder to show that one can sometimes recover such structural properties from the induced arithmetic ones, that is, there are corresponding *inverse results*. For the examples given here, this is possible even when one relaxes the condition of rationality on  $F$  and only assumes that  $F$  is *D-finite*, with rationality being obtained as part of the conclusion: for finiteness this was shown by van der Poorten and Shparlinski [PS96] and by Bell and Chen [BC17]; for the stated growth restriction it is due to Bell, Nguyen, and Zannier [BNZ20, Theorem 1.3]; for the stated multiplicative restriction, it is a result of Bézivin [Béz86, Théorème 4], who generalized results of Pólya [Pól21], Benzaghou [Ben70], and himself [Béz87] which dealt with the case  $M = 1$ .

Recently, numerous new such inverse results have been established, often by using deep theorems from number theory: for (*multivariate*) *D-finite* series there are growth restriction results by Bell, Chen, Nguyen, and Zannier [BNZ20; BNZ23; Bel+24]. Here, a full growth-based classification is conjectured, but currently unproven [BNZ23, Question 4.3]; it is related to long-standing open problems on Siegel *E*-functions. Results for multiplicative restrictions have been proven by Bell and the second author [BS23a]. For power series *Mahler functions*, a complete growth-based classification of five distinct structural classes has been established by Adamczewski, Bell, and the second author [ABS23], characterizing in particular *k-automatic* and *k-regular* sequences by their growth.

Among the  $k$ -regular sequences, a more refined growth-based classification was given by Bell, Coons, and Hare [BCH16, Theorem 2], who in fact establish a growth-based characterization of finitely generated subsemigroups of  $\mathbb{Z}^{d \times d}$  [BCH16, Theorem 1].

A *noncommutative rational series* is a function  $f: X^* \rightarrow K$  from a free monoid  $X^*$  on a finite alphabet  $X$  to a field  $K$  that can be computed by a *weighted finite automaton* (WFA) [BR11; Sak09]. The one-letter case corresponds exactly to LRSs. For noncommutative series, an inverse result shows that the Pólya restriction (all nonzero values contained in  $\Gamma$ , with  $\Gamma \leq K^\times$  finitely generated) corresponds precisely to those series computable by *unambiguous automata*. This was conjectured by Reutenauer in 1979 [Reu79] and recently proven by Bell and the second author [BS21].

The motivation, and goal, of the present work is a generalization of the main results of [BS21] to establish a full correspondence between an ambiguity hierarchy (unambiguous, finitely ambiguous, polynomially ambiguous) of weighted automata and a hierarchy based on arithmetic restrictions on the outputs of these automata (or the semigroups arising from their minimal linear representations). We achieve a surprisingly complete and natural correspondence for the special case of *invertible* WFA, that is, WFA for which the transition matrices are invertible (Theorem 1.6 and Fig. 1). It leads to decidability of all of these properties (Corollary 1.7).

In fact, the problem leads us to consider arithmetic restrictions on linear groups: we obtain arithmetic characterizations of finitely generated linear groups that are epimorphic images of *monomial representations*, respectively, *block-triangular representations with monomial diagonal blocks*. In characteristic zero, the results also hold for non-finitely generated linear groups. As we will explain below, in a sense, these results can be seen as an extension of the classical Burnside–Schur theorem (torsion linear groups are locally finite): we characterize multiplicative-type restrictions, and the classic Burnside–Schur theorem characterizes a corresponding finiteness restriction. Corresponding to that, we encounter analogous limitations in positive characteristic.

**Linear groups.** Let  $\Gamma$  be a *finitely generated* multiplicative subgroup of a field  $K$ . In this case we always set  $\Gamma_0 := \Gamma \cup \{0\}$ . For  $M \in \mathbb{Z}_{\geq 0}$  the  $M$ -fold sumset of  $\Gamma_0$  is denoted by  $M \cdot \Gamma_0 := \{ \sum_{i=1}^M g_i : g_i \in \Gamma_0 \}$ ; we write the explicit dot to distinguish it from the dilation of  $\Gamma_0$  by  $M$ . A set  $S \subseteq K$  is *Bézivin* if it is a subset of a set of the form  $M \cdot \Gamma_0$  with  $\Gamma$  finitely generated, and it is *Pólya* if it is a subset of some  $\Gamma_0$  (that is, if one can take  $M = 1$ ).

A matrix group  $G \leq \mathrm{GL}_d(K)$  is *Bézivin* if there is a Bézivin set containing every entry of every  $A \in G$ ; the Pólya property for  $G$  is defined analogously. The Pólya property is clearly not preserved under a change of basis, but the Bézivin property is (Lemma 2.1).

Let us stress that even the Bézivin property depends on the particular (faithful) representation of  $G$  as a linear group, that is, a group  $G$  may have embeddings into

some  $\mathrm{GL}_d(K)$  that are Bézivin, while others are not. For instance, the representation  $(\mathbb{Z}, +) \rightarrow \mathrm{GL}_1(\mathbb{Q})$ ,  $n \mapsto 2^n$  is trivially Bézivin, but  $(\mathbb{Z}, +) \rightarrow \mathrm{GL}_2(\mathbb{Q})$ ,  $n \mapsto \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}$  is not even locally Bézivin. Thus, the (local) Bézivin property is not a property of the abstract group, but really of its chosen linear representation.

A monomial matrix (or generalized permutation matrix, or weighted permutation matrix) is a matrix  $A$  for which every row and column has at most one nonzero entry. Of course, if  $A$  is invertible, it has exactly one nonzero entry in each row and column. Naturally, any finitely generated group of monomial matrices is Bézivin, even Pólya. Our first main result is a sharpening of this observation together with an inverse result.<sup>1</sup>

The field  $K$  is **uniformly power-splitting** for  $G \leq \mathrm{GL}_d(K)$  if there exists some  $N \geq 1$  such that  $\lambda^N \in K$  for every eigenvalue of every matrix  $A \in G$ . As usual, a group property holds **locally** if it holds for every finitely generated subgroup.

**Theorem 1.1.** *Let  $K$  be a field and  $G \leq \mathrm{GL}_d(K)$ . Suppose that  $\mathrm{char} K = 0$  or that  $G$  is finitely generated. Then the following statements are equivalent.*

- (a) *There exists  $T \in \mathrm{GL}_d(K)$  such that  $TGT^{-1}$  is locally Pólya and  $K$  is uniformly power-splitting for  $G$ .*
- (b) *The group  $G$  is locally Bézivin and  $K$  is uniformly power-splitting for  $G$ .*
- (c) *The group representation  $G \hookrightarrow \mathrm{GL}_d(K)$  is an epimorphic image of a monomial representation of  $G$  (over  $K$ ).*
- (d) *The group  $G$  has a finite-index subgroup of simultaneously diagonalizable (over  $K$ ) matrices.*

An arbitrary group  $G$  has a faithful monomial representation over  $K$  if and only if it has a finite-index subgroup  $H$  that is isomorphic to a subgroup of some  $(K^\times)^d$  (a monomial representation is obtained as an induced representation from a representation of  $H$ ). In particular, any such group is virtually abelian. If  $G$  is finitely generated, the uniform power-splitting hypothesis in (a) and (b) can be dropped and the equivalence still holds.

The interesting direction here is that one can deduce from the arithmetic restriction (b) structural properties of the representation. We also observe that the local Pólya property and the local Bézivin property characterize the same types of groups, however, the Pólya property only holds for a suitable choice of basis, while the Bézivin property holds in any basis and is therefore much more natural in this context.

The proof of Theorem 1.1 makes essential use of the fundamental theorem on unit equation in characteristic zero, a deep theorem from Diophantine number theory proven

<sup>1</sup>More refined statements of our main theorems will be given in Section 3, after introducing some more technical notions in Section 2.

by Evertse [Eve84] and van der Poorten and Schlickewei [PS91], that can nowadays be deduced from Schmidt’s subspace theorem. In positive characteristic, solutions of unit equations are effectively described by a theorem of Derksen and Masser [DM12], as suitable orbits of compositions of twisted Frobenius automorphisms, and by a theorem of Adamczewski and Bell [AB12, Theorem 3.1], as  $p$ -automatic sets. We use the result by Derksen and Masser. The necessary background is given in Section 5. Using a corollary that works in any characteristic (Proposition 5.6), we are able to give a proof of Theorem 1.1 that works for any characteristic.

*Remark 1.2.* While finishing the write-up of the present paper, we were made aware of the recent papers [Cor+22b; Cor+22a] and preprint [Cor+23] by Corvaja, Demeio, Rapinchuk, Ren, and Zannier. The special case of Theorem 1.1 in which  $G$  is finitely generated and  $\text{char } K = 0$ , yields a generalization of one of their main theorems [Cor+23, Theorem 1.13]. Let us discuss the connection and differences in detail.

Corvaja, Demeio, Rapinchuk, Ren, and Zannier introduce the notions of a *Purely Exponential Parametrization* and (PEP) sets. From their definitions, it is obvious that every (PEP) subset of  $K^{d \times d}$  is contained in a set of the form  $(M \cdot \Gamma_0)^{d \times d}$ ; conversely, every set of the form  $(M \cdot \Gamma_0)^{d \times d}$  can be seen to be a (PEP) set (by using enough parameters in the purely exponential parametrization). So, if  $G$  is finitely generated and  $\text{char } K = 0$ , one could replace (b) in Theorem 1.1 by “ $G$  is a subset of a (PEP) set”.

In notation adapted to the present paper, [Cor+23, Theorem 1.13] states that, for a field  $K$  of characteristic zero and a group  $G \leq \text{GL}_d(K)$ , the following statements are equivalent.

- (A)  $G$  is a (PEP) set.
- (B)  $G$  has bounded generation (BG) and every element of  $G$  is diagonalizable (over  $\overline{K}$ ).
- (C)  $G$  is finitely generated and the connected component  $\tilde{G}^0$  of the Zariski closure  $\tilde{G}$  is a torus.

If  $G$  is finitely generated, then (A)  $\Leftrightarrow$  (C) is almost (b)  $\Leftrightarrow$  (d) of Theorem 1.1. A difference is that (A) requires  $G$  to be a (PEP) set, whereas (b) only requires containment in a (PEP) set. So Theorem 1.1(b) a priori makes no assumption on the existence of certain elements, making (b)  $\Rightarrow$  (d) stronger. Of course, then we do not quite get (C)  $\Rightarrow$  (A), but this is the “easier” direction (although it requires a clever observation to see that finite unions of (PEP) sets are again (PEP) sets [Cor+23, Proposition 4.3]).

The stronger implication (b)  $\Rightarrow$  (d) that we deduce is crucial in the derivation of our second main theorem and the applications to automata, because it is easy to see that the involved sets are Bézivin (i.e., subsets of (PEP) sets) but not that they are indeed full (PEP) sets, as the involved properties do not allow us to deduce enough information

about existence of certain elements: this occurs in Proposition 8.4 in the implications (a)  $\Rightarrow$  (b) and (b)  $\Rightarrow$  (c), and in (1) of Proposition 9.3.

However, the arguments of [Cor+23] can also yield the stronger implication in characteristic zero [Cor24]. Variants of our main results hold more generally for semigroups of automorphisms (Section 3), and these more general results are also used in our application to weighted automata.

For not necessarily finitely generated groups  $G$ , the implication (A)  $\Rightarrow$  (C) yields finite generation. Our Theorem 1.1 does not directly imply that Bézivin groups in characteristic 0 are finitely generated, however this can be deduced from (d) using an additional small observation (see Lemma 5.4 and Corollary 7.15). On the other hand, the article [Cor+23] says nothing about the local properties or positive characteristic.

Given the similarity of Theorem 1.1 and [Cor+23, Theorem 1.13], let us note that both proofs are powered by unit equations, however the approach appears to be quite different: Corvaja, Demeio, Rapinchuk, Ren, and Zannier first derive a quantitative point counting theorem [Cor+23, Theorem 1.5]. For this, they need stronger results on unit equations. They also make use of the theory of *generic elements* developed by Prasad and Rapinchuk [PR03; PR14; PR17].

By contrast, we only make use of the basic finiteness result on unit equations (of course, then we also do not get a quantitative point counting theorem). We work directly with linear groups, having no need to pass to linear algebraic groups or to use the theory of generic elements. The key to our proof is instead Corollary 6.5: this allows us to iteratively diagonalize suitable powers of matrices, while forcing a compatible structure on the remaining group (Lemmas 6.6 and 6.8). In this way, step by step, the locally Bézivin group is compelled to reveal its diagonalizable finite-index subgroup to us.

Interestingly, the motivation of [Cor+23] comes from a very different direction, namely the study of the bounded generation (BG) property for linear groups, but one ends up characterizing the same class of groups. In particular, a group consisting of diagonalizable matrices (over  $\overline{K}$ ) has BG if and only if it is finitely generated Bézivin. In characteristic zero we have nothing new to add to this angle. In positive characteristic it was already known that BG groups are virtually abelian by a theorem of Abért, Lubotzky, and Pyber [ALP03, Theorem 1.1]. Since our results hold in any characteristic, in the spirit of [Cor+23], we now also easily recover this theorem from Theorem 1.1 (see Section 4.4). In fact, we get a characterization of linear BG groups in positive characteristic (Corollary 4.3).

The **spectrum** of a group  $G \leq \mathrm{GL}_d(K)$  is the set of its eigenvalues, as a subset of the algebraic closure  $\overline{K}$ . The group  $G$  has **finitely generated spectrum** if its spectrum is contained in a finitely generated subgroup of  $\overline{K}^\times$ . Since connected solvable linear groups are simultaneously triangularizable over the algebraic closure  $\overline{K}$  by the Lie–Kolchin

Theorem [Bor91, Theorem 10.5], it is not hard to see that every virtually solvable group, and thus in particular every block-triangular group with monomial diagonal blocks, has locally finitely generated spectrum (Lemma 8.2). Our second main theorem is an inverse result for this arithmetic restriction.

**Theorem 1.3.** *Let  $K$  be a field and  $G \leq \mathrm{GL}_d(K)$ . Suppose that  $\mathrm{char} K = 0$  or that  $G$  is finitely generated. Then the following statements are equivalent.*

- (a) *The group  $G$  has locally finitely generated spectrum and  $K$  is uniformly power-splitting for  $G$ .*
- (b) *The representation  $G \hookrightarrow \mathrm{GL}_d(K)$  is the epimorphic image of an (upper) block-triangular representation with monomial diagonal blocks (over  $K$ ).*

The next corollary shows that, unlike the Bézivin property, having locally finitely generated spectrum does *not* depend on the particular representation of  $G$ .

**Corollary 1.4** ([Ber05]). *Let  $K$  be a field and  $G \leq \mathrm{GL}_d(K)$  a linear group. Suppose that  $\mathrm{char} K = 0$  or that  $G$  is finitely generated. Then  $G$  has locally finitely generated spectrum if and only if it is virtually solvable.*

Corollary 1.4 has previously been deduced by Bernik [Ber05] in characteristic zero as a consequence of the theory of generic elements by Prasad and Rapinchuk. Bernik even obtains virtual solvability under the—a priori—weaker condition that the spectrum is contained in a finitely generated field. Our proof does not make use of the theory of generic elements or Bernik’s result.

For non-finitely generated groups, Theorems 1.1 and 1.3 do not hold in positive characteristic in general: the group  $\mathrm{GL}_d(\overline{\mathbb{F}_p})$  is not virtually solvable for  $d \geq 2$ , yet, being a torsion group, it is locally finite by Burnside–Schur, hence it has locally finitely generated spectrum and is even locally Bézivin.

However, by Tits’ alternative, a finitely generated linear group is either virtually solvable or contains a non-cyclic free subgroup. Combining this with Corollary 1.4, the following does hold in any characteristic even for non-finitely generated groups.

**Corollary 1.5.** *Let  $K$  be a field. A group  $G \leq \mathrm{GL}_d(K)$  has a non-cyclic free subgroup if and only if it does not have locally finitely generated spectrum.*

To put Theorem 1.3 into perspective, let us compare it to some other results characterizing linear groups by restrictions on their spectrum. We say that  $G \leq \mathrm{GL}_d(\overline{K})$  is **tame** if its spectrum consists of roots of unity. Equivalently, the group  $G$  has *locally finite spectrum*.<sup>2</sup> Tame groups show up in [BCH16] [ABS23, Section 8]. Torsion groups are

<sup>2</sup>A finitely generated group is defined over a finitely generated field and consists of matrices of a fixed dimension. Hence, only finitely many roots of unity are possible for each finitely generated subgroup.



tame, and a well-known variant of Burnside–Schur shows, using a trace argument, that absolutely irreducible tame groups are locally finite (see for instance [Lam91, Lemma 9.3]; compare to Proposition 8.4). In characteristic zero, absolutely irreducible tame groups are even finite, because torsion groups are finite.

Let us now assume that  $\text{char } K = 0$  and  $K = \overline{K}$  is algebraically closed. Let  $G \leq \text{GL}_d(\overline{K})$  be tame. Choosing a suitable basis, we can assume that  $G$  is block-triangular with absolutely irreducible diagonal blocks. Now these diagonal blocks are *finite*. Explicitly,

$$G = \begin{pmatrix} S_1 & K^{d_1 \times d_2} & \dots & K^{d_1 \times d_r} \\ 0 & S_2 & \dots & K^{d_2 \times d_r} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & S_r \end{pmatrix}$$

with finite  $S_i \leq \text{GL}_{d_i}(\overline{K})$ . Irreducible representations of finite groups are always epimorphic images of the regular representation, and we get that  $G$  has a block-triangular representation with permutation matrices in the diagonal blocks.

In this spirit, Theorem 1.3 extends this characterization from groups having *locally finite spectrum* to groups having *locally finitely generated spectrum*.

At the other end of the spectrum (pun intended), if the spectrum of an irreducible group  $G \leq \text{GL}_d(\overline{K})$  is contained in the field  $K$ , then  $G$  is conjugate to a subgroup of  $\text{GL}_d(K)$  by a theorem of Bernik [Ber07; RY19]. Thus, in Bernik’s theorem, a weaker restriction on the spectrum controls representability over a subfield.

Many other spectral properties of matrix (semi)groups have been studied. One that is particularly close in spirit to the locally finitely generated spectrum property is the *submultiplicative spectrum* property [Rad00; Kra05; Gru+12]: A semigroup  $S \subseteq \mathbb{C}^{d \times d}$  has submultiplicative spectrum if, for all  $A, B \in S$  every eigenvalue of  $AB$  is of the form  $\alpha\beta$  with  $\alpha$  an eigenvalue of  $A$  and  $\beta$  an eigenvalue of  $B$ . Semigroups with submultiplicative spectrum have locally finitely generated spectrum, but submultiplicativity is much more restrictive. If  $S \subseteq \mathbb{C}^{d \times d}$  is irreducible with submultiplicative spectrum, then there exists a finite nilpotent group  $G \leq \mathbb{C}^{d \times d}$  such that  $\mathbb{C}^\times S = \mathbb{C}^\times G$  [Rad00, Proposition 2.7].

Incidentally, the Burnside–Schur Theorem as well as the mentioned results extend to matrix *semigroups*  $S \subseteq K^{d \times d}$ , although in the generalizations of the Burnside–Schur Theorem, the restriction to finite generation is now essential, even in characteristic zero (see [Ste12] for a proof and a comprehensive list of earlier proofs). Extending Theorems 1.1 and 1.3 to semigroups is beyond the scope of the present work, but may be feasible and would allow us to remove the invertibility restriction in the WFA case (see Section 11).

In a related spirit to Theorem 1.1, Bernik, Mastnak, and Radjavi [BMR11] and Cigler and Drnovšek [CD13] have shown that irreducible semigroups  $S \subseteq \mathbb{C}^{d \times d}$  with *nonnegative diagonal entries* can be conjugated to a monomial semigroup.



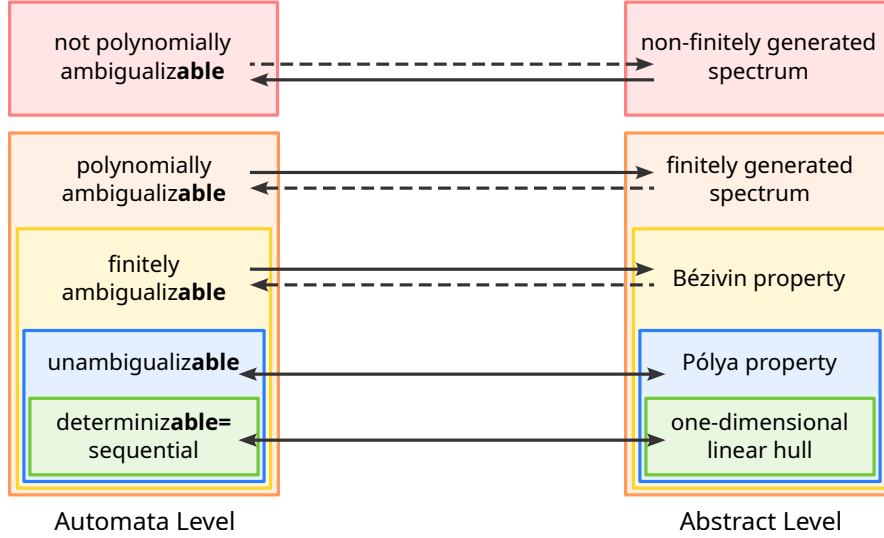


FIGURE 1. The natural complexity hierarchy for rational series (functions computable by WFA) based on ambiguity is reflected in arithmetic properties of the outputs, respectively the semigroup of a minimal linear representation. Full arrows represent unconditional implications, the implications of the dashed arrows are currently known in the invertible case (Theorem 1.6, Proposition 9.3, and [BS21]). The picture represents the case of algebraically closed fields, for simplicity.

**Ambiguity of WFA.** Let  $X$  be a finite alphabet and  $K$  a field. The fundamental Kleene–Schützenberger Theorem shows that a function  $f: X^* \rightarrow K$  is a *noncommutative rational series* if and only if it can be computed by a WFA (this holds even when  $K$  is just a semiring instead of a field) [BR11, Chapter 1] [Sak09, Chapter III.2].

WFA can equivalently be described through their linear representations. In this context, a *linear representation* is a tuple  $(u, \mu, v)$  of vectors  $u \in K^{1 \times d}$ ,  $v \in K^{d \times 1}$  and a monoid homomorphism  $\mu: X^* \rightarrow K$  such that  $f(w) = u\mu(w)v$  for every word  $w \in X^*$ . We say that a WFA (or its linear representation) is *invertible* if  $\mu(X^*) \subseteq \text{GL}_d(K)$ . Represented in a suitable way as a directed graph, WFA can be viewed as a computational model; as such they are a central object of study in theoretical computer science, we refer to [Sak09; DKV09; BR11] as starting points into the vast literature.

Among WFA, there is a natural complexity hierarchy, determined by how many successful runs a given word can have, leading to deterministic, unambiguous, finitely ambiguous, polynomially ambiguous, and exponentially ambiguous WFA (see Definition 9.1). The ambiguity of a given WFA does not depend on the weights; it can be elegantly characterized in terms of certain features of the underlying directed graph. These features can be algorithmically decided by classical work of Weber and Seidl [WS91].

However, a given rational series  $f: X^* \rightarrow K$  can be computed by many different WFA. Over a field, there is, up to a base change, a unique minimal one (minimal number of states, corresponding to minimal dimension of the linear representation). Unfortunately, the minimal WFA will typically *not* achieve a low ambiguity class, as the low ambiguity comes at the expense of increasing the number of states. One therefore faces the question: given a rational  $f: X^* \rightarrow K$ , represented by *some* WFA, is it decidable what the lowest ambiguity class of a WFA that recognizes  $f$  is? In particular, one can ask if  $f$  is computable by a *deterministic* WFA. This question has a long history for WFA over tropical semirings (see the survey [LS06] as a starting point). The general problem is still open in this setting [LS06, Problem 2], but *within* the class of polynomially ambiguous WFA the problem has since been resolved by Kirsten and Lombardy [KL09].

For fields, there was a long-standing conjecture of Reutenauer that was recently solved in [BS21], and that shows that computability by unambiguous WFA corresponds exactly to the rational series having the Pólya property. Together with the computability of the (linear) Zariski closure, established for matrix groups by Derksen and Jeandel [DJK05a] and for matrix semigroups by Hrushovski, Ouaknine, Pouly, and Worrell [Hru+18; Hru+23] (see also [Nos+22; BS23b]), this shows that both determinizability and unambigualizability over fields is decidable, resolving [LS06, Problem 1]. First complexity bounds were given by Benalioua, Lhote, and Reynier [BLR24], and in the subclass of polynomially weighted automata by Jecker, Mazowiecki, and Purser [JMP24].

We mention [Kir12; Col15; CP17; Dav+17; MR17; CP20; Kos22; Hru+23; Kos24] as further pointers into the recent literature on determinizability and unambigualizability of weighted automata. An interesting connection between rational series and one-dimensional topological theories was recently established by Im and Khovanov [IK22; IK24].

Assuming the WFA is invertible, our group-theoretic results yield a perfect correspondence between the best possible ambiguity class of a rational  $f: X^* \rightarrow K$ , and corresponding arithmetic properties on the values, respectively, the semigroup of a minimal linear representation, as depicted in Fig. 1. Two WFA are equivalent if they compute the same series.

**Theorem 1.6.** *Let  $\mathcal{A} = (u, \mu, v)$  be a minimal linear representation of an invertible WFA over a field  $K$ , and let  $M \in \mathbb{Z}_{\geq 0}$ .*

- (1) *The WFA is equivalent to an  $M$ -ambiguous WFA if and only if there exists a finitely generated  $\Gamma \leq K^\times$  such that  $u\mu(X^*)v \subseteq M \cdot \Gamma_0$ . In particular, it is equivalent to a finitely ambiguous WFA if and only if the outputs form a Bézivin set.*
- (2) *The WFA is equivalent to a polynomially ambiguous WFA if and only if  $\mu(X^*)$  has finitely generated spectrum and  $K$  is power-splitting for  $\mu(X^*)$ .*

In fact, in these cases the equivalent automaton  $\widehat{\mathcal{A}}$  can be chosen in such a way that the transition matrices are monomial, respectively, block-triangular with monomial diagonal blocks (Proposition 9.4). In the special case  $M = 1$ , Theorem 1.6(1) recovers the main result of [BS21], when restricted to invertible WFA (no restriction is necessary in [BS21]).

Using Theorem 1.6, the complexity classes become decidable. (The restriction on the field can be loosened, see Section 10.)

**Corollary 1.7.** *Let  $\mathcal{A} = (u, \mu, v)$  be a linear representation of an invertible WFA over a number field  $K$ .*

- (1) *It is decidable whether  $\mathcal{A}$  is equivalent to a finitely ambiguous, respectively, a polynomially ambiguous WFA.*
- (2) *If  $\mathcal{A}$  is equivalent to a finitely ambiguous WFA, then the minimal ambiguity  $M$  of such a WFA is computable.*

It is also possible to compute an equivalent finitely ambiguous, respectively, polynomially ambiguous WFA if it exists. In the finitely ambiguous case, this WFA can be computed to be of the minimal possible ambiguity  $M$ . While we show that the problems are decidable, our algorithms are presumably not particularly efficient or practical, as they are based on the computation of the (linear) Zariski closure of a finitely generated group of matrices. In several special cases better algorithms are known.

**Proofs.** In [BS21] unit equations were used together with the linear Zariski topology to great effect in the unambiguous (Pólya) case of WFA. Because the Pólya property depends on the initial and terminal weights of a WFA, it does not reduce to a purely (semi)group-theoretic question.

Unfortunately, the key lemma [BS21, Lemma 4.2] does not appear to generalize in a suitable form to attack Bézivin set problems we face here. Instead, we need to inspect the matrix groups much more closely, applying unit equations in a novel but very natural way, in the key results Proposition 6.4 and Corollary 6.5.

The heavy lifting for the Bézivin case is done in Sections 6 and 7. The heart of the arguments is contained in Proposition 6.4 and Lemmas 6.6 and 6.8. The finitely generated spectrum case is dealt with in Section 8. As we hope the results may be useful to people with a variety of mathematical backgrounds, we strive to provide detailed and mostly self-contained proofs, where this is reasonably possible. In contrast to [BS21], we avoid using the height machinery also in positive characteristic; considerations using the natural density in subsets of the integers end up being sufficient.

**Acknowledgements.** We thank Ganna (Anja) Kudryavtseva for participating in extensive discussions that helped flesh out the main arguments in this paper. We also thank

Laure Daviaud, Urban Jezernik, Filip Mazowiecki, and David Purser for discussions and Janez Bernik for making us aware of the papers [Cor+22b; Cor+22a; Cor+23]. Additional thanks go to Pietro Corvaja for confirming that the arguments in [Cor+23] can also be used to deduce our version of Theorem 1.1 and to Andrei Rapinchuk for pointing out that Corollary 1.4 has been observed before.

Smertnig was supported by the Slovenian Research and Innovation Agency (ARIS) program P1-0288. Part of the research was conducted during a research stay of Puch at IMFM in Ljubljana, funded by ARIS program P1-0288 and Polish National Science Centre SONATA BIS-12 grant number 2022/46/E/ST6/00230.

## 2. NOTIONS AND PRELIMINARIES

For  $m, n \in \mathbb{Z}_{\geq 0}$ , we write  $[m, n] := \{x \in \mathbb{Z} : m \leq x \leq n\}$  for the discrete interval and set  $[n] := [1, n]$ . If  $K$  is a field, we denote by  $\mu_n(K)$  the group of all  $n$ -th roots of unity, by  $\mu_n^*(K)$  the set of primitive  $n$ -th roots of unity, and set  $\mu(K) := \bigcup_{n \geq 1} \mu_n(K)$ .

It will be useful to make use of the language of linear representations of groups (and sometimes semigroups). Let  $K$  be field, embedded in a fixed algebraic closure  $\overline{K}$ , and let  $V$  be a finite-dimensional vector space over  $K$ . A  $(K)$ -representation of a semigroup  $S$  is a semigroup homomorphism  $\rho: S \rightarrow \text{End}(V)$ , or, more explicitly, it is the pair  $(V, \rho)$ . We only consider finite-dimensional representations. A representation is *faithful* if  $\rho$  is injective. A *morphism* (or  *$S$ -equivariant map*) between representations  $(V, \rho)$  and  $(W, \sigma)$  is a linear map  $\Psi: V \rightarrow W$  such that  $\Psi\rho(s) = \sigma(s)\Psi$  for all  $s \in S$ .

*Almost exclusively* we will only consider representations  $S \rightarrow \text{GL}(V)$  with invertible images (which is of course a severe restriction), with Proposition 6.4 and Corollary 6.5 being the notable exceptions. For the problems we consider, this will turn out to essentially be equivalent to considering group representations  $G \rightarrow \text{GL}(V)$ . However, the more general statements are more natural in the applications to automata theory.

A matrix semigroup  $S \subseteq \text{GL}_d(K)$  is *Bézivin* if there is a Bézivin set containing every entry of every  $A \in S$ ; the Pólya property for  $S$  is defined analogously.

**Lemma 2.1.** *If  $S \subseteq \text{GL}_d(K)$  is a Bézivin semigroup and  $T \in \text{GL}_d(K)$ , then also  $TST^{-1}$  is Bézivin.*

*Proof.* Let  $\Gamma \leq K^\times$  be finitely generated and let  $M \geq 0$  be such that all entries of all matrices  $A \in S$  are contained in  $M \cdot \Gamma_0$ . Suppose  $T = (t_{ij})_{1 \leq i, j \leq d}$  and  $T^{-1} = (t'_{ij})_{1 \leq i, j \leq d}$ . If  $A = (a_{ij})_{1 \leq i, j \leq d} \in S$ , then the  $rs$ -entry of  $TST^{-1}$  is  $\sum_{i, j=1}^d t_{ri} a_{ij} t'_{js}$ . Taking  $\Gamma'$  to be the group generated by  $\Gamma$  together with all  $t_{ij}$  and  $t'_{ij}$ , we see that all elements of  $TST^{-1}$  have their entries in  $d^2 M \cdot \Gamma'_0$ .  $\square$

Thus, the Bézivin property is independent of the choice of basis (although the constant  $M$  and the group  $\Gamma$  may change). Given a representation  $\rho: S \rightarrow \text{End}(V)$  of a semigroup  $S$ , we therefore say that  $\rho$  has **finitely generated spectrum**, respectively the Bézivin property, if this is the case for  $\rho(S)$  for some choice of basis on  $V$  (and hence for all choices of bases on  $V$ ).

A semigroup  $S \subseteq \text{GL}(V)$  is **irreducible** if it acts irreducibly on  $V$ , that is, there is no nonzero proper  $S$ -invariant subspace of  $V$ . The semigroup is **absolutely irreducible** if it is irreducible when considered over  $\overline{K}$ , that is, as a subsemigroup of  $\text{GL}(\overline{K} \otimes_K V)$ .

The notion of a steady endomorphism will be of central importance in our proofs.

**Definition 2.2.** *An endomorphism  $A \in \text{End}(V)$  is **steady** if*

- (1) *every subspace  $U \subseteq V$  that is  $A^n$ -invariant for some  $n \geq 1$  is  $A$ -invariant, and*
- (2) *all eigenvalues of  $A$  are contained in  $K$ .*

The following basic properties of steadiness are easily deduced using the Jordan normal form (see Section 6 below for proofs).

**Lemma 2.3.** *Let  $A \in \text{End}(V)$ .*

- (1) *There exists some  $n \geq 1$  such that  $A^n$  is steady if and only if there exists some  $m \geq 1$  such that  $\lambda^m \in K$  for all eigenvalues  $\lambda \in \overline{K}$  of  $A$ .*
- (2) *If  $A$  is steady and  $\lambda, \mu \in K$  are eigenvalues with  $\lambda/\mu$  a root of unity, then  $\lambda = \mu$ .*
- (3) *If  $\text{char } K > 0$ , then every steady  $A \in \text{End}(V)$  is diagonalizable.*

Since property (1) of Lemma 2.3 will be important, we make the following definitions.

**Definition 2.4.** *Let  $S \subseteq \text{GL}(V)$  be a semigroup.*

- (1) *A field  $L \supseteq K$  is **power-splitting** for  $S$  if for every  $A \in S$  and every eigenvalue  $\lambda \in \overline{K}$  of  $A$ , there exists some  $m \geq 1$  such that  $\lambda^m \in K$ .*
- (2) *A field  $L \supseteq K$  is **uniformly power-splitting** for  $S$  if there is an  $N \geq 1$  such that  $\lambda^N \in K$  for every eigenvalue  $\lambda$  of every  $A \in S$ .*

Given a representation  $\rho: S \rightarrow \text{GL}(V)$  every decomposition  $V = V_1 \oplus \cdots \oplus V_r$  gives rise to a block-decomposition of  $\rho$ . We denote by  $\rho_{ij}$  the corresponding map  $\rho_{ij}: S \rightarrow \text{Hom}_K(V_j, V_i)$ . Each diagonal block  $\rho_{ii}$  is of course itself a representation of  $S$ .

**Definition 2.5.** *A semigroup representation  $\rho: S \rightarrow \text{GL}(V)$  is*

- (1) ***monomial** if there exists a decomposition  $V = V_1 \oplus \cdots \oplus V_d$  into one-dimensional spaces, such that for every  $s \in S$  and every  $j \in [d]$ , there exists  $i \in [d]$  such that  $\rho(s)V_j \subseteq V_i$ .*

- (2) (*upper*) *block-triangular with monomial diagonal blocks* if there exists a decomposition  $V = V_1 \oplus \cdots \oplus V_r$  such that  $\rho_{ij} = 0$  for  $i < j$  and  $\rho_{ii}$  is monomial for all  $i \in [r]$ .

In other words, a representation is *monomial* if and only if, after a suitable choice of basis, every matrix of the representation has exactly one nonzero entry in each row and each column. A representation is *block-triangular with monomial diagonal blocks* if, after a suitable choice of basis, every matrix has that shape.

In general, locally Bézivin representations (even of finitely generated semigroups) will not be monomial. For instance, every finite matrix group is trivially Bézivin, but not necessarily monomial. For this reason, the following weaker notions are useful.

**Definition 2.6.** *Let  $\rho: S \rightarrow \text{GL}(V)$  be a semigroup representation. A weakly epimonomial decomposition (of  $V$  for  $\rho$ ) is a vector space decomposition  $V = V_1 \oplus \cdots \oplus V_r$  such that*

- (1) *every  $V_i$  is contained in an eigenspace of all steady  $\rho(s)$  ( $s \in S$ ),*
- (2) *every  $\rho(t)$ , with  $t \in S$ , permutes the spaces  $V_1, \dots, V_r$ , and*
- (3) *the field  $K$  is power-splitting for  $\rho(S)$ .*

*The representation  $\rho$  is weakly epimonomial if there exists a weakly epimonomial decomposition.*

In different terminology, the spaces  $V_1, \dots, V_r$  are an imprimitive system for  $\rho$ , consisting of subspaces of joint-eigenspaces of the steady elements. Given a weakly monomial decomposition, we call

$$D := \{ \rho(s) \in \rho(S) : \text{every } V_i \text{ is contained in an eigenspace of } \rho(s) \}$$

the *diagonal* of the  $\rho$  with respect to the decomposition. The diagonal  $D$  is a commutative subsemigroup of  $\rho(S)$  (subgroup if  $S = G$  is a group), and it embeds into  $(K^\times)^r$ .

After a choice of basis that is compatible with the decomposition  $V = V_1 \oplus \cdots \oplus V_r$ , the set  $D$  is indeed precisely the subset of diagonal matrices in  $\rho(S)$  that are scalar on each  $V_i$ . Restricting a weakly epimonomial representation to a subsemigroup  $S'$ , trivially gives again a weakly epimonomial representation of  $S'$ .

We will later see: if  $\rho: S \rightarrow \text{GL}(V)$  is weakly epimonomial and  $G = \langle \rho(S) \rangle$  is the group generated by  $\rho(S)$ , then  $G \hookrightarrow \text{GL}(V)$  is weakly epimonomial as well (Lemma 7.1).

**Definition 2.7.** (1) *Let  $\rho: S \rightarrow \text{GL}(V)$  be a semigroup representation with diagonal  $D$ , and let  $G = \langle \rho(S) \rangle$ . A weakly epimonomial decomposition for  $\rho$  is *epimonomial* if  $G/\langle D \rangle$  is finite.*

- (2) *A semigroup representation  $\rho: S \rightarrow \text{GL}(V)$  is *epimonomial* if it has an epimonomial decomposition.*

If  $S \subseteq \mathrm{GL}(V)$  is a subsemigroup, we apply all of this terminology to  $S$  as well, via the representation  $S \hookrightarrow \mathrm{GL}(V)$  arising from the inclusion. Note that these notions depend on the representation, respectively, the particular embedding of  $S$  into  $\mathrm{GL}(V)$ , and as such are not properties of the abstract semigroup  $S$  (or abstract group  $G$ ).

If a group representation  $\rho: G \rightarrow \mathrm{GL}(V)$  is epimonomial with diagonal  $D$ , it is easy to see that first restricting to  $\rho^{-1}(D)$ , and then considering the induced representation on  $G$ , gives a monomial representation for  $G$ . Specifically, we have the following (see Lemma 7.11 for a proof).

**Proposition 2.8.** *If  $\rho: S \rightarrow \mathrm{GL}(V)$  is an epimonomial semigroup representation, then there exists a monomial representation  $\widehat{\rho}: S \rightarrow \mathrm{GL}(\widehat{V})$  and an epimorphism  $(\widehat{V}, \widehat{\rho}) \twoheadrightarrow (V, \rho)$ .*

### 3. STATEMENTS OF MAIN RESULTS

We now state our main results in full generality. Let  $K$  be a field and  $V$  a finite-dimensional vector space. Our first main theorem characterizes locally Bézivin representations (over any field).

**Theorem 3.1.** *Let  $\rho: S \rightarrow \mathrm{GL}(V)$  be a semigroup representation. Then the following statements are equivalent.*

- (a) *There exists a basis of  $V$  with respect to which  $\rho$  is locally Pólya.*
- (b) *The representation  $\rho$  is locally Bézivin.*
- (c) *The representation  $\rho_L: S \rightarrow \mathrm{GL}(L \otimes_K V)$  is locally Bézivin for some extension field  $L/K$ , and  $K$  is a power-splitting field for  $\rho_L(S)$ .*
- (d) *The representation  $\rho$  is weakly epimonomial.*
- (e) *The representation  $\rho$  is locally epimonomial.*

For finitely generated  $S$ , we will get the following corollary.

**Corollary 3.2.** *If  $S$  is finitely generated and  $G = \langle \rho(S) \rangle$  is the group generated by  $\rho(S)$ , then the following statements are equivalent.*

- (a) *There exists a basis of  $V$  with respect to which  $\rho$  is Pólya.*
- (b) *The representation  $\rho$  is Bézivin.*
- (c) *The representation  $\rho_L: S \rightarrow \mathrm{GL}(L \otimes_K V)$  is Bézivin for some extension field  $L/K$ , and  $K$  is a power-splitting field for  $\rho_L(S)$ .*
- (d) *The representation  $\rho$  is epimonomial.*



- (e) *The group  $G$  is virtually simultaneously diagonalizable (over  $K$ ).*
- (f) *The representation  $\rho$  is the epimorphic image of a monomial representation of  $S$ .*

If  $S$  is not finitely generated, we still get a similar result in characteristic 0.

**Corollary 3.3.** *If  $\text{char } K = 0$  and  $G = \langle \rho(S) \rangle$ , the following statements are equivalent.*

- (a) *The representation  $\rho$  satisfies the equivalent conditions of Theorem 3.1 and  $K$  is uniformly power-splitting for  $\rho(S)$ .*
- (b) *The group  $G$  is virtually simultaneously diagonalizable (over  $K$ ).*
- (c) *The representation  $\rho$  is the epimorphic image of a monomial representation of  $S$ .*

The proof of Corollary 3.3 makes use of the Jordan–Schur Theorem and the variant of the Burnside–Schur Theorem that shows that a linear torsion group of finite exponent over a field of characteristic zero is finite.

For representations with locally finitely spectrum, we obtain the following.

**Theorem 3.4.** *Let  $\rho: S \rightarrow \text{GL}(V)$  be a semigroup representation. Then the following statements are equivalent.*

- (a) *The representation  $\rho$  has locally finitely generated spectrum and  $K$  is power-splitting for  $\rho(S)$ .*
- (b) *The representation  $\rho$  is block-triangular with locally epimonomial diagonal blocks.*

Combining Theorem 3.4 with Corollaries 3.2 and 3.3 we obtain the following.

**Corollary 3.5.** *If*

- *the semigroup  $S$  is finitely generated, or*
- *the field  $K$  is uniformly power-splitting for  $\rho(S)$  and  $\text{char } K = 0$ ,*

*then the statements of Theorem 3.4 are in addition equivalent to*

- (c) *There exists a representation  $\widehat{\rho}: S \rightarrow \text{GL}(\widehat{V})$  that is block-triangular with monomial diagonal blocks and an epimorphism  $(\widehat{V}, \widehat{\rho}) \twoheadrightarrow (V, \rho)$ .*

Here, it is *not* possible to remove the power-splitting hypothesis when  $S$  is finitely generated (Example 4.1).

Theorems 1.1 and 1.3 and Corollary 1.4 are now easily deduced.

*Proof of Theorem 1.1.* We consider the faithful representation  $\rho: G \hookrightarrow \text{GL}_d(K)$ . If  $G$  is finitely generated, then the equivalences, without the uniform power-splitting hypothesis in (a) and (b), hold by Corollary 3.2 applied to  $\rho$ . Since  $K$  is trivially uniformly power-splitting for an epimonomial  $G$ , also the uniform power-splitting hypothesis holds.

Now suppose  $\text{char } K = 0$ . Then the claims hold by Corollary 3.3.  $\square$

*Proof of Theorem 1.3.* Again consider  $\rho: G \hookrightarrow \text{GL}_d(K)$ . If  $G$  has locally finitely generated spectrum and  $K$  is uniformly power-splitting for  $G$ , then Corollary 3.5 yields the claim.

In the converse direction, to apply Corollary 3.5 again, it suffices to observe that  $K$  is uniformly power-splitting for  $G$ . Let  $V := K^{d \times 1}$  and let  $\widehat{\rho}: G \rightarrow \text{GL}(\widehat{V})$  be a block-triangular representation with monomial diagonal blocks such that  $\rho$  is an epimorphic image of  $\widehat{\rho}$ . Let  $\pi: \widehat{V} \rightarrow V$  denote the  $G$ -equivariant epimorphism.

There exists  $N \geq 1$  such that  $\widehat{\rho}(g)^N$  is upper-triangular for each  $g \in G$ . Thus, the field  $K$  is uniformly power-splitting for  $\widehat{\rho}$ . Considered as a vector space epimorphism, the map  $\pi$  splits, so that  $\widehat{V} \cong \ker(\pi) \oplus V$ . Since  $\ker(\pi)$  is  $G$ -invariant, the elements of  $\widehat{\rho}(G)$  have a block-triangular form with respect to this splitting, and the spectrum of  $\rho(g)$  is a subset of the spectrum of  $\widehat{\rho}(g)$ . Hence, the field  $K$  is also uniformly power-splitting for  $G$ .  $\square$

*Proof of Corollary 1.4.* Suppose  $G$  has locally finitely generated spectrum. Since  $\overline{K}$  is trivially uniformly power-splitting for  $\overline{\rho}(G)$ , there exists a faithful block-triangular representation of  $G$  over  $\overline{K}$  with monomial diagonal blocks (Corollary 3.5). The finite-index subgroup whose diagonal blocks are themselves diagonal is solvable. Hence, the group  $G$  is virtually solvable. Conversely, every virtually solvable group has locally finitely generated spectrum (Lemma 8.2).  $\square$

We will prove Theorem 3.1 and Corollaries 3.2 and 3.3 in Section 7, after doing the heavy-lifting in Section 6. Theorem 3.4 and Corollary 3.5 will be proved in Section 8.

#### 4. GUIDING EXAMPLES AND SPECIAL CASES

We gather some easy examples that illustrate the conditions and limitations of the theorems as well as some interesting special cases. The section provides some context for the (uniform) power-splitting conditions and the limitations to characteristic zero in some of our theorems. The section can otherwise safely be skipped.

**4.1. One-letter WFA: Linear Recurrence Sequences (LRS).** Let  $K$  be a field. It is well known that a  $K$ -valued sequence  $(a(n))_{n \geq 0}$  is an LRS if and only if there exist  $d \geq 0$ , a matrix  $A \in K^{d \times d}$ , and vectors  $u \in K^{1 \times d}$  and  $v \in K^{d \times 1}$  such that  $a(n) = uA^n v$  for all  $n \geq 0$ . Equivalently, the generating function  $\sum_{n=0}^{\infty} a(n)x^n$  is a rational function  $P/Q \in K(x)$  with  $Q(0) = 1$ . Equivalently, the sequence  $(a(n))_{n \geq 0}$  is generated by a one-letter weighted finite automaton: if  $X = \{x\}$  is the alphabet, then  $a(n)$  is the output of the word  $x^n$ .

*Example 4.1.* Let  $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \in \text{GL}_2(\mathbb{Q})$ , let  $u = (1, 0)$ , and let  $v = (0, 1)^T$ . Then  $F_n := uA^n v$  is the sequence of Fibonacci numbers. The eigenvalues of  $A$  are  $\varphi_{\pm} := \frac{1 \pm \sqrt{5}}{2}$ . Since no

power of  $\varphi_+$  is in  $\mathbb{Q}$  (for instance, because  $\varphi_+^n = \varphi_+ F_n + F_{n-1}$  and  $\varphi_+ \notin \mathbb{Q}$ ), the group generated by  $A$  is not power-splitting over  $\mathbb{Q}$ . Hence, it does not even have a block-triangular representation with monomial diagonal blocks. On the other hand, over the field  $L = \mathbb{Q}(\varphi_+)$ , the matrix  $A$  is diagonalizable. Thus, the group  $\langle A \rangle$  and the sequence  $F_n = \frac{1}{\sqrt{5}}\varphi_+^n - \frac{1}{\sqrt{5}}\varphi_-^n$  are Bézivin over  $L$ . This shows that the power-splitting condition in the main theorems cannot be removed.

Consider an arbitrary LRS of the form  $a(n) = uA^n v$ . By dropping finitely many initial terms, we can assume  $A \in \mathrm{GL}_d(K)$ . We also assume that  $d$  is minimal for the given LRS. If  $K$  is algebraically closed of characteristic 0, then  $(a(n))_{n \geq 0}$  is Bézivin if and only if  $A$  is diagonalizable: this follows from a theorem of Bézivin [Béz86, Théorème 4] or as a special case of Proposition 9.3(2) together with Theorem 1.1.

Similarly, for an arbitrary field  $K$ , the sequence  $(a(n))_{n \geq 0}$  is Bézivin if and only if some power of  $A$  is diagonalizable over  $K$ . In this case (and only then), one can find a linear representation  $(\widehat{u}, \widehat{A}, \widehat{v})$  of the same LRS  $(a(n))_{n \geq 0}$  with a monomial matrix  $\widehat{A}$  by Proposition 9.4. The LRS case, corresponding to one-letter WFA, was also recently studied in detail by Kostolányi [Kos22; Kos24].

By passing to an extension field (in which we can put  $A$  into Jordan normal form) an LRS can always be computed by a polynomially ambiguous WFA, while computability by a finitely ambiguous WFA over an extension field corresponds to diagonalizability of  $A$ . In particular, the exponentially ambiguous case only appears for alphabets with at least two letters.

**4.2. Non-finitely generated groups.** For non-finitely generated groups two issues appear: one needs to assume *uniform* power-splitting and characteristic zero is essential.

*Example 4.2.* (1) Consider the embedding  $\mathbb{C}^\times \rightarrow \mathrm{GL}_2(\mathbb{R})$  given by  $a + bi \mapsto \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$ . Restricting to  $\mu(\mathbb{C})$ , we get a locally finite, and hence locally Bézivin, representation of  $\mu(\mathbb{C})$  over  $\mathbb{R}$ . However, the group  $\mu(\mathbb{C})$  does not have a faithful monomial representation over  $\mathbb{R}$ : suppose  $j: \mu(\mathbb{C}) \rightarrow \mathrm{GL}_d(\mathbb{R})$  is a faithful monomial representation. Then  $j(\zeta)^{d!}$  is diagonal for each  $\zeta \in \mu(\mathbb{C})$ . Because  $j(\zeta)^{d!}$  has finite order, its diagonal entries must be  $\pm 1$ , showing that  $j(\zeta)^{2d!} = I$ . This contradicts the existence of elements of arbitrarily large order in  $\mu(\mathbb{C})$ . Of course, the field  $\mathbb{R}$  is power-splitting but not uniformly power-splitting for our original representation of  $\mu(\mathbb{C})$ .

(2) Fix  $n \geq 1$  and let  $R_n := \{z \in \mathbb{C}^\times : z^n \in \mathbb{R}\}$ . Then  $R_n$  is infinite abelian and  $R_n$  can be embedded into  $\mathrm{GL}_2(\mathbb{R})$  as a uniformly power-splitting locally Bézivin subgroup; indeed with the embedding as in (1), we have

$$R_n = \bigcup_{\zeta \in \mu_n(\mathbb{C})} \zeta \mathbb{R}^\times,$$

(identifying  $\mathbb{R}^\times \subseteq \mathrm{GL}_2(\mathbb{R})$  with the scalar matrices). Hence, the group  $R_n$  does have a monomial representation over  $\mathbb{R}$  (Corollary 3.3). Indeed, here  $\mathbb{R}^\times$  is simultaneously diagonal of finite index.

- (3) Let  $K$  be a field of positive characteristic  $p > 0$ . Consider the faithful representation  $\rho: (K, +) \mapsto \mathrm{GL}_2(K)$  defined by

$$\rho(x) = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}.$$

Since  $\mathrm{im} \rho$  is an abelian  $p$ -torsion group, it is locally finite. In particular, therefore  $\rho$  is locally Bézivin.

Suppose that  $\pi: (K, +) \rightarrow \mathrm{GL}(V)$  is a faithful monomial representation on a vector space of dimension  $d$ . Then there is a group homomorphism  $\psi: \mathrm{im} \pi \rightarrow \mathfrak{S}_d$  to the symmetric group, mapping an element of  $\mathrm{im} \pi$  to its underlying permutation. Let  $D := \pi^{-1}(\ker(\psi))$ . Then  $D$  maps to diagonal matrices under  $\pi$ , hence  $D \hookrightarrow (K^\times)^d$ . Now  $D$  is  $p$ -torsion, and hence embeds into the  $p$ -torsion subgroup of  $(K^\times)^d$ , that is  $D \hookrightarrow \mu_p(K)^d$ . Thus, the group  $D$  is finite, and also of finite index in  $K$ . Hence, also  $K$  is finite.

Thus, over an infinite field of positive characteristic, the group  $(K, +)$  has a faithful locally Bézivin representation and  $K$  is uniformly power splitting for that representation, but  $(K, +)$  has no faithful monomial representations, showing that Corollary 3.3 does not hold in general in positive characteristic.

**4.3.  $k$ -regular and  $k$ -automatic sequences.** In the 90s, Allouche and Shallit introduced the notion of a  $k$ -regular sequence as a generalization of a  $k$ -automatic sequence [AS03] (here  $k \in \mathbb{Z}_{\geq 2}$  is an arbitrary but fixed parameter). One of the equivalent definitions is the following: a sequence  $(a(n))_{n \geq 0}$  with values in the field  $K$  is  $k$ -regular if there exist  $d \geq 0$ , matrices  $A_0, \dots, A_{k-1} \in K^{d \times d}$  and vectors  $u \in K^{1 \times d}$ ,  $v \in K^{d \times 1}$  such that,

$$a\left(\sum_{i=0}^l n_i k^i\right) = u A_{n_0} \cdots A_{n_l} v,$$

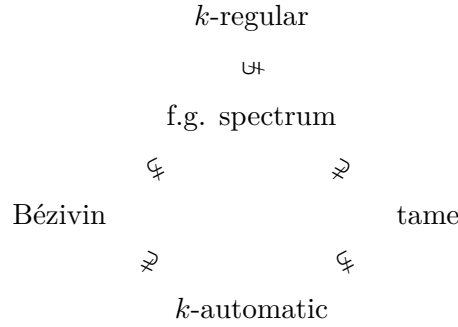
whenever  $l \geq 0$ , and  $n_0, \dots, n_l \in [0, k-1]$ . In other words, there is a WFA on the alphabet of  $k$ -adic digits  $X = [0, k-1]$ , such that whenever a word  $w \in X^*$  represents the number  $n$  in base  $k$ , then  $a(n)$  is the output of the WFA for the word  $w$ .<sup>3</sup> We suppose that the dimension  $d$  is minimal for the sequence, that is, the WFA is minimal.

<sup>3</sup>The word  $w$  may have leading zeroes, but all such words representing the same number produce the same output, in this definition. In defining the class of  $k$ -regular sequences, it does not actually matter whether the WFA reads the base  $k$  representation from least to most significant digit or the other way around.

The class of  $k$ -regular sequences contains the  $k$ -automatic sequences as a subset; conversely, the generating function of any  $k$ -regular sequence is  $k$ -Mahler (see [ABS23] for definitions).

Let  $K = \overline{\mathbb{Q}}$ . Then the class of  $k$ -Mahler functions can be separated into five distinct classes based on the asymptotic growth of their coefficients, as measured by the logarithmic Weil height [ABS23]. The three slowest growing classes correspond to  $k$ -regular sequences:  $O(\log n)$ ,  $O(\log \log n)$  and  $O(1)$ . Here the  $O(\log \log n)$  class correspond to the matrix semigroup  $S$  generated by  $A_0, \dots, A_{k-1}$  being **tame**, that is, the semigroup has locally finite spectrum. The class  $O(1)$  corresponds to  $k$ -automatic sequences, which is equivalent to  $S$  being finite, and is trivially Bézivin.

Let us restrict to the consideration of  $k$ -regular sequences for which the matrices  $A_0, \dots, A_{k-1}$  are invertible. Then the classification by the Bézivin property and finitely generated spectrum, established in this paper, interfaces with the growth classification of [ABS23] to yield the following picture.



**4.4. Bounded generation.** A motivation in [Cor+22b; Cor+22a; Cor+23] is the study of the bounded generation (BG) property for linear groups. A group  $G$  has BG, if there exist  $A_1, \dots, A_n \in G$  such that  $G = \langle A_1 \rangle \cdots \langle A_n \rangle$ . One of the main results of Corvaja, Demeio, Rapinchuk, Ren, and Zannier [Cor+23, Theorem 1.13] is that, in *characteristic zero*, a linear group  $G$ , all of whose elements are diagonalizable, has BG if and only if  $G$  is finitely generated and the connected component of its Zariski closure is a torus. In particular, such a group is virtually abelian.

This is based on the observation that a BG group whose elements are diagonalizable is a PEP set [Cor+23, Example 1.4]; the same observation shows that a BG group is Bézivin so that Theorem 1.1 recovers the result, but this does not give anything new.

In *positive characteristic*, Abért, Lubotzky, and Pyber had already shown before that every linear BG group is virtually abelian [ALP03, Theorem 1.1]. Here it is not necessary to assume that the elements are diagonalizable. Since our Theorem 1.1 also holds in positive characteristic, we obtain a strengthening of [ALP03, Theorem 1.1].

**Corollary 4.3.** *Suppose  $K$  has positive characteristic. A group  $G \leq \mathrm{GL}(V)$  has bounded generation (BG) if and only if it is finitely generated and virtually simultaneously diagonalizable (over  $\overline{K}$ ).*

*Proof.* First suppose that  $G$  has BG. We identify  $\mathrm{GL}(V) \cong \mathrm{GL}_d(K)$  and may assume  $K = \overline{K}$ . Say  $G = \langle A_1 \rangle \cdots \langle A_n \rangle$  with  $A_i \in \mathrm{GL}_d(\overline{K})$ . Then  $G$  is obviously finitely generated by  $A_1, \dots, A_n$ . For each  $A_i$ , choose  $T_i \in \mathrm{GL}_d(\overline{K})$  such that  $B_i := T_i A_i T_i^{-1}$  is in Jordan normal form. Since the characteristic is positive, there exists some power  $q = p^e$  such that  $B_i^q$  is diagonal for all  $i \in [n]$  (see, for instance, the proof of Lemma 2.3). Now

$$G = \bigcup_{r_1, \dots, r_n \in [0, q-1]} \bigcup_{m_1, \dots, m_n \in \mathbb{Z}} (T_1^{-1} B_1^{qm_1} B_1^{r_1} T_1) \cdots (T_n^{-1} B_n^{qm_n} B_n^{r_n} T_n).$$

For fixed  $r_1, \dots, r_n$ , these matrices have their entries in some Bézivin set (over  $\overline{K}$ ), because the  $B_i^q$  are diagonal and everything else is constant. But then the finite union  $G$  is also Bézivin. Now Theorem 1.1 implies the claim.

For the converse direction, it is sufficient to note that  $G$  is virtually abelian, and every virtually abelian group is easily seen to have BG.  $\square$

## 5. UNIT EQUATIONS: LINEAR EQUATIONS OVER FINITELY GENERATED GROUPS

Throughout this section, let  $K$  be a field and let  $\Gamma \leq K^\times$  be a *finitely generated* subgroup. The theory of *unit equations* concerns itself with the study of solutions  $(x_0, \dots, x_n) \in \Gamma^{n+1}$  to a linear homogeneous equation

$$a_0 X_0 + \cdots + a_n X_n = 0 \quad (a_i \in K). \quad (1)$$

The restriction to solutions in  $\Gamma$  turns this into a problem in Diophantine number theory. Obviously, any  $\Gamma$ -multiple of a solution is again a solution, so it makes sense to consider projective solutions  $(x_0 : \cdots : x_n) \in \mathbb{P}^n(\Gamma)$  which have a representative with  $x_0, \dots, x_n \in \Gamma$ . If, for a solution  $\mathbf{x} = (x_0 : \cdots : x_n)$ , there is a subset  $\emptyset \neq I \subsetneq [0, n]$  such that  $\sum_{i \in I} a_i x_i = 0$ , then  $\mathbf{x}$  is *degenerate*. Otherwise, the solution  $\mathbf{x}$  is *non-degenerate*.

In a degenerate solution, one can multiply subsums by different scalars, potentially generating infinitely many solutions. However, for non-degenerate solutions and  $\mathrm{char} K = 0$ , we have the following celebrated theorem. It was proven independently by Evertse [Eve84] and van der Poorten–Schlickewei [PS82] for number fields, and extended to fields of characteristic zero in [PS91]. The theorem can also be deduced from Schmidt’s subspace theorem. See [EG15, Chapter 6] or [BG06, Theorem 7.4.1] for details.

**Theorem 5.1** (Main Theorem on Unit Equations). *Suppose  $\mathrm{char} K = 0$ . Then Eq. (1) has only finitely many non-degenerate solutions in  $\mathbb{P}^n(\Gamma)$ .*

Now suppose  $\text{char } K = p > 0$ . Then the conclusion of the theorem does not hold, because the Frobenius automorphism can generate infinitely many solutions. However, these solutions can still be partitioned into finitely many orbits by a theorem of Derksen and Masser [DM12, Theorem 3], that we describe in the following. Let  $\sqrt{\Gamma} \leq K^\times$  be the group of all  $\gamma \in K^\times$  for which there exists  $n \geq 1$  with  $\gamma^n \in \Gamma$ . A  $\sqrt{\Gamma}$ -automorphism is a map

$$\psi: \mathbb{P}^n(K) \rightarrow \mathbb{P}^n(K), (x_0 : \dots : x_n) \mapsto (\alpha_0 x_0 : \dots : \alpha_n x_n)$$

with  $\alpha_0, \dots, \alpha_n \in \sqrt{\Gamma}$ . For  $q$  a power of  $p$ , let

$$\varphi_q(x_0 : \dots : x_n) := (x_0^q : \dots : x_n^q).$$

For  $\sqrt{\Gamma}$ -automorphisms  $\psi_1, \dots, \psi_k$  and  $\mathbf{y} \in \mathbb{P}^n(K)$  let

$$[\psi_1, \dots, \psi_k]_q \{\mathbf{y}\} := \left\{ (\psi_1^{-1} \varphi_q^{e_1} \psi_1)(\psi_2^{-1} \varphi_q^{e_2} \psi_2) \dots (\psi_k^{-1} \varphi_q^{e_k} \psi_k)(\mathbf{y}) : e_1, \dots, e_k \geq 0 \right\}. \quad (2)$$

Thus, the set  $[\psi_1, \dots, \psi_k]_q \{\mathbf{y}\}$  is obtained from the point  $\mathbf{y}$  by applying powers of Frobenius automorphisms, twisted by suitable  $\sqrt{\Gamma}$ -automorphisms.

**Theorem 5.2** ([DM12, Theorem 3]). *Let  $\text{char } K = p > 0$ . If  $\sqrt{\Gamma}$  is a finitely generated group and  $S \subseteq \mathbb{P}(\Gamma)$  is the set of non-degenerate solutions of Eq. (1), then there exists some power  $q$  of  $p$  such that  $S$  is contained in a finite union of sets of the form as in Eq. (2) (with  $k \leq n - 1$ ).*

If  $K$  is a finitely generated field and  $\Gamma$  is finitely generated, then  $\sqrt{\Gamma}$  is also finitely generated (see, for instance, [BS21, Lemma 4.1] for a proof of this well-known fact). When applying Theorem 5.2, we will always be able to work in a finitely generated field.

An easy consequence of Theorems 5.1 and 5.2 is the following useful lemma.

**Lemma 5.3.** *Let  $\lambda \in K$  and let  $p = \text{char } K \geq 0$ . Suppose there exist  $a_0, \dots, a_M \in K^\times$  such that*

$$a_0 \lambda^n \in a_1 \Gamma_0 + \dots + a_M \Gamma_0$$

*for infinitely many  $n \in \mathbb{Z} \setminus p\mathbb{Z}$ . Then there exists  $m \geq 1$  such that  $\lambda^m \in \Gamma_0$ .*

*Proof.* We assume  $\lambda \neq 0$ , as the claim is trivially true otherwise. Working in the field generated by  $a_0, \dots, a_M$  and  $\Gamma$  over the prime field of  $K$ , we may also assume that  $K$  is finitely generated. By assumption there exist an infinite set  $\Omega \subseteq \mathbb{Z} \setminus p\mathbb{Z}$  and maps  $\gamma_1, \dots, \gamma_M: \Omega \rightarrow \Gamma_0$  such that  $a_0 \lambda^n - a_1 \gamma_1(n) - \dots - a_M \gamma_M(n) = 0$  for all  $n \in \Omega$ . We view the tuples  $(\lambda^n, \gamma_1(n), \dots, \gamma_M(n))$  as solutions to a unit equation in  $M + 1$  variables over the group generated by  $\Gamma \cup \{\lambda\}$ . Since each solution can be partitioned into non-degenerate solutions of subequations, and there are only finitely many ways to partition the set  $[0, M]$ , there exists a subset  $I \subseteq [M]$  and an infinite set  $\Omega' \subseteq \Omega$  such that  $a_0 \lambda^n - \sum_{i \in I} a_i \gamma_i(n) = 0$



is non-degenerate for all  $n \in \Omega'$ . Because  $a_0, \lambda \neq 0$ , necessarily  $I \neq \emptyset$ . After renumbering the  $\gamma_i$ , without restriction  $I = [l]$  with  $l \in [M]$ .

*First suppose*  $\text{char } K = 0$ . Theorem 5.1 implies that  $\{\lambda^n/\gamma_1(n) : n \in \Omega'\}$  is finite. Hence, there are  $n \neq n' \in \Omega'$  with  $\lambda^n/\gamma_1(n) = \lambda^{n'}/\gamma_1(n')$ . Then  $\lambda^{n-n'} = \gamma_1(n)/\gamma_1(n') \in \Gamma$ .

*Now suppose*  $\text{char } K = p > 0$ . Then Theorem 5.2 implies that  $\{(\lambda^n : \gamma_1(n) : \dots : \gamma_l(n)) : n \in \Omega'\}$  is contained in a union of finitely many sets of the type as in Eq. (2). Hence, there exists an infinite subset  $\Omega'' \subseteq \Omega'$  all of whose solutions are contained in a single such set. On this set, we can express the ratio of the first two coordinates in the form

$$\frac{\lambda^n}{\gamma_1(n)} = \underbrace{\beta_1^{q^{e_1(n)}-1} \beta_2^{q^{e_1(n)}(q^{e_2(n)}-1)} \dots \beta_k^{q^{e_1(n)+\dots+e_{k-1}(n)}(q^{e_k(n)}-1)}}_{\in \sqrt{\Gamma}} x^{q^{e_1(n)+\dots+e_k(n)}},$$

with suitable  $\beta_1, \dots, \beta_k \in \sqrt{\Gamma}$ , and  $q$  a power of  $p$ . Hence, we get  $\lambda^n/\gamma(n) = x^{q^{e(n)}}$  for some  $\gamma: \Omega'' \rightarrow \sqrt{\Gamma}$  and  $e(n) := e_1(n) + \dots + e_k(n)$ . Again take  $n \neq n' \in \Omega''$ , and without restriction let  $e(n) \leq e(n')$ . Then  $\lambda^{nq^{e(n')-e(n)}}/\gamma(n)^{q^{e(n')-e(n)}} = x^{q^{e(n')}} = \lambda^{n'}/\gamma(n')$ . Now  $p \nmid n'$  and  $n \neq n'$  imply  $nq^{e(n')-e(n)} \neq n'$ , and hence  $\lambda \in \sqrt{\Gamma}$ .  $\square$

We also note the following consequence for later use in Corollary 7.15.

**Lemma 5.4.** *Let  $\Lambda \leq K^\times$  be a group. If there exists  $M \geq 0$  such that*

$$\Lambda \subseteq M \cdot \Gamma_0,$$

*then  $\Lambda$  is finitely generated.*

*Proof.* Let  $K_0$  be the prime field of  $K$ . By the assumption, the group  $\Lambda$  is contained in the finitely generated field  $K_0(\Gamma)$ , and we may without restriction assume  $K = K_0(\Gamma)$ . Because  $\lambda^n \in M \cdot \Gamma_0$  for every  $\lambda \in \Lambda$  and  $n \in \mathbb{Z}$ , Lemma 5.3 implies  $\Lambda \subseteq \sqrt{\Gamma}$ . Since  $\sqrt{\Gamma}$  is finitely generated abelian, so is  $\Lambda$ .  $\square$

It is well known that  $\mathbb{Z}^d$  cannot be covered by finitely many cosets of subgroups of strictly smaller rank. This can be viewed as a consequence of the fact that  $\mathbb{Q}^d$  cannot be covered by finitely many proper affine subspaces, or as a special case of the fact that a group cannot be covered by finitely many cosets of subgroups of infinite index [Neu54, Lemma 4.1]. To be able to deal with positive characteristic, we use an easy variation of this result dealing with infinite unions, as long as the cosets are sufficiently sparse.

The upper natural density (or just density) of  $A \subseteq \mathbb{Z}$  is

$$d(A) := \limsup_{N \rightarrow \infty} \frac{|A \cap [-N, N]|}{2N+1}.$$

It holds that  $d(A \cup B) \leq d(A) + d(B)$  and  $d(A+y) = d(A)$  for all  $y \in \mathbb{Z}$ . Finite sets have zero density.

**Lemma 5.5.** *Let  $d \geq 0$  and let  $\varphi_1, \dots, \varphi_n: \mathbb{Z}^d \rightarrow \mathbb{Z}$  be nonzero group homomorphisms. If  $F_1, \dots, F_n \subseteq \mathbb{Z}$  are sets of zero density, then*

$$\bigcup_{i=1}^n \varphi_i^{-1}(F_i) \not\subseteq \mathbb{Z}^d.$$

*Proof.* We first show that we can reduce to the case  $\ker \varphi_j \not\subseteq \ker \varphi_k$  for  $j \neq k$ . Suppose  $\ker \varphi_j \subseteq \ker \varphi_k$  for  $j \neq k$ . Because  $\text{im } \varphi_j$  and  $\text{im } \varphi_k$  both have rank 1, the kernels both have rank  $d-1$ . Thus, the quotient  $\ker \varphi_k / \ker \varphi_j$  is finite. Then  $\varphi_j(\ker(\varphi_k))$  is a finite subgroup of  $\text{im } \varphi_j \cong \mathbb{Z}$ , and hence  $\varphi_j(\ker(\varphi_k)) = \{0\}$ . We conclude  $\ker \varphi_j = \ker \varphi_k$ .

Then

$$\varphi_j^{-1}(F_j) \cup \varphi_k^{-1}(F_k) = \varphi_k^{-1}(\tilde{F}_j \cup F_k) \quad \text{with} \quad \tilde{F}_j := \varphi_k(\varphi_j^{-1}(F_j)).$$

Let  $K := \ker \varphi_j = \ker \varphi_k$  and let  $\mathbf{x} \in \mathbb{Z}^d$  be such that  $\mathbf{x} + K$  generates  $\mathbb{Z}^d/K \cong \mathbb{Z}$ . Let  $n_j, n_k \in \mathbb{Z}$  be such that  $\varphi_j(\mathbf{x}) = n_j$  and  $\varphi_k(\mathbf{x}) = n_k$ . Then  $\tilde{F}_j \subseteq \frac{n_k}{n_j} F_j \cap \mathbb{Z}$ , so  $\tilde{F}_j$  has zero density. Upon replacing  $F_k$  by  $\tilde{F}_j \cup F_k$ , we may therefore drop the homomorphism  $\varphi_j$ . Repeating this argument, we ultimately get  $\ker \varphi_j \not\subseteq \ker \varphi_k$  for  $j \neq k$ .

We proceed by induction on  $d$  to show the claim of the lemma. The case  $d = 0$  is vacuously true because then necessarily  $n = 0$  and the union is empty while  $\mathbb{Z}^0 = \{0\}$ . Now let  $d \geq 1$  and suppose the claim holds for  $d-1$ .

The image  $\text{im } \varphi_1$  is a nonzero subgroup of  $\mathbb{Z}$ . Such subgroups have positive density, hence  $\text{im } \varphi_1 \not\subseteq F_1$ . Let  $\mathbf{x} \in \mathbb{Z}^d \setminus \varphi_1^{-1}(F_1)$  and set  $y_i := \varphi_i(\mathbf{x})$ . Then

$$\bigcup_{i=2}^n (\varphi_i^{-1}(F_i - y_i) \cap \ker(\varphi_1)) \subseteq \ker(\varphi_1).$$

Since  $\varphi_i|_{\ker(\varphi_1)} \neq 0$ , the induction hypothesis implies that the inclusion is proper. Let  $\mathbf{x}' \in \ker(\varphi_1)$  with  $\varphi_i(\mathbf{x}') \notin F_i - y_i$  for all  $i \in [2, n]$ . Then  $\mathbf{x} + \mathbf{x}' \notin \bigcup_{i=1}^n \varphi_i^{-1}(F_i)$ .  $\square$

We derive a somewhat crude corollary to Theorems 5.1 and 5.2 that holds in any characteristic and is sufficient for our purposes.

**Proposition 5.6.** *Let  $S \subseteq \Gamma^{n+1}$  be the set of all non-degenerate solutions to the unit equation Eq. (1). If  $\pi: \Gamma \rightarrow \mathbb{Z}$  is a group homomorphism, then*

$$F_{ij} = \{ \pi(x_i/x_j) : (x_0, \dots, x_n) \in S \}$$

*has zero density for all  $i, j \in [0, n]$ .*

*Proof.* In characteristic 0, each set  $F_{ij}$  is finite by Theorem 5.1. Suppose  $\text{char } K = p > 0$ . Replacing  $K$  be the field generated by  $\Gamma$  and  $a_0, \dots, a_n$  over  $\mathbb{F}_p$ , we may assume that  $K$  is a finitely generated field, and hence that  $\sqrt{\Gamma}$  is a finitely generated group. By Theorem 5.2, the set  $\mathbb{P}(S)$  is a finite union of sets of the form  $[\psi_1, \dots, \psi_k]_q \{\mathbf{y}\}$  as in Eq. (2). Fixing such a set and  $i, j \in [0, n]$ , it therefore suffices to show that

$$P_l := \{ \pi(x_i/x_j) : (x_0 : \dots : x_n) \in [\psi_{k-l+1}, \dots, \psi_k]_q \{\mathbf{y}\} \}$$

has zero density for all  $l \in [0, k]$ . To do so, we prove  $|P_l \cap [-N, N]| \in O(\log(N)^l)$  by induction on  $l$ .

First note that, for convenience, we can assume that  $\pi$  is defined on all of  $\sqrt{\Gamma}$ . Indeed, since  $\sqrt{\Gamma}$  is finitely generated abelian, the quotient  $\sqrt{\Gamma}/\Gamma$  has finite exponent  $e$ . Extending first  $\pi$  to  $\pi: \sqrt{\Gamma} \rightarrow \mathbb{Q}$  (using that  $\mathbb{Q}$  is injective as an abelian group), we have  $\pi(\sqrt{\Gamma}) \subseteq \frac{1}{e}\mathbb{Z}$ . Replacing  $\pi$  by  $e\pi$  (it is sufficient to prove the claim for this homomorphism) we have  $\pi: \sqrt{\Gamma} \rightarrow \mathbb{Z}$ .

The set  $P_0$  is a singleton, so it contains  $O(1)$  elements. Let  $l \geq 1$ , and suppose  $|P_{l-1} \cap [-N, N]| = O(\log(N)^{l-1})$ . Let  $\alpha_0, \dots, \alpha_n \in \sqrt{\Gamma}$  be such that  $\psi_{k-l+1}(x_0 : \dots : x_n) = (\alpha_0 x_0 : \dots : \alpha_n x_n)$ . For  $(x_0 : \dots : x_n) \in \mathbb{P}^n(\Gamma)$  we have

$$\pi\left(\frac{\alpha_i^{-1}(x_i \alpha_i)^{q^e}}{\alpha_j^{-1}(x_j \alpha_j)^{q^e}}\right) = q^e(y + c) - c \quad \text{with } c = \pi(\alpha_i/\alpha_j) \text{ and } y = \pi(x_i/x_j).$$

If  $q^e(y + c) - c \in [-N, N]$ , then  $|q^e(y + c)| \leq N + |c|$ . We get  $|y + c| \leq N + |c|$  and if  $y + c \neq 0$ , then also  $e \leq \log_q(N + |c|)$ . Then  $|y| \leq N' := N + 2|c|$ . Now there are at most  $O(\log(N')^{l-1})$  such choices for  $y \in P_{l-1}$  and  $O(\log(N'))$  choices for  $e$  (if  $y + c \neq 0$ , but if  $y + c = 0$  then  $e$  does not matter). Thus

$$|P_l \cap [-N, N]| \in O(\log(N')^l) = O(\log(N)^l). \quad \square$$

The argument we used in positive characteristic is similar to the one in [BS21, Lemma 4.4]. In either case, the fact that the non-degenerate solutions are parametrized by a finite union of orbits of the type in Eq. (2), shows that while there are infinitely many non-degenerate solutions, there are *few* of them, in a suitable quantitative sense.

## 6. LOCALLY BÉZIVIN IMPLIES WEAKLY EPIMONOMIAL

Throughout this section, let  $K$  be a field, embedded in some fixed algebraic closure  $\overline{K}$ , and let  $V$  be a finite-dimensional  $K$ -vector space. In this section we establish one of the main structural implications, namely that locally Bézivin groups are weakly epimonomial.

We start by verifying the basic properties of steady endomorphisms (Lemma 2.3).

*Proof of Lemma 2.3.* (1) If  $\lambda \in \overline{K}$  is an eigenvalue of  $A$ , then  $\lambda^n$  is an eigenvalue of  $A^n$ . If  $A^n$  is steady, then  $\lambda^n \in K$  by the definition of steadiness.

Suppose conversely that  $m \geq 1$  is such that  $\lambda^m \in K$  for all eigenvalues  $\lambda \in \overline{K}$ . Choosing a suitable basis of  $V$ , the endomorphism  $A^m$  can be put into Jordan normal form over  $K$ , and we identify it with its corresponding matrix. Now take  $n$  a multiple of  $m$  such that

- $(\lambda/\mu)^n = 1$  if  $\lambda, \mu \in \overline{K}$  are eigenvalues of  $A$  with  $\lambda/\mu$  a root of unity,
- $A^n$  has no nontrivial nilpotent block (so  $\ker(A^n) = \ker(A^{n+j})$  for all  $j \geq 0$ ), and
- if  $\text{char } K = p > 0$ , then  $A^n$  is diagonal.

The last property can be achieved because in a  $p^e$ -power of a Jordan block with eigenvalue  $\lambda$ , the entries in the  $j$ -th upper diagonal ( $j \geq 0$ ) are of the form  $\binom{p^e}{j} \lambda^{p^e-j}$ . Choosing  $e$  sufficiently large, the binomial coefficients are divisible by  $p$ , and the off-diagonal entries become 0.

With such a choice of  $n$ , it is not hard to see that  $A^n$  is steady (see for instance [BS23b, Lemmas 11 and 42]).

(2) Suppose  $\lambda \neq \mu$  and  $(\lambda/\mu)^n = 1$  for  $n \geq 1$ . If  $v, w \in V$  with  $Av = \lambda v$  and  $Aw = \mu w$ , then  $K(v+w)$  is  $A^n$ -invariant but not  $A$ -invariant.

(3) As in (1) we can assume that  $A$  is a matrix in Jordan normal form. Suppose  $A$  is not diagonal. Then  $A$  contains an upper-triangular Jordan block  $J_k(\lambda)$  of size  $k \geq 2$  and with  $\lambda \in K$ . Taking this block without restriction in the upper left corner, and denoting by  $e_1, \dots, e_d$  the standard basis, we have  $A^n e_2 = \lambda^n e_2 + n\lambda^{n-1} e_1$ . Thus, the space  $Ke_2$  is  $A^n$ -invariant but not  $A$ -invariant.  $\square$

**Lemma 6.1.** *Let  $S \subseteq \text{GL}(V)$  be a locally Bézivin semigroup.*

- (1) *If  $\text{char } K = 0$ , then every  $A \in S$  is diagonalizable over  $\overline{K}$ .*
- (2) *Every steady  $A \in S$  is diagonalizable over  $K$ .*

*Proof.* (1) Let  $A \in S$ . Fixing a suitable basis of  $\overline{K} \otimes_K V$  we can assume that  $A$  is an invertible matrix in Jordan normal form. Suppose that  $A$  has an upper-triangular Jordan block  $J_k(\lambda)$  with eigenvalue  $\lambda$  and size  $k \geq 2$ . Then the  $(1,2)$  entry of  $J^n$  is  $n\lambda^{n-1}$ . Because the subsemigroup generated by  $A$  is Bézivin, we have  $n\lambda^{n-1} \in M \cdot \Gamma_0$  for a finitely generated  $\Gamma \leq \overline{K}^\times$  and  $M \geq 0$ . Hence, for every  $n \geq 0$ , we have  $n \in M \cdot \Gamma'_0$  with  $\Gamma' = \Gamma \cup \{\lambda\}$ . However, this is impossible, say, by applying [BS23a, Lemma 4.7] with the linear polynomial  $y$ .

(2) By definition, all eigenvalues of  $A$  are contained in  $K$ . If  $\text{char } K = 0$ , then (1) implies that  $A$  is diagonalizable over  $K$ . If  $\text{char } K > 0$ , then  $A$  is diagonalizable by (3) of Lemma 2.3.  $\square$

*Remark 6.2.* Statement (1) of Lemma 6.1 also follows from a theorem of Bézivin [Béz86]. He showed that a univariate  $D$ -finite power series in  $\overline{K}[[x]]$  whose coefficients form a Bézivin set is, firstly, already rational, and secondly has only simple poles. The coefficients of a rational series are given by an LRS, and the  $n$ -th coefficient can be represented in the form  $uA^n v$  with  $u \in \overline{K}^{1 \times d}$ ,  $v \in \overline{K}^{d \times 1}$ , and  $A \in \overline{K}^{d \times d}$ . Given any non-diagonalizable  $A \in \text{GL}_d(\overline{K})$ , using the Jordan normal form, we can pick  $u, v$  such that the rational series with coefficients  $uA^n v$  does *not* have simple poles and hence apply [Béz86, Théorème 4] to obtain Lemma 6.1(1).

*Example 6.3.* Claim (1) of Lemma 6.1 is false in characteristic  $p > 0$ . Indeed,

$$J_2(\lambda)^n = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}^{\pm n} = \begin{pmatrix} \lambda^{\pm n} & \pm n \lambda^{\pm n-1} \\ 0 & \lambda^{\pm n} \end{pmatrix} \quad \text{for } n \in \mathbb{Z},$$

shows that the group generated by a Jordan block of size two is Bézivin, because  $n$  takes only finitely many values modulo  $p$ . Note however that the  $p$ -th power of the Jordan block is diagonal (and steady).

Suppose  $V = V_1 \oplus \cdots \oplus V_r$  with corresponding projections  $P_i: V \rightarrow V$  such that  $P_i|_{V_i}$  is the identity and  $P_i|_{V_j} = 0$  for  $j \neq i$ . If  $B \in \text{End}(V)$ , we denote by  $[B]_{ij} := P_i \circ B \circ P_j$  the  $ij$  “block” of  $B$ , so that  $B = \sum_{i,j=1}^r [B]_{ij}$ . If we fix bases of the  $V_i$  and identify  $B$  with its corresponding matrix, this indeed corresponds to a block of  $B$ .

We formulate the next proposition for a semigroup of endomorphisms because this gives us a slightly stronger statement, with implications for Corollary 11.1 later.

**Proposition 6.4.** *Let  $S \subseteq \text{End}(V)$  be a Bézivin semigroup. Let  $A \in S \cap \text{GL}(V)$  be steady and let  $V = V_1 \oplus \cdots \oplus V_r$  be the decomposition of  $V$  into  $A$ -eigenspaces. Suppose that there exist  $v = v_1 + \cdots + v_r \in V$  (with  $v_i \in V_i$ ), a linear map  $\psi: V \rightarrow K$ , a constant  $M \geq 0$  and a finitely generated  $\Gamma \leq K^\times$  such that*

$$\psi(Bv) \in M \cdot \Gamma_0 \quad \text{for all } B \in S.$$

*Then, for all  $l \geq 0$  and  $B_1, \dots, B_l \in S$ , it holds that*

$$|\{(i_0, \dots, i_l) \in [r]^{l+1} : \psi([B_1]_{i_0 i_1} [B_2]_{i_1 i_2} \cdots [B_l]_{i_{l-1} i_l} v_{i_l}) \neq 0\}| \leq M.$$

Thus, if we evaluate  $\psi(B_1 B_2 \cdots B_l v)$  as a block product corresponding to the decomposition  $V = V_1 \oplus \cdots \oplus V_r$ , the number of nonzero terms in the sum is bounded by  $M$ , no matter how large  $l$  is.

*Proof.* By Lemma 6.1, the automorphism  $A$  is diagonalizable over  $K$ , so that the required decomposition of  $V$  indeed exists. Let  $\lambda_i$  be the eigenvalue of  $A$  on  $V_i$ . Keep in mind that  $\lambda_1, \dots, \lambda_r$  are pairwise distinct.

We use multi-index notation  $\mathbf{n} = (n_0, \dots, n_l) \in \mathbb{Z}^{l+1}$ ,  $\mathbf{i} = (i_0, \dots, i_l) \in [r]^{l+1}$ ,  $\boldsymbol{\lambda}(\mathbf{i}) = (\lambda_{i_0}, \dots, \lambda_{i_l})$ , and set

$$\boldsymbol{\lambda}(\mathbf{i})^{\mathbf{n}} := \lambda_{i_0}^{n_0} \cdots \lambda_{i_l}^{n_l} \quad \text{and} \quad B(\mathbf{i}) := [B_1]_{i_0 i_1} [B_2]_{i_1 i_2} \cdots [B_l]_{i_{l-1} i_l}.$$

We have to show  $\{ \mathbf{i} \in [r]^{l+1} : \psi(B(\mathbf{i}) v_{i_l}) \neq 0 \} \leq M$ .

For all  $\mathbf{n} \in \mathbb{Z}^{l+1}$ ,

$$A^{n_0} B_1 A^{n_1} B_2 A^{n_2} B_3 \cdots A^{n_{l-1}} B_l A^{n_l} = \sum_{\mathbf{i} \in [r]^{l+1}} \boldsymbol{\lambda}(\mathbf{i})^{\mathbf{n}} B(\mathbf{i}).$$

By assumption, there exist functions  $\gamma_1, \dots, \gamma_M: \mathbb{Z}^{l+1} \rightarrow \Gamma_0$  such that

$$\sum_{\mathbf{i} \in [r]^{l+1}} \lambda(\mathbf{i})^{\mathbf{n}} \psi(B(\mathbf{i})v_{i_l}) = \gamma_1(\mathbf{n}) + \dots + \gamma_M(\mathbf{n}) \quad \text{for all } \mathbf{n} \in \mathbb{Z}^{l+1}. \quad (3)$$

We consider the equations (3) as solutions to a unit equation over a group  $\Gamma'$ , where  $\Gamma'$  is generated by  $\Gamma$  together with

- the elements  $\psi(B(\mathbf{i})v_{i_l})$  for  $\mathbf{i} \in [r]^{l+1}$ , and
- the elements  $\lambda_j$  for  $j \in [r]$ .

Note that these data remain fixed as  $\mathbf{n}$  varies, so that  $\Gamma'$  is indeed finitely generated and the same for all  $\mathbf{n}$ .

For all subsets  $\Omega \subseteq [r]^{l+1}$  and  $\Omega' \subseteq [M]$ , let  $N(\Omega, \Omega') \subseteq \mathbb{Z}^{l+1}$  be the set of all those  $\mathbf{n}$  for which

$$\sum_{\mathbf{i} \in \Omega} \lambda(\mathbf{i})^{\mathbf{n}} \psi(B(\mathbf{i})v_{i_l}) = \sum_{k \in \Omega'} \gamma_k(\mathbf{n}),$$

is a non-degenerate solution to a unit equation (that is, equality does not hold for a nonempty proper subsum). Because we can partition any solution to Eq. (3) into non-degenerate solutions, certainly  $\mathbb{Z}^{l+1} \subseteq \bigcup_{\Omega, \Omega'} N(\Omega, \Omega')$ .

We now consider one such set  $N(\Omega, \Omega')$  with  $|\Omega| \geq 2$ . Let  $\mathbf{i} \neq \mathbf{j} \in \Omega$ . By non-degeneracy, necessarily  $\psi(B(\mathbf{i})v_{i_l}), \psi(B(\mathbf{j})v_{j_l}) \neq 0$ . Consider the group homomorphism

$$\varphi_{\mathbf{i}, \mathbf{j}}: \mathbb{Z}^{l+1} \rightarrow \Gamma', \quad \mathbf{n} \mapsto \frac{\lambda(\mathbf{i})^{\mathbf{n}}}{\lambda(\mathbf{j})^{\mathbf{n}}} = \left( \frac{\lambda_{i_0}}{\lambda_{j_0}} \right)^{n_0} \dots \left( \frac{\lambda_{i_l}}{\lambda_{j_l}} \right)^{n_l}.$$

Because  $\mathbf{i} \neq \mathbf{j}$ , there exists some  $\nu \in [0, l]$  such that  $i_\nu \neq j_\nu$ . Then also  $\lambda_{i_\nu} \neq \lambda_{j_\nu}$ , since these eigenvalues are pairwise distinct. But  $A$  is steady, so  $\lambda_{i_\nu}/\lambda_{j_\nu}$  has infinite order in  $K^\times$  by (2) of Lemma 2.3. Thus, the image  $\text{im } \varphi_{\mathbf{i}, \mathbf{j}} \leq \Gamma'$  has a torsion-free component. Hence, there exists a homomorphism  $\pi_{\mathbf{i}, \mathbf{j}}: \Gamma' \rightarrow \mathbb{Z}$  that is nonzero on  $\text{im } \varphi_{\mathbf{i}, \mathbf{j}}$ . By Proposition 5.6, the set  $F_{\mathbf{i}, \mathbf{j}} := \pi_{\mathbf{i}, \mathbf{j}} \circ \varphi_{\mathbf{i}, \mathbf{j}}(N(\Omega, \Omega'))$  has zero density.

Now

$$N := \bigcup_{\substack{\Omega, \Omega' \\ |\Omega| \geq 2}} N(\Omega, \Omega') \subseteq \bigcup_{\mathbf{i}, \mathbf{j}} (\pi_{\mathbf{i}, \mathbf{j}} \circ \varphi_{\mathbf{i}, \mathbf{j}})^{-1}(F_{\mathbf{i}, \mathbf{j}}).$$

By Lemma 5.5 necessarily  $N \not\subseteq \mathbb{Z}^{l+1}$ .

Fix  $\mathbf{n} \in \mathbb{Z}^{l+1} \setminus N$ . Partitioning the solution Eq. (3) into non-degenerate subequations for this specific  $\mathbf{n}$ , we can, for every  $\mathbf{i} \in [r]^{l+1}$ , find a subset  $\Omega'(\mathbf{i}) \subseteq [M]$  such that  $\mathbf{n} \in N(\{\mathbf{i}\}, \Omega'(\mathbf{i}))$  and such that  $\Omega'(\mathbf{i}) \cap \Omega'(\mathbf{j}) = \emptyset$  for  $\mathbf{i} \neq \mathbf{j}$ . If  $\varphi(B(\mathbf{i})v_{i_l}) \neq 0$ , then necessarily  $\Omega'(\mathbf{i}) \neq \emptyset$ . The pairwise disjointness of these sets implies  $|\{\mathbf{i} \in [r]^{l+1} : \varphi(B(\mathbf{i})v_{i_l}) \neq 0\}| \leq M$ .  $\square$

Proposition 6.4 in this form will only be needed at the end, when we deal with automata. In dealing with linear groups, we use the following corollary.

**Corollary 6.5.** *Let  $S \subseteq \text{End}(V)$  be a Bézivin semigroup. Suppose that there exists a steady  $A \in S \cap \text{GL}(V)$ . Let  $V = V_1 \oplus \cdots \oplus V_r$  be the decomposition of  $V$  into  $A$ -eigenspaces. Then there exists a constant  $C \in \mathbb{Z}_{\geq 0}$  such that, for all  $l \geq 0$  and  $B_1, \dots, B_l \in S$ , it holds that*

$$|\{(i_0, \dots, i_l) \in [r]^{l+1} : [B_1]_{i_0 i_1} [B_2]_{i_1 i_2} \cdots [B_l]_{i_{l-1} i_l} \neq 0\}| \leq C.$$

In other words, in evaluating  $B_1 \cdots B_l$  as a block-matrix product with respect to that decomposition of  $V$ , there are at most  $C$  nonzero terms (independent of  $l$ ).

*Proof.* We use the same notation as in the proof of Proposition 6.4. Now we have to show  $|\{\mathbf{i} \in [r]^{l+1} : B(\mathbf{i}) \neq 0\}| \leq C$ . Let  $e_1, \dots, e_d$  be a basis of  $V$  with dual basis  $e_1^*, \dots, e_d^* : V \rightarrow K$ . It suffices to show that for all  $\mu, \nu \in [d]$ , there exists at most  $M \geq 0$  multi-indices  $\mathbf{i}$  for which  $e_\mu^* B(\mathbf{i}) e_\nu \neq 0$ . Then the required bound holds with  $C = Md^2$ . But for fixed  $\mu, \nu$  this follows from Proposition 6.4 by setting  $\psi = e_\mu^*$  and  $v = e_\nu$ .  $\square$

**Lemma 6.6.** *Let  $S \subseteq \text{GL}(V)$  be a locally Bézivin semigroup, let  $A \in S$  be steady, and let  $V = V_1 \oplus \cdots \oplus V_r$  be the decomposition of  $V$  into eigenspaces of  $A$ .*

- (1) *For every  $B \in S$ , there exist decompositions  $V_i = V_{i1} \oplus \cdots \oplus V_{is_i}$  such that  $B$  permutes the spaces  $\{V_{ij} : i \in [r], j \in [s_i]\}$ .*
- (2) *If  $B$  is steady, then  $BV_{ij} = V_{ij}$  for every  $i, j$  in a decomposition as in (1). In particular, it holds that  $BV_i = V_i$  for all  $i \in [r]$ .*

*Proof.* (1) For every sequence  $\mathbf{m} = (m_j)_{j \geq 0}$  with  $m_j \in [r]$ , let

$$W(\mathbf{m}) = \{v \in V : B^j v \in V_{m_j} \text{ for all } j \geq 0\}.$$

Observe that  $W(\mathbf{m})$  is a vector space and  $BW(m_0, m_1, m_2, \dots) \subseteq W(m_1, m_2, \dots)$ . We will show

$$V = \bigoplus_{\mathbf{m}} W(\mathbf{m}). \quad (4)$$

Because  $W(\mathbf{m}) \subseteq V_{m_0}$ , and thus in particular for each  $\mathbf{m}$  and  $i \in [r]$  either  $W(\mathbf{m}) \cap V_i = 0$  or  $W(\mathbf{m}) \subseteq V_i$ , it will then follow that

$$V_i = \bigoplus_{W(\mathbf{m}) \subseteq V_i} W(\mathbf{m}).$$

Renaming the finitely many nonzero spaces in Eq. (4) to  $V_{i1}, \dots, V_{is_i}$  will then yield claim (1).

First, we show that the sum in Eq. (4) is direct. Let

$$0 = \sum_{\mathbf{m}} w(\mathbf{m}) \quad (5)$$



with  $w(\mathbf{m}) \in W(\mathbf{m})$ , almost all of which are zero. Supposing, for the sake of contradiction, that some  $w(\mathbf{m})$  is not zero, we can also take the sum in Eq. (5) in such a way that the number of nonzero summands is minimal. Then there are at least two nonzero summands, say  $w(\mathbf{m}')$  and  $w(\mathbf{m}'')$ . We may suppose that  $\mathbf{m}' = (m'_k)_{k \geq 0}$  and  $\mathbf{m}'' = (m''_k)_{k \geq 0}$  differ in the  $j$ -th place. Applying  $B^j$  to Eq. (5),

$$0 = \sum_{\mathbf{m}} B^j w(\mathbf{m}) = \sum_{i=1}^r \underbrace{\sum_{\substack{\mathbf{m}=(m_k)_{k \geq 0} \\ m_j=i}} B^j w(\mathbf{m})}_{\in V_i}.$$

Because the sum  $V_1 \oplus \cdots \oplus V_r$  is direct and  $B$  is invertible, we obtain

$$0 = \sum_{\substack{\mathbf{m}=(m_k)_{k \geq 0} \\ m_j=m'_j}} w(\mathbf{m}) = \sum_{\substack{\mathbf{m}=(m_k)_{k \geq 0} \\ m_j=m''_j}} w(\mathbf{m}).$$

Now  $w(\mathbf{m}')$  appears in the first sum, but not the second, and conversely for  $w(\mathbf{m}'')$ . Thus, both sums are nonempty, but shorter than the one in Eq. (5), contradicting the minimality of its choice.

We still need to show  $V = \sum_{\mathbf{m}} W(\mathbf{m})$ . As in Corollary 6.5, we write  $[A]_{ij}$  and  $[B]_{ij}$  for blocks of  $A$ , respectively, of  $B$  corresponding to the decomposition  $V_1 \oplus \cdots \oplus V_r$ . For  $l \geq 1$  and  $\mathbf{i} = (i_0, \dots, i_l) \in [r]^{l+1}$ , define

$$B(\mathbf{i}) := B_{i_l i_{l-1}} \cdots B_{i_2 i_1} B_{i_1 i_0}.$$

and let

$$\Omega(l) := \{ \mathbf{i} \in [r]^{l+1} : B(\mathbf{i}) \neq 0 \}.$$

By Corollary 6.5, applied to the Bézivin subsemigroup  $\langle A, B \rangle$  of  $S$ , there exists  $C \geq 0$  such that  $|\Omega(l)| \leq C$  for all  $l$ . If  $\mathbf{i} = (i_0, \dots, i_l) \in \Omega(l)$ , then  $B \cdot B(\mathbf{i}) \neq 0$ , since  $B$  is invertible. Hence, there must exist  $i_{l+1}$  with  $B(i_0, i_1, \dots, i_{l+1}) \neq 0$  and hence  $\mathbf{i}' = (i_1, \dots, i_{l+1}) \in \Omega(l+1)$ .

Thus, the cardinality  $|\Omega(l)|$  is non-decreasing in  $l$  and must stabilize, say at some  $l = l_0$ . But then every  $\mathbf{i} \in \Omega(l)$  with  $l \geq l_0$  can be extended *uniquely* to some  $\mathbf{i}' \in \Omega(l+1)$ . This uniqueness implies  $B \cdot B(\mathbf{i})V \subseteq V_{i_{l+1}}$ . Doing this for all  $l \geq l_0$ , we find a sequence  $\mathbf{m} = (m_k)_{k \geq 0}$  such that  $B^k B(\mathbf{i})V \subseteq V_{m_k}$  for all  $k \geq 0$ . Hence, for every  $\mathbf{i} \in \Omega(l_0)$ , there exists a sequence  $\mathbf{m}$  such that  $B(\mathbf{i})V \subseteq W(\mathbf{m})$ .

Surjectivity of  $B$  implies  $\sum_{\mathbf{i} \in \Omega(l_0)} B(\mathbf{i})V = V$ , and thus

$$V = \sum_{\mathbf{i} \in \Omega(l_0)} B(\mathbf{i})V \subseteq \sum_{\mathbf{m}} W(\mathbf{m}).$$

(2) Suppose that we have a decomposition as in (1). Then there exists some  $n \geq 1$  such that  $B^n V_{ij} = V_{ij}$  for all  $i, j$ . Since  $B$  is steady, therefore  $BV_{ij} = V_{ij}$ .  $\square$

**Lemma 6.7.** *If  $S \subseteq \mathrm{GL}(V)$  is a locally Bézivin semigroup, then  $K$  is a power-splitting field for  $S$ .*

*Proof.* Fix  $A \in S$  and consider  $\bar{V} = \bar{K} \otimes_K V$ . Without restriction, we can consider the finitely generated subgroup  $S$  generated by  $A$ , so that  $S$  is Bézivin. We identify  $\mathrm{GL}(\bar{V}) \cong \mathrm{GL}_d(\bar{K})$  using a basis in which  $A$  has Jordan normal form. Then there exist a finitely generated subgroup  $\Gamma \leq K^\times$  and coefficients  $a_1, \dots, a_m \in \bar{K}^\times$  such that all entries of elements of  $S$  are contained in  $\sum_{i=1}^M a_i \Gamma_0$ . We stress that  $\Gamma$  can be taken over the base field  $K$  by our assumption, while the coefficients can only be taken in  $\bar{K}$ . If  $\lambda$  is an eigenvalue of  $A$ , then  $\lambda^n \in \sum_{i=1}^M a_i \Gamma_0$  for all  $n \in \mathbb{Z}$ . Lemma 5.3 shows  $\lambda^m \in \Gamma_0 \subseteq K$  for some  $m \geq 1$ .  $\square$

Given a set  $\mathcal{S} \subseteq \mathrm{End}(V)$ , we say that  $W \subseteq V$  is a joint-eigenspace of  $\mathcal{S}$  if for every  $A \in \mathcal{S}$  there exists an eigenspace  $W_A$  such that

$$W = \bigcap_{A \in \mathcal{S}} W_A.$$

**Lemma 6.8.** *Let  $S \subseteq \mathrm{GL}(V)$  be a locally Bézivin semigroup. Let  $V_1, \dots, V_r$  be the joint-eigenspaces of all steady elements of  $S$ . Then  $V_1 \oplus \dots \oplus V_r$  is a weakly epimonomial decomposition for  $S$ .*

The weakly epimonomial decomposition in the lemma is the coarsest possible. Since each component of any weakly epimonomial decomposition is contained in an eigenspace of all steady elements, every weakly epimonomial decomposition of  $S$  must be a refinement of the one obtained here.

*Proof.* We first show by induction on  $n$ : if  $A_1, \dots, A_n \in S$  are steady, then there exists a decomposition  $V = V_1 \oplus \dots \oplus V_r$  such that every  $V_i$  is a joint-eigenspace of  $A_1, \dots, A_n$ .

For  $n = 0$ , the claim holds vacuously. Let  $n \geq 1$  and suppose the claim holds for  $n - 1$ . Let  $i \in [r]$ . We know  $V_i = W_{i,1} \cap \dots \cap W_{i,n-1}$  with each  $W_{i,j}$  an eigenspace of  $A_j$  for  $j \in [n - 1]$ . By (2) of Lemma 6.6, each  $W_{i,j}$  is invariant under  $A_n$ . Hence, each  $V_i$  is invariant under  $A_n$ . Since  $A_n$  is diagonalizable over  $K$  by Lemma 6.1, the restriction  $A_n|_{V_i}$  is also diagonalizable. We can therefore decompose  $V_i$  into eigenspaces of  $A_n|_{V_i}$ , and since each such space is of the form  $U \cap V_i$  with  $U$  an eigenspace of  $A_n$ , the claim of the induction follows.

We can now decompose  $V = V_1 \oplus \dots \oplus V_r$  in such a way that each  $V_i$  is a joint-eigenspace of *all* steady  $A \in S$ . Indeed, decomposing iteratively, at each step each  $V_i$  is either already contained in an eigenspace of each steady  $A \in S$ , and we are done, or there exists some steady  $A \in S$  for which this is not the case. In the second case, we use  $A$  to decompose  $V_i$  further. Since  $V$  is finite-dimensional, this process stops after finitely many refinements.

Let  $B \in S$ . We now show that  $B$  permutes the spaces  $V_1, \dots, V_r$ . By (1) of Lemma 6.6, we can refine  $V_i = V_{i1} \oplus \dots \oplus V_{is_i}$  so that  $B$  permutes the spaces  $\{V_{ij} : i \in [r], j \in [s_i]\}$ . Let  $f, g$  be the functions for which  $BV_{ij} = V_{f(i,j),g(i,j)}$ . We will be done if we can show that  $f(i, j)$  does not actually depend on  $j$ , because then  $BV_i \subseteq V_{f(i,*)}$ .

Suppose there exist  $i' \in [r]$  and  $j' \neq j'' \in [s_{i'}]$  such that  $f(i', j') \neq f(i', j'')$ . By construction of the  $V_i$ , there exists a steady  $A \in S$  such that  $A$  has distinct eigenvalues on  $V_{f(i', j')}$  and  $V_{f(i', j'')}$ . By Lemma 6.7 the field  $K$  is power-splitting for  $S$ . Hence, there exists  $n \geq 1$  such that  $B^n$  is steady (using (1) of Lemma 2.3). Then each  $V_i$  is also a subspace of an eigenspace of  $B^n$ . Let  $\alpha_i$  be the eigenvalue of  $A$  on  $V_i$ , and  $\beta_i$  the eigenvalue of  $B^n$ .

Since  $B$  permutes the spaces  $V_{ij}$ , we have

$$B^{n-1}AB|_{V_{ij}} = \alpha_{f(i,j)}B^n|_{V_{ij}} = \alpha_{f(i,j)}\beta_i \text{id}_{V_{ij}} \quad \text{for } i \in [r] \text{ and } j \in s_i.$$

Let  $C := (B^{n-1}AB)^m$  with  $m \geq 1$  such that  $C$  is steady (again using Lemmas 2.3 and 6.7). Then  $C$  has eigenvalue  $\alpha_{f(i', j')}^m \beta_{i'}^m$  on  $V_{i'j'}$  and eigenvalue  $\alpha_{f(i', j'')}^m \beta_{i'}^m$  on  $V_{i'j''}$ . Because  $A$  is steady, these two eigenvalues are distinct (using (2) of Lemma 2.3). This contradicts the choice of  $V = V_1 \oplus \dots \oplus V_r$ . (In other words, the automorphism  $C$  would allow us to further refine the decomposition, contradicting the maximally refined choice).  $\square$

## 7. CHARACTERIZATION OF LOCALLY BÉZIVIN GROUPS

Let  $K$  be a field and  $V$  a finite-dimensional vector space. We are now in a position to establish the characterization of locally Bézivin groups. Let us first show that a weakly epimonomial decomposition of a semigroup of invertible endomorphisms lifts to the group.

**Lemma 7.1.** *Let  $S \subseteq \text{GL}(V)$  be a semigroup and let  $G = \langle S \rangle \leq \text{GL}(V)$ .*

- (1) *If  $S$  is weakly epimonomial, then  $G$  is weakly epimonomial.*
- (2) *If  $S$  is weakly epimonomial as a subsemigroup of  $\text{GL}(L \otimes_K V)$  for some extension field  $L$  and  $K$  is power-splitting for  $S$ , then  $K$  is also power-splitting for  $G$ .*

*Proof.* (1) We have to verify the properties in Definition 2.6 for the inclusion  $G \hookrightarrow \text{GL}(V)$ . Since property (2) holds for all  $A \in S$ , it also holds for all inverses and products, and hence holds for all  $A \in G$ .

Let  $D \subseteq S$  be the diagonal. Let  $\widehat{D} \leq \text{GL}(V)$  consists of all  $A \in \text{GL}(V)$  for which every  $V_i$  is contained in an eigenspace of  $A$ . Thus, we have  $D = S \cap \widehat{D}$ . After fixing bases of  $V_1, \dots, V_r$  to obtain a basis of  $V$ , the elements of  $\widehat{D}$  are the diagonal matrices that are scalar when restricted to each  $V_i$ . If  $P \in \text{GL}(V)$  permutes the spaces  $V_1, \dots, V_r$ , then  $P^{-1}\widehat{D}P \subseteq \widehat{D}$ .

To verify (1) and (3) of Definition 2.6, we first check:

- (i) every element of  $G$  is of the form  $AB^{-1}$  with  $A \in S$  and  $B \in \widehat{D}$ , where the eigenvalues of  $B$  are contained in the group generated by the spectrum of  $S$ .
- (ii) if  $A \in S$  and  $B \in \widehat{D}$ , then some power of  $AB^{-1}$  is in  $\widehat{D}$ .

To verify (i), note that if  $A \in S$ , then weak epimonomiality of  $S$  together with Lemma 2.3 implies  $A^n \in D$  for some  $n \geq 1$ . If  $B := A^n$ , then  $A(A^{n-1}B^{-1}) = \text{id}$ , so that  $A^{-1} = A^{n-1}B^{-1}$ . Thus, as a monoid, the group  $G$  is generated by  $S$  and  $D^{-1}$ . Explicitly, elements of  $G$  can be represented as  $A_1B_1^{-1}A_2B_2^{-1}\cdots A_kB_k^{-1}$  with  $A_i \in S \cup \{\text{id}\}$  and  $B_i \in D \cup \{\text{id}\}$ . We can express  $B_1^{-1}A_2 = A_2(A_2^{-1}B_1^{-1}A_2)$  with  $(A_2^{-1}B_1^{-1}A_2) \in \widehat{D}$ , because  $A_2$  permutes  $V_1, \dots, V_r$ . Iterating this, we find  $A_1B_1^{-1}A_2B_2^{-1}\cdots A_kB_k^{-1} = A_1\cdots A_kB^{-1}$  with  $B \in \widehat{D}$  (but possibly  $B \notin G$ ).

For the claim about the eigenvalues, note that the eigenvalues of  $P^{-1}B^{-1}P$ , with  $B \in S$  and  $P$  permuting the spaces  $V_1, \dots, V_r$ , are contained in the group  $\Gamma \leq K^\times$  generated by all the eigenvalues of  $S$ . Since the matrices of the form  $P^{-1}B^{-1}P$  are moreover diagonal, the same holds true for a product of such matrices.

To see (ii), note

$$(AB^{-1})^n = A^n \underbrace{(A^{-(n-1)}BA^{n-1})(A^{-(n-2)}BA^{n-2})\cdots(A^{-1}B^{-1}A)B^{-1}}_{\in \widehat{D}}. \quad (6)$$

Since  $S$  is weakly epimonomial, there exists some  $n$  such that  $A^n$  is steady (by Lemma 2.3), hence  $A^n \in D$ . We get  $(AB^{-1})^n \in \widehat{D}$ .

Now (i) and (ii) together imply that  $K$  is power-splitting for  $G$ , verifying (3) of weak epimonomiality. Let  $C \in G$  be steady. Since we can express  $C = AB^{-1}$  with  $A \in S$  and  $B \in \widehat{D}$ , each  $V_i$  is contained in an eigenspace of some power  $C^n$  of  $C$  (again using (i) and (ii)). Hence, every subspace of  $V_i$  is invariant under  $C^n$ , and by steadiness, also under  $C$ . It follows that  $V_i$  is contained in an eigenspace of  $C$ .

(2) Let  $C \in G$ . Let  $\widehat{V} = L \otimes_K V$  and let  $\widehat{V}_1 \oplus \cdots \oplus \widehat{V}_r$  be a weakly epimonomial decomposition for  $S \subseteq \text{GL}(\widehat{V})$ . Applying (i) from the proof of (1), we can again express  $C = AB^{-1}$  with  $A \in S$  and  $B \in \text{GL}(\widehat{V})$  such that each  $\widehat{V}_i$  is contained in an eigenspace of  $B$ . The eigenvalues of  $B$  are contained in the group  $\Gamma \leq L^\times$  generated by all the eigenvalues of  $S$ . By Lemma 2.3, there exists  $n \geq 1$  such that  $A^n$  is steady, and hence, in particular, diagonal with respect to any basis obtained by refining  $\widehat{V}_1 \oplus \cdots \oplus \widehat{V}_r$ . Eq. (6) shows that  $(AB^{-1})^n$  is also diagonal with all eigenvalues in  $\Gamma$ . By the power-splitting hypothesis for  $S$ , some power of  $(AB^{-1})^n$  has all its eigenvalues in  $K$ .  $\square$

**Proposition 7.2.** *If  $G \leq \text{GL}(V)$  has a weakly epimonomial decomposition with diagonal  $D$ , then  $D \trianglelefteq G$  and  $G/D$  is a linear torsion group.*

*Proof.* Let  $V = V_1 \oplus \cdots \oplus V_r$  be the given weakly epimonomial decomposition. By definition, the diagonal  $D$  contains all steady  $A \in G$ , and  $D$  is clearly a subgroup of  $G$ . To see that it is normal, let  $A \in D$  and  $B \in G$ . Let  $i \in [r]$  and let  $k \in [r]$  be the unique index for which  $BV_i = V_k$ . If  $\lambda$  is the eigenvalue of  $A$  on  $V_k$ , then  $B^{-1}AB|_{V_i} = B^{-1}\lambda B|_{V_i} = \lambda \text{id}|_{V_i}$ . Hence, we have  $B^{-1}AB \in D$ .

For every  $B \in G$  some power  $B^n$  is steady (using Lemma 2.3), and hence  $B^n \in D$ . Thus, the quotient  $G/D$  is a torsion group. We show that it is a linear group, that is, that it embeds into some  $\text{GL}(W)$  with  $W$  a finite-dimensional  $K$ -vector space. Note that  $G$  acts on  $\text{End}(V)$  by conjugation, that is, there is a group homomorphism  $\Phi: G \rightarrow \text{GL}(\text{End}(V))$ ,  $B \mapsto \Phi_B$  with  $\Phi_B(X) = BXB^{-1}$  for all  $X \in \text{End}(V)$ . Since  $B$  permutes  $V_1, \dots, V_r$ , this restricts to a representation  $\varphi$  of  $G$  on  $W := \text{End}(V_1) \oplus \cdots \oplus \text{End}(V_r)$  (or, after fixing a basis, on block-diagonal matrices). Explicitly  $\varphi: G \rightarrow \text{GL}(W)$ ,  $B \mapsto \varphi_B$  is given by

$$\varphi_B(X_1, \dots, X_r) = (BX_{\pi(1)}B^{-1}|_{V_1}, \dots, BX_{\pi(r)}B^{-1}|_{V_r}),$$

with  $\pi$  the permutation for which  $B^{-1}V_i = V_{\pi(i)}$ , and  $X_i \in \text{End}(V_i)$ . Because the center of  $\text{End}(V_i)$  consists of scalar multiples of the identity, we see that  $B \in \ker \varphi$  if and only if  $\pi = \text{id}$  and  $B|_{V_i}$  is a scalar multiple of the identity for each  $i \in [r]$ . Hence, we find  $\ker(\varphi) = D$ , and  $G/D \cong \text{im } \varphi$  is a linear group.  $\square$

*Remark 7.3.* Let  $\tilde{G}$  denote the Zariski closure of  $G$ . Then  $\tilde{D}(K)$  consists of elements for which each  $V_i$  is contained in an eigenspace, and hence  $D = \tilde{D}(K) \cap G$  by the definition of the diagonal (here  $\tilde{D}(K)$  denotes the set of  $K$ -rational points of the scheme  $\tilde{D}$ ). Since  $\tilde{D} \trianglelefteq \tilde{G}$  is a closed normal subgroup, Chevalley's theorem on linear algebraic groups implies that  $\tilde{G}/\tilde{D}$  is a linear algebraic group as well, and therefore  $G/D$  is a linear group. The previous proof makes the arguments in Chevalley's theorem explicit in this special case without reference to linear algebraic groups. Since the argument is short, and also has the benefit of making explicit the dimension of  $W$  (which is useful for estimates as in [BS23b, Appendix C]), we have chosen to give this explicit variant.

We obtain a characterization of locally Bézivin (semi)groups that holds over any field. We need the following classical theorems of Burnside and Schur.

**Theorem 7.4** (Burnside–Schur). *Let  $G \leq \text{GL}(V)$ .*

- (1) (Schur) *If  $G$  is a torsion group, then  $G$  is locally finite.*
- (2) (Burnside) *If  $\text{char } K = 0$  and  $G$  has finite exponent, then  $G$  is finite.*

**Theorem 7.5.** *Let  $S \subseteq \text{GL}(V)$  be a semigroup and let  $G = \langle S \rangle$ . Then the following statements are equivalent.*

- (a) *The group  $G$  is locally Bézivin.*

- (b) *The semigroup  $S$  is locally Bézivin.*
- (c) *The semigroup  $S$  is weakly epimonomial.*
- (d) *The group  $G$  is weakly epimonomial.*
- (e) *The group  $G$  is locally epimonomial.*
- (f) *The group  $G$  is locally virtually simultaneously diagonalizable (over  $K$ ).*

The last statement means that every finitely generated subgroup  $H \leq G$  has a finite-index subgroup  $D \leq H$  that is simultaneously diagonalizable.

*Proof.* (a)  $\Rightarrow$  (b): Trivial.

(b)  $\Rightarrow$  (c): By Lemma 6.8.

(c)  $\Rightarrow$  (d): By Lemma 7.1.

(d)  $\Rightarrow$  (e): Let  $H \leq G$  be a finitely generated subgroup. Then  $H$  is also weakly epimonomial. Let  $D \triangleleft H$  be the diagonal with respect to some weakly epimonomial decomposition. By Proposition 7.2 the quotient  $H/D$  is a finitely generated linear torsion group. Burnside–Schur (Theorem 7.4) implies that  $H/D$  is finite.

(e)  $\Rightarrow$  (f): Let  $H \leq G$  be finitely generated. Then  $H$  is epimonomial with a diagonal  $D \triangleleft H$  that has finite index, and  $D$  is simultaneously diagonalizable (over  $K$ ).

(f)  $\Rightarrow$  (a): Let  $H \leq G$  be finitely generated and let  $D \leq H$  be a finite-index subgroup of simultaneously diagonalizable elements. Let  $B_1, \dots, B_n$  be a system of representatives for the left cosets of  $H/D$ . It suffices to show that each element of  $B_i D$  has entries in  $M \cdot \Gamma_0$  for some  $M \geq 0$  and a finitely generated  $\Gamma \leq K^\times$ . Now  $D$ , as a finite-index subgroup in the finitely generated group  $H$ , is also finitely generated [Rob96, 1.6.11]. Since the elements of  $D$  are moreover simultaneously diagonalizable, the claim follows.  $\square$

In particular, we see that the local Bézivin property for a semigroup  $S \subseteq \mathrm{GL}(V)$  is equivalent to the local Bézivin property for the group  $G = \langle S \rangle$ . Hence, we may usually deal with groups for the remainder of the section.

**Corollary 7.6.** *A finitely generated  $G \leq \mathrm{GL}(V)$  is Bézivin if and only if it is epimonomial.*

*Proof.* By Theorem 7.5.  $\square$

By Theorem 7.5 and Lemma 6.8 every weakly epimonomial  $G \leq \mathrm{GL}(V)$  gives the coarsest (hence canonical) decomposition  $V = V_1 \oplus \dots \oplus V_r$  with  $V_i$  an intersection of eigenspaces of all steady elements of  $G$ . However, even if  $G$  is epimonomial, it may not be epimonomial with respect to this particular decomposition, as the next example shows.

*Example 7.7.* Let

$$G := \begin{pmatrix} \mu(\mathbb{C}) & 0 \\ 0 & 1 \end{pmatrix} \leq \mathrm{GL}_2(\mathbb{C}).$$

Then  $G$  is an infinite torsion group and the identity is the only steady element. The weakly epimonomial decomposition of  $\mathbb{C}^2$  arising from the steady elements is therefore just  $\mathbb{C}^2$  itself, and  $G$  is not epimonomial with respect to the corresponding diagonal. However, clearly  $\mathbb{C}^2 = \mathbb{C} \oplus \mathbb{C}$  is an epimonomial decomposition.

We now show that, in Theorem 7.5, we can also relax the locally Bézivin property to hold over an extension field, as long as  $K$  is power-splitting for  $S$ .

**Proposition 7.8.** *For a finitely generated semigroup  $S \subseteq \mathrm{GL}(V)$  the following statements are equivalent.*

- (a) *The semigroup  $S$  is Bézivin.*
- (b) *The semigroup  $S \subseteq \mathrm{GL}(L \otimes_K V)$  is Bézivin for every extension field  $L/K$  and  $K$  is a power-splitting field for  $S$ .*
- (c) *The semigroup  $S \subseteq \mathrm{GL}(L \otimes_K V)$  is Bézivin for some extension field  $L/K$  and  $K$  is a power-splitting field for  $S$ .*

*Proof.* (a)  $\Rightarrow$  (b): Clearly, the semigroup  $S$  is still Bézivin over  $L$ . By Lemma 6.7 the field  $K$  is power-splitting for  $S$ .

(b)  $\Rightarrow$  (c): Trivial.

(c)  $\Rightarrow$  (a): Let  $G = \langle S \rangle$ . We identify  $\mathrm{GL}(V) = \mathrm{GL}_d(K)$  by choosing a basis of  $V$ . Extending scalars (that is, viewing  $G$  as a subgroup of  $\mathrm{GL}_d(L)$ ), we obtain an  $L$ -representation of  $G$ . By Theorem 7.5 the group  $G$  is weakly epimonomial as a subgroup of  $\mathrm{GL}_d(L)$ . Then (2) of Lemma 7.1 shows that  $K$  is power-splitting for  $G$ .

By (f) of Theorem 7.5, there exists a finite-index subgroup  $T \leq G$  that is simultaneously diagonalizable over  $L$ . Let  $B \in \mathrm{GL}_d(L)$  be such that all elements of  $BTB^{-1}$  are diagonal. As a finite-index subgroup of a finitely generated group, the group  $T$  is also finitely generated. Since  $K$  is power-splitting for  $G$ , for each  $A \in BTB^{-1}$  there exists some  $n(A)$  such that  $A^{n(A)}$  has all entries in  $K$ . Let  $T' \leq T$  be the subgroup generated by all these  $A^{n(A)}$ . Since  $T$  is finitely generated abelian, so is  $T'$ . Since  $T/T'$  is torsion, it is finite.

Let  $\Gamma \leq K^\times$  be the (finitely generated) group generated by all entries of the matrices in  $BT'B^{-1}$ . Transforming back into the original basis and keeping in mind that  $T'$  has finite index in  $G$ , there exist  $c_1 = 1, c_2, \dots, c_m \in L$  such that every entry of every element of  $G$  is contained in  $\sum_{i=1}^m c_i \Gamma_0$ . Among these elements  $c_1, \dots, c_m$ , we pick a maximal  $K$ -linearly independent set; after reindexing, let this be  $c_1 = 1, c_2, \dots, c_n$  with  $n \leq m$ .



Express each  $c_i$  (with  $i \in [m]$ ) as  $c_i = \sum_{j=1}^n \gamma_{i,j} c_j$  with  $\gamma_{i,j} \in K$ . Then the entries of elements of  $G$  are contained in

$$K \cap \sum_{i=1}^m c_i \Gamma_0 = K \cap \sum_{j=1}^n \left( \sum_{i=1}^m \gamma_{i,j} \right) c_j \Gamma_0 \subseteq \sum_{i=1}^m \gamma_{i,1} \Gamma_0,$$

where the last inclusion follows from  $K$ -linear independence of  $c_1, \dots, c_n$ . Letting  $\Gamma' \leq K^\times$  be the subgroup generated by  $\Gamma$  and all  $\gamma_{i,1}$ , all entries of all elements of  $G$  are contained in  $m \cdot \Gamma'_0$ , and thus  $G$  is Bézivin. Hence, so is  $S$ .  $\square$

The following is not essential, but allows us to add the Pólya property to the characterization in Theorem 1.1, for completeness.

**Lemma 7.9.** *If  $G \leq \mathrm{GL}(V)$  is epimonomial, then there exists a basis of  $V$  with respect to which  $G$  is locally Pólya.*

*Proof.* Let  $V = V_1 \oplus \dots \oplus V_r$  be an epimonomial decomposition of  $V$ , and let  $D \trianglelefteq G$  be the diagonal of this decomposition. Choose a basis of  $V$  by choosing bases on each  $V_i$ , and identify  $G = \mathrm{GL}_d(K)$  using this basis. Then the elements of  $G$  have a monomial block-structure with respect to the decomposition  $V = V_1 \oplus \dots \oplus V_r$ . The elements of  $D$  are diagonal, and each diagonal block is a scalar multiple of the corresponding identity matrix. Let  $A_1, \dots, A_n$  be a set of representatives for  $G/D$ . If  $C \in D$ , then  $A_i C$  is obtained from  $A_i$  by multiplying each block by a scalar.

Let  $H \leq G$  be finitely generated and assume without restriction  $A_1, \dots, A_n \in H$ . Then  $H \cap D$  is finitely generated, as a finite-index subgroup of  $H$ . Let  $\Gamma \leq K^\times$  be generated by the nonzero entries of  $A_1, \dots, A_n$  together with the entries of a set of generators of  $H \cap D$ . Then  $H \subseteq \Gamma_0^{d \times d}$ .  $\square$

**7.1. Maximal separating elements.** For the coarsest weakly epimonomial decomposition of a weakly epimonomial group, it is sometimes useful to know that there is a single element that creates this decomposition.

**Lemma 7.10.** *Let  $G \leq \mathrm{GL}(V)$  be weakly epimonomial and  $V = V_1 \oplus \dots \oplus V_r$  with  $V_i$  the joint-eigenspaces of all steady elements of  $G$ . Then there exists a steady  $A \in G$  such that the eigenvalues  $\lambda_1, \dots, \lambda_r$  of  $A$  on  $V_1, \dots, V_r$  are pairwise distinct.*

*Proof.* The spaces  $V_i$  are intersections of eigenspaces of steady  $A \in G$ . Thus, there exist  $N \geq 0$  and steady  $A_1, \dots, A_N$  such that, for every  $i \neq i'$  there exists some  $A_j$  having distinct eigenvalues on  $V_i$  and  $V_{i'}$ .

Let  $\lambda_{ij}$  be the eigenvalue of  $A_j$  on  $V_i$ . For all  $\mathbf{n} = (n_1, \dots, n_N) \in \mathbb{Z}^N$ , we consider the product

$$B(\mathbf{n}) := \prod_{j \in [N]} A_j^{n_j},$$

which has on  $V_i$  the eigenvalue

$$\mu_i(\mathbf{n}) := \prod_{j \in [N]} \lambda_{ij}^{n_j}.$$

We seek  $\mathbf{n}$  such that  $\mu_i(\mathbf{n})/\mu_{i'}(\mathbf{n})$  has infinite order for all  $i \neq i'$ .

Note that  $\varphi_{ii'}: \mathbb{Z}^N \mapsto \overline{K}^\times$ ,  $\mathbf{n} \mapsto \mu_i(\mathbf{n})/\mu_{i'}(\mathbf{n})$  is a group homomorphism. For every pair  $i \neq i'$  there exists some  $j$  with  $\lambda_{ij} \neq \lambda_{i'j}$ . Since  $A_j$  is steady, the ratio  $\lambda_{ij}/\lambda_{i'j}$  even has infinite order. Thus, the image  $\text{im } \varphi_{ii'}$  has a torsion-free part. Composing  $\varphi_{ii'}$  with  $\pi: \overline{K}^\times \rightarrow \overline{K}^\times/\mu(\overline{K})$ , we see that  $\ker(\pi \circ \varphi_{ii'})$  has rank strictly smaller than  $N$ .

Thus, we get  $\bigcup_{i \neq i'} \ker(\pi \circ \varphi_{ii'}) \subsetneq \mathbb{Z}^N$ . Taking any  $\mathbf{n}$  outside this union ensures that  $\mu_i(\mathbf{n})/\mu_{i'}(\mathbf{n})$  has infinite order for all  $i \neq i'$ . The element  $B(\mathbf{n})$  is steady and has pairwise distinct eigenvalues on  $V_1, \dots, V_n$ .  $\square$

**7.2. Lifting to monomial representations.** We now show that epimonomial representations are epimorphic images of monomial representations.

**Lemma 7.11.** *Let  $\rho: G \rightarrow \text{GL}(V)$  be a representation and let  $T \leq G$  be a finite-index subgroup such that  $\rho(T)$  is simultaneously diagonalizable over  $K$ . Then the induced representation  $\pi := \text{Ind}_T^G(\rho|_T): G \rightarrow \text{GL}(W)$  is monomial and there exists a  $G$ -equivariant epimorphism  $(W, \pi) \twoheadrightarrow (V, \rho)$ .*

*Proof.* We check that  $\pi$  is a monomial representation. Let  $V = V_1 \oplus \dots \oplus V_r$  with  $\dim(V_i) = 1$  and such that  $\rho(T)$  is diagonal with respect to this decomposition. Let  $a_1, \dots, a_n$  represent the left cosets of  $G/T$ . For each  $b \in G$  and  $i \in [n]$ , let  $c_{i,b} \in T$  and  $f(i, b) \in [n]$  be such that

$$ba_i = a_{f(i,b)} c_{i,b}.$$

Recall that the induced representation is defined on the  $nr$ -dimensional space

$$W = \bigoplus_{i=1}^n a_i V = \bigoplus_{i=1}^n \bigoplus_{j=1}^r a_i V_j.$$

For each  $b \in G$  the endomorphism  $\pi(b) \in \text{GL}(W)$  is defined by

$$\pi(b)(a_i v) = a_{f(i,b)} (\rho(c_{i,b})(v)) \quad \text{for } v \in V.$$

Since  $\rho(c_{i,b})$  is diagonal, there exists  $\lambda_{i,b} \in K$  for which  $\rho(c_{i,b})(v_j) = \lambda_{i,b} v_j$  for  $v_j \in V_j$ , so in fact  $\pi(b)(a_i v_j) = a_{f(i,b)} (\lambda_{i,b} v_j)$ . Hence, the representation  $\pi$  is monomial.

Define a linear epimorphism

$$\psi: W \twoheadrightarrow V, \quad (a_i v_i)_{i \in [n]} \mapsto \sum_{i=1}^n \rho(a_i) v_i \quad (\text{for } v_i \in V).$$

Then  $\psi$  is  $G$ -equivariant, because

$$\begin{aligned} \psi(\pi(b)(a_i v_i)_{i \in [n]}) &= \psi\left(\left(a_{f(i,b)}(\rho(c_{i,b})(v_i))\right)_{i \in [n]}\right) = \sum_{i=1}^n \rho(a_{f(i,b)} c_{i,b})(v_i) = \sum_{i=1}^n \rho(b a_i)(v_i) \\ &= \rho(b)\left(\sum_{i=1}^n \rho(a_i) v_i\right) = \rho(b)\left(\psi((a_i v_i)_{i \in [n]})\right). \end{aligned} \quad \square$$

**7.3. Not necessarily finitely generated groups.** Example 4.2 shows that not every locally Bézivin group has a monomial representation. In (1) of Example 4.2, the issue is that  $\mathbb{R}$  is power-splitting for  $\mu(\mathbb{C}) \hookrightarrow \mathrm{GL}_2(\mathbb{R})$  but not uniformly so. In characteristic 0, we can still get the following two results for not necessarily finitely generated groups, as long as we assume *uniform* power-splitting. First, we have the following easy observation.

**Corollary 7.12.** *Suppose  $\mathrm{char} K = 0$  and  $|\mu(K)| < \infty$ . For  $G \leq \mathrm{GL}(V)$  the following statements are equivalent.*

- (a) *The group  $G$  satisfies the equivalent conditions of Theorem 7.5 and  $K$  is uniformly power splitting for  $G$ .*
- (b) *The group  $G$  is epimonomial.*

*Proof.* (a)  $\Rightarrow$  (b): By assumption  $G$  is a weakly epimonomial. Let  $D \trianglelefteq G$  be a diagonal with respect to a weakly epimonomial decomposition of  $V$ . Then, using Proposition 7.2, the quotient  $G/D$  is a linear torsion group, and we show that it has finite exponent. By assumption, there exists  $N \geq 1$  such that every  $A^N$ , with  $A \in G$ , has its eigenvalues in  $K$ . Let  $M := |\mu(K)|N$ . If  $A \in G$ , then  $A^N$  is diagonalizable over  $K$  (Lemma 6.1). Raising  $A^N$  to the power  $|\mu(K)|$  ensures  $\lambda/\mu$  cannot be a nontrivial root of unity for two distinct eigenvalues  $\lambda, \mu$  of  $A$ . Therefore, the power  $A^M$  is steady, and hence  $A^M \in D$ . Because  $\mathrm{char} K = 0$ , every linear torsion group with finite exponent is finite (Theorem 7.4).

(b)  $\Rightarrow$  (a): Let  $D$  be a diagonal corresponding to an epimonomial decomposition of  $V$ . Since  $G/D$  is finite, there exists  $N \geq 1$  such that  $A^N \in D$  for all  $A \in G$ . Let  $A \in G$  and let  $\lambda \in \overline{K}$  be an eigenvalue of  $A$ . Then  $\lambda^N$  is an eigenvalue of  $A^N$ . But the eigenvalues of  $A^N$  are in  $K$  by definition of  $D$ , hence  $\lambda^N \in K$ .  $\square$

Achieving an analogous result when  $\mu(K)$  may be infinite is harder. We need a classical theorem of Jordan.

**Theorem 7.13** (Jordan). *Let  $K$  be a field of characteristic 0. Then there is a function  $f: \mathbb{Z}_{\geq 0} \rightarrow \mathbb{Z}_{\geq 0}$  such that every finite subgroup  $G \leq \mathrm{GL}_d(K)$  has an abelian subgroup of index at most  $f(d)$ .*

**Theorem 7.14.** *Suppose  $\mathrm{char} K = 0$  and let  $G \leq \mathrm{GL}(V)$ . For  $G \leq \mathrm{GL}(V)$  the following statements are equivalent.*

- (a) *The group  $G$  satisfies the equivalent conditions of Theorem 7.5 and  $K$  is uniformly power splitting for  $G$ .*
- (b) *The group  $G$  is virtually simultaneously diagonalizable over  $K$ .*

*Proof.* (a)  $\Rightarrow$  (b): By Theorem 7.5 there exists a weakly epimonomial decomposition  $V = V_1 \oplus \cdots \oplus V_r$  of  $V$ . Let  $d_i := \dim V_i$ . Let  $D$  be the diagonal of  $G$ . Then  $G/D$  is a linear torsion group that embeds into  $\mathrm{GL}(W)$  with  $W = \mathrm{End}(V_1) \oplus \cdots \oplus \mathrm{End}(V_r)$  (see Proposition 7.2). Let  $G' \leq G$  be the subgroup of all  $B \in G$  with  $BV_i = V_i$  for all  $i \in [r]$ . Then  $(G : G') \leq r!$  and it suffices to prove that  $G'$  is virtually simultaneously diagonalizable. Note that  $D \subseteq Z(G')$  by choice of  $G'$ .

Let  $\mathbb{P}_{\mathrm{fin}}(G')$  denote the set of all finite subsets of  $G$ . For every  $S \in \mathbb{P}_{\mathrm{fin}}(G')$  let  $G_S = \langle S \rangle$ . Then  $G' = \bigcup_{S \in \mathbb{P}_{\mathrm{fin}}(G')} G_S$ . The decomposition of  $V$  is weakly epimonomial for each  $G_S$  as well, and  $D_S := D \cap G_S$  is the corresponding diagonal. The map  $G \rightarrow \mathrm{GL}(W)$  with kernel  $D$  therefore gives rise to an embedding  $G_S/D_S \hookrightarrow \mathrm{GL}(W)$ , so that each  $G_S/D_S$  embeds into the fixed  $\mathrm{GL}(W)$ . As a finitely generated linear torsion group, each  $G_S/D_S$  is finite by Burnside–Schur (Theorem 7.4). Now Jordan’s Theorem (Theorem 7.13) implies that there exist a constant  $C$  and subgroups  $D_S \leq H_S \leq G_S$  such that  $H_S/D_S$  is abelian and  $(G_S : H_S) \leq C$  for all  $S \in \mathbb{P}_{\mathrm{fin}}(G')$ .

With a similar compactness argument as in a proof of Jordan–Schur [Tao14, Chapter 2.1] we can choose the groups  $H_S$  so that  $H_T = H_S \cap G_T$  whenever  $T \subseteq S$ : Let  $\mathfrak{X}_S$  denote the set of all subgroups  $H_S$  with  $D_S \leq H_S \leq G_S$  such that  $H_S/D_S$  is abelian and  $(G_S : H_S) \leq C$ . We endow  $\mathfrak{X}_S$  with the discrete topology. Since each  $\mathfrak{X}_S$  is finite, hence compact, the product space  $\mathfrak{X} = \prod_{S \in \mathbb{P}_{\mathrm{fin}}(G')} \mathfrak{X}_S$  is compact by Tychonoff’s Theorem. For each  $S \in \mathbb{P}_{\mathrm{fin}}(G')$  the set

$$\mathfrak{C}(S) := \left\{ (H_T)_{T \in \mathbb{P}_{\mathrm{fin}}(G')} \in \mathfrak{X} : H_T = G_T \cap H_S \text{ for } T \subseteq S \right\}$$

is closed (because it is defined by only finitely many conditions). Each  $\mathfrak{C}(S)$  is nonempty, because we may take  $H_S \in G_S$  with  $H_S/D_S$  abelian and  $(G_S : H_S) \leq C$  and then  $H_T := H_S \cap G_T$  for  $T \subseteq S$  will also satisfy  $(G_T : H_T) \leq C$ . Since  $\mathfrak{C}(S_1) \cap \cdots \cap \mathfrak{C}(S_n) \supseteq \mathfrak{C}(S)$  for  $S = S_1 \cup \cdots \cup S_n$ , any finite intersection of such sets is also nonempty. The finite intersection property for compact sets implies  $\bigcap_{S \in \mathbb{P}_{\mathrm{fin}}(G')} \mathfrak{C}(S) \neq \emptyset$ . Taking any element  $(H_S)_{S \in \mathbb{P}_{\mathrm{fin}}(G')}$  in this intersection, we have  $H_T = H_S \cap G_T$  for all  $S, T \in \mathbb{P}_{\mathrm{fin}}(G')$  with  $T \subseteq S$ .

Let  $H = \bigcup_{S \in \mathbb{P}_{\mathrm{fin}}(G')} H_S$ . Then  $(G' : H) \leq C$ : suppose  $A_1, \dots, A_n \in G'$  are in distinct cosets modulo  $H$ . With  $S = \{A_1, \dots, A_n\}$  we have  $A_1, \dots, A_n \in G_S$ . Then  $A_1, \dots, A_n$  must represent distinct cosets of  $H_S$  in  $G_S$ , hence  $n \leq C$ . Furthermore, by construction  $H/D$  is abelian with  $D \subseteq Z(G')$ . In particular, the group  $H$  is nilpotent of nilpotency index at most 2. By definition of  $D$ , each element of  $D$ , when restricted to  $V_i$ , consists

of a scalar multiple of the identity on  $V_i$ . Let  $[A, B] := ABA^{-1}B^{-1}$  be the multiplicative commutator. For all  $A, B \in H$  it holds that  $\det([A, B]) = 1$  and  $[A, B] \in D$ . Thus, we have  $[A, B]|_{V_i} = \zeta_i \text{id}$  for some  $\zeta_i \in \mu_{d_i}(K)$ . By nilpotency class at most 2, we have the identity  $[XY, Z] = [X, Z][Y, Z]$  for all  $X, Y, Z \in H$ . Therefore  $[A^{d_i}, B]|_{V_i} = ([A, B]|_{V_i})^{d_i} = \zeta_i^{d_i} \text{id}|_{V_i} = \text{id}|_{V_i}$ , so  $[A^{d_i}, B] = \text{id}$ .

By Lemma 6.1, all elements of  $H$  are diagonalizable over  $\overline{K}$ . Since  $K$  is uniformly power-splitting for  $G$ , there exists an  $N \geq 1$  such that  $A^N$  is diagonalizable over  $K$  for each  $A \in H$ . We may choose  $N$  so that also  $\text{lcm}(d_1, \dots, d_r) \mid N$ . Then  $[A^N, B] = [A, B]^N = \text{id}$  shows that  $B$  and  $A^N$  commute for every  $B \in H$ . Hence, each  $B \in H$  leaves every eigenspace of  $A^N$  invariant. Since  $H \leq G'$ , also each  $V_i$  is  $H$ -invariant. We can therefore refine  $V_i = W_{i1} \oplus \dots \oplus W_{is_i}$  into a decomposition of joint-eigenspaces of all  $A^N$ , where  $A \in H$ . Then each  $W_{ij}$  is  $H$ -invariant; in particular the refined decomposition  $V = \bigoplus_{i,j} W_{ij}$  is a weakly epimonomial decomposition for  $H$ .

Let  $T$  be the diagonal of  $H$  with respect to the refined decomposition. Then  $H/T$  is a linear torsion group by Proposition 7.2, and it has exponent dividing  $N$ . Since  $\text{char } K = 0$ , this already implies that  $H/T$  is finite (Theorem 7.4). Since  $G/G'$ ,  $G'/H$ , and  $H/T$  are finite, we conclude that  $G/T$  is finite.

(b)  $\Rightarrow$  (a): Identifying  $\text{GL}(V) \cong \text{GL}_d(K)$  along a suitable basis, we find a finite-index subgroup  $T \leq G$  of diagonal matrices. Hence, property (f) of Theorem 7.5 holds. Taking  $N = (G : T)$  we see that  $\lambda^N$  is in  $K$  for every eigenvalue of every element of  $G$ .  $\square$

Now it is easy to see that if we assume  $G$  to be not just locally Bézivin, but Bézivin, it is actually necessarily finitely generated.

**Corollary 7.15.** *Suppose  $\text{char } K = 0$  and let  $G \leq \text{GL}(V)$ . If  $G$  is Bézivin, then  $G$  is finitely generated.*

*Proof.* Since the claimed property for  $G$  is purely group theoretic, and the Bézivin property is preserved under scalar extension, we may without restriction assume  $K = \overline{K}$ , so  $K$  is uniformly power-splitting for  $G$  for trivial reasons.

By Theorem 7.14 there exists a finite-index subgroup  $T \leq G$  that is simultaneously diagonalizable. Identify  $\text{GL}(V) = \text{GL}_d(K)$  using a basis in which  $T$  is diagonal. Let  $\Gamma \leq K^\times$  be finitely generated and let  $M \geq 0$  be such that  $G \subseteq (M \cdot \Gamma_0)^{d \times d}$ . Let  $\Lambda \leq K^\times$  be the subgroup consisting of all diagonal entries of elements of  $T$ . Then  $\Lambda \subseteq M \cdot \Gamma_0$ . Lemma 5.4 implies that  $\Lambda$  is finitely generated. Hence, the subgroup  $T$  is finitely generated, and because  $G/T$  is finite, also  $G$  is finitely generated.  $\square$

**7.4. Proofs of Theorem 3.1 and Corollaries 3.2 and 3.3.** With the results we proved, the main theorems on locally Bézivin representations now follow easily.

- For the proof of Theorem 3.1, note that (a)  $\Rightarrow$  (b) is trivial, and (b)  $\Leftrightarrow$  (c) follows from Proposition 7.8. Now (b)  $\Rightarrow$  (d) follows from Theorem 7.5. For (d)  $\Rightarrow$  (e) first note that weak epimonomiality passes from  $\rho(S)$  to the group it generates by Lemma 7.1. Then the claim again follows from Theorem 7.5. The final implication (e)  $\Rightarrow$  (a) follows from Lemma 7.9.
- We again work with  $G = \langle \rho(S) \rangle$ . In Corollary 3.2 the equivalence of (b), (d), and (e) follows from Theorem 7.5. The implication (a)  $\Rightarrow$  (b) is trivial, and (d)  $\Rightarrow$  (a) holds by Lemma 7.9. The equivalence of these conditions with (c) follows from Proposition 7.8.

Now (e)  $\Rightarrow$  (f) follows from Lemma 7.11. For the converse, (f)  $\Rightarrow$  (e), let  $(\widehat{V}, \widehat{\rho})$  be a monomial representation that maps epimorphically onto  $(V, \rho)$ , say via an  $S$ -equivariant  $\pi: \widehat{V} \rightarrow V$ . Because of  $\pi\widehat{\rho}(s) = \rho(s)\pi$  for all  $s \in S$ , we get a semigroup homomorphism  $\psi: \widehat{\rho}(S) \rightarrow \rho(S)$ , which extends to a group homomorphism  $\psi: \widehat{G} \rightarrow G$  with  $\widehat{G} = \langle \widehat{\rho}(S) \rangle$  and  $G = \langle \rho(S) \rangle$ . If  $l = \dim(\widehat{V})$ , then there is a group homomorphism  $\widehat{G} \rightarrow \mathfrak{S}_l$  to the symmetric group, mapping a monomial matrix to its underlying permutation matrix. Its kernel  $D$  has finite index in  $\widehat{G}$  and is simultaneously diagonal. Choose a basis  $e_1, \dots, e_l$  of  $\widehat{V}$  on which  $D$  is diagonal, and take as basis of  $V$  a subset of  $\{\pi(e_i) : i \in l\}$ . Then  $\psi(D)$  is a diagonal finite-index subgroup of  $G$ .

- In Corollary 3.3 the equivalence (a)  $\Leftrightarrow$  (b) holds by Theorem 7.14. The implication (b)  $\Rightarrow$  (c) follows again from Lemma 7.11, and the converse follows in the same way as we just argued for Corollary 3.2.

We record that (f) of Corollary 3.2 establishes the following.

**Corollary 7.16.** *Let  $\widehat{\rho}: S \rightarrow \mathrm{GL}(\widehat{V})$  and  $\rho: S \rightarrow \mathrm{GL}(V)$  be semigroup representations. If there exists an  $S$ -equivariant epimorphism  $\pi: \widehat{\rho} \rightarrow \rho$  and  $\widehat{\rho}$  satisfies the equivalent conditions of Theorem 7.5, then so does  $\rho$ .*

## 8. CHARACTERIZATION OF LOCALLY FINITELY GENERATED SPECTRUM

Again let  $K$  be a field and  $V$  a finite-dimensional vector space. To establish the characterization of linear groups with locally finitely generated spectrum, we first show that virtually solvable groups have this property. We use the following observation.

**Lemma 8.1.** *Let  $\Gamma' \leq \overline{K}^\times$  be finitely generated and  $N \geq 1$ . Then*

$$\Gamma := \{ \gamma \in \overline{K}^\times : \gamma^N \in \Gamma' \}$$

*is finitely generated as well.*

*Proof.* There is a homomorphism of abelian groups  $\varphi: \Gamma \rightarrow \Gamma'$ ,  $\gamma \mapsto \gamma^N$ . Since  $\Gamma'$  is finitely generated and  $\ker(\varphi) = \mu_N(K)$  is finite, also  $\Gamma$  is finitely generated.  $\square$

**Lemma 8.2.** *If  $G$  is a virtually solvable group, then every representation  $\rho: G \rightarrow \mathrm{GL}(V)$  has locally finitely generated spectrum.*

*Proof.* We may assume that  $K$  is algebraically closed and replace  $G$  by  $\rho(G)$ . Since  $G$  is virtually solvable, a theorem of Mal'cev [Weh73, Theorem 3.6] implies that there exists a finite-index subgroup  $U \leq G$  that is simultaneously triangularizable.<sup>4</sup>

Let  $H$  be a finitely generated subgroup of  $G$ . Then  $H \cap U$  has finite index in  $H$ , and is thus also finitely generated. Let  $\Gamma' \leq \overline{K}^\times$  be generated by the eigenvalues of the elements of  $H \cap U$ . Because  $H \cap U$  is simultaneously triangularizable, the group  $\Gamma'$  is finitely generated. Let  $\Gamma \leq \overline{K}^\times$  be generated by all eigenvalues of  $H$  and let  $N := \exp(G/U)$ . Then  $\Gamma^N = \{\gamma^N : \gamma \in \Gamma\} \leq \Gamma'$  and hence  $\Gamma$  is finitely generated by Lemma 8.1.  $\square$

In the absolutely irreducible case locally Bézivin and locally finitely generated spectrum (plus power-splitting) coincide. Before showing this, we need one more lemma.

**Lemma 8.3.** *Let  $L/K$  be a field extension. If  $\Lambda \leq L^\times$  is a finitely generated subgroup such that  $\Lambda/(\Lambda \cap K^\times)$  is torsion and  $M \geq 0$ , then there exists a finitely generated  $\Gamma \leq K^\times$  such that  $K \cap M \cdot \Lambda \subseteq M \cdot \Gamma$ .*

*Proof.* Let  $\Gamma' := \Lambda \cap K^\times$ . By assumption, the quotient  $\Lambda/\Gamma'$  is a finitely generated abelian torsion group, and hence finite. Let  $A \subseteq \Lambda$  be a set of representatives for  $\Lambda/\Gamma'$  with  $1 \in A$ . Let  $A' \subseteq A$  be a  $K$ -basis of the finite-dimensional vector space  $\mathrm{span}_K A$ . We may without restriction assume  $1 \in A'$ . Expressing each  $\alpha \in A$  as  $\alpha = \sum_{\alpha' \in A'} c_{\alpha, \alpha'} \alpha'$  with  $c_{\alpha, \alpha'} \in K$ , we have

$$\Lambda \subseteq \bigcup_{\alpha \in A} \alpha \Gamma' \subseteq \bigcup_{\alpha \in A} \sum_{\alpha' \in A'} c_{\alpha, \alpha'} \alpha' \Gamma' \subseteq \sum_{\alpha' \in A'} \alpha' \Gamma,$$

where  $\Gamma$  is the group generated by  $\Gamma'$  and the finitely many coefficients  $c_{\alpha, \alpha'}$  that are nonzero. Thus,

$$M \cdot \Lambda \subseteq \sum_{\alpha' \in A'} \alpha' (M \cdot \Gamma).$$

Linear independence of  $A'$  implies  $M \cdot \Lambda \cap K \subseteq M \cdot \Gamma$ .  $\square$

**Proposition 8.4.** *Let  $S \subseteq \mathrm{GL}(V)$  be an absolutely irreducible semigroup. Then the following statements are equivalent.*

- (a) *The semigroup  $S$  has locally finitely generated spectrum and  $K$  is a power-splitting field for  $S$ .*
- (b) *For every finitely generated subsemigroup  $S' \subseteq S$ , the set  $\mathrm{Tr}(S')$  is Bézivin.*
- (c) *The semigroup  $S$  satisfies the equivalent conditions of Theorem 7.5.*

<sup>4</sup>Alternatively, consider the Zariski closure  $\tilde{G}$  and apply the Lie–Kolchin Theorem [Bor91, Theorem 10.5] to see that the connected component of the identity is triangularizable.



*Proof.* (a)  $\Rightarrow$  (b): By assumption  $S'$  has finitely generated spectrum. Let  $\Lambda \leq \overline{K}^\times$  be finitely generated such that all eigenvalues of elements of  $S'$  are contained in  $\Lambda$ . Then  $\text{Tr}(S') \subseteq d \cdot \Lambda \cap K$ . Since  $K$  is power-splitting for  $S$ , the group  $\Lambda/(\Lambda \cap K^\times)$  is torsion. Lemma 8.3 implies that  $\text{Tr}(S')$  is Bézivin.

(b)  $\Rightarrow$  (c): We verify that  $S$  is locally Bézivin. Without restriction  $\text{GL}(V) = \text{GL}_d(K)$ . Because  $S$  is absolutely irreducible, there exists a  $K$ -basis  $A_1, \dots, A_{d^2} \in S$  of  $K^{d \times d}$  (see for instance [CR90, Theorems 3.32 and 3.43] for this classical result by Burnside). The map

$$K^{d \times d} \rightarrow K^{d^2}, \quad X \mapsto (\text{Tr}(XA_1), \dots, \text{Tr}(XA_{d^2}))$$

is a vector space isomorphism. In particular there exist  $c_{ijk} \in K$  such that  $x_{ij} = \sum_{k=1}^{d^2} c_{ijk} \text{Tr}(XA_k)$  for all  $X = (x_{ij}) \in K^{d \times d}$ . Because of (b), the sets  $\sum_{k=1}^{d^2} c_{ijk} \text{Tr}(XA_k)$  are Bézivin.

(c)  $\Rightarrow$  (a): By (f) of Theorem 7.5, the group  $G := \langle S \rangle$  is in particular virtually abelian. Then Lemma 8.2 implies that  $G$  has locally finitely generated spectrum. Proposition 7.8 implies that  $K$  is power-splitting for  $G$ . Both properties then also hold for  $S$ .  $\square$

In dealing with the locally finitely generated spectrum case, we will need to pass to a finite extension field  $L$  of  $K$ . It is not trivial, but still true, that the resulting weakly epimonomial  $L$ -representations are weakly epimonomial when viewed as  $K$ -representations, *as long as  $K$  is power-splitting for  $S$* . This is the following lemma.

**Lemma 8.5.** *Let  $L/K$  be a finite field extension and let  $W$  be a finite-dimensional  $L$ -vector space. If  $S \subseteq \text{GL}(LW)$  is a weakly epimonomial subsemigroup and  $K$  is power-splitting for  $S$ , then also  $S \subseteq \text{GL}(KW)$  is weakly epimonomial.*

*Proof.* For clarity, we say  $A \in \text{GL}(LW)$  is  $L$ -steady when  $A$  is steady as an  $L$ -vector space endomorphism, and  $K$ -steady when  $A$  is considered as a  $K$ -vector space endomorphism. If  $A$  is  $K$ -steady, then it is also  $L$ -steady: indeed, all eigenvalues of  $A$  are in  $L$ , because they are in  $K$ , and if  $U \subseteq W$  is an  $A^n$ -invariant  $L$ -subspace of  $W$ , then  $U$  is also an  $A^n$ -invariant  $K$ -subspace of  $W$ , so that  $AU \subseteq U$  by  $K$ -steadiness of  $A$ .

Let  $LW = W_1 \oplus \dots \oplus W_r$  be a weakly epimonomial decomposition of  $W$ . So each  $W_i$  is contained in an  $L$ -eigenspace of every  $L$ -steady element of  $S \subseteq \text{GL}(LW) \subseteq \text{GL}(KW)$ . Property (2) of Definition 2.6 does not depend on the base field, and (3) holds by assumption. We have to verify (1) when this decomposition is considered as one into  $K$ -vector spaces. However, if  $A \in S$  is  $K$ -steady, then  $A$  is  $L$ -steady, and hence  $W_i$  is contained in an  $L$ -eigenspace of  $A$ , which is then also contained in a  $K$ -eigenspace.  $\square$

We prove our main results on representations with locally finitely generated spectrum.

*Proof of Theorem 3.4.* (a)  $\Rightarrow$  (b): Passing to a suitable finite extension field  $L$  of  $K$  we can find in  $\widehat{V} := L \otimes_K V$  a chain of  $S$ -invariant  $L$ -vector spaces

$$0 = W_0 \subsetneq W_1 \subsetneq \cdots \subsetneq W_r = \widehat{V},$$

such that each  $W_i/W_{i-1}$  is absolutely irreducible. The resulting  $L$ -representation  $\bar{\rho}: S \rightarrow \mathrm{GL}(\widehat{V})$  is block-triangular with respect to this chain. Since  $\bar{\rho}$  has locally finitely generated spectrum as an  $L$ -representation (as the eigenvalues do not change under scalar extension), the diagonal blocks  $\bar{\rho}_{ii}$  also have locally finitely generated spectrum. By Proposition 8.4, each  $\bar{\rho}_{ii}$  is weakly epimonomial.

We now consider  $\bar{\rho}$  as a  $K$ -representation and for clarity we change the name to  $\widehat{\rho}: S \rightarrow \mathrm{GL}(\widehat{V})$ . By Lemma 8.5 the diagonals  $\widehat{\rho}_{ii}$  (with respect to the block structure from before) are still weakly epimonomial, hence locally epimonomial by Theorem 7.5.

Let  $b_1 = 1, b_2, \dots, b_m \in L$  be a  $K$ -basis for  $L$ , so that  $\widehat{V} \cong \bigoplus_{i=1}^m b_i V$ , as  $K$ -vector space. Then  $\widehat{\rho}(s)(b_1 v_1 + \cdots + b_m v_m) = b_1 \rho(s)(v_1) + \cdots + b_m \rho(s)(v_m)$  for all  $s \in S$  and  $v_i \in V$ . The inclusion  $\iota: V \hookrightarrow \widehat{V}, v \mapsto 1 \cdot v$  is split by the epimorphism  $\pi: \widehat{V} \rightarrow V$  that maps  $b_1$  to 1 and  $b_i$  to 0 for  $i \neq 1$ . Since  $\pi$  is  $S$ -equivariant, we get an epimorphism  $(\widehat{V}, \widehat{\rho}) \rightarrow (V, \rho)$ .

In  $V$ , we obtain a chain of  $S$ -invariant subspaces

$$0 = \pi(W_0) \subseteq \pi(W_1) \subseteq \cdots \subseteq \pi(W_r) = V.$$

This chain induces a block-triangular structure on  $\rho$ . The diagonal block of  $\rho$  on  $\pi(W_i)/\pi(W_{i-1})$  is the image of the diagonal block  $\widehat{\rho}_{ii}$  of  $\widehat{\rho}$  on  $W_i/W_{i-1}$ , and it is therefore locally epimonomial (Corollary 7.16).

(b)  $\Rightarrow$  (a): Let  $S' \subseteq S$  be a finitely generated subsemigroup, and let  $G = \langle \rho(S') \rangle$ . Then  $G$  has epimonomial diagonal blocks. In a suitable basis, therefore  $G$  has a finite-index subgroup  $U$  consisting of upper-triangular matrices. Since  $U$  is solvable, the group  $G$  is virtually solvable. From Lemma 8.2 it follows that  $G$  has locally finitely generated spectrum. Because the diagonal blocks of  $G$  are epimonomial, also  $K$  is power-splitting for  $G$ .  $\square$

*Proof of Corollary 3.5.* By Theorem 3.4 the representation  $S \hookrightarrow \mathrm{GL}(V)$  can be assumed to have a block-triangular structure with locally epimonomial diagonal blocks. Let  $G = \langle S \rangle$  and consider  $\rho: G \hookrightarrow \mathrm{GL}(V)$ . The diagonal blocks  $\rho_{ii}$  are virtually simultaneously diagonalizable by Corollaries 3.2 and 3.3. Let  $D_{ii} \triangleleft \rho_{ii}(G)$  be a finite-index simultaneously diagonalizable subgroup for each  $i$ , and consider

$$N = \begin{pmatrix} D_{11} & \rho_{12}(G) & \cdots & \rho_{1n}(G) \\ 0 & D_{22} & \cdots & \rho_{2n}(G) \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & D_{rr} \end{pmatrix}$$

Then  $N \leq \rho(G)$  is a subgroup of finite index. The induced representation  $\text{Ind}_N^G(\tau|_N)$  is block-triangular with monomial diagonal blocks and has an epimorphism to  $\tau$  by Lemma 7.11.  $\square$

## 9. APPLICATION TO AUTOMATA

Let  $K$  be a field and  $X$  an alphabet (a nonempty finite set). We set our notation for weighted finite automata (WFA), but will emphasize the (equivalent) viewpoint of *linear representations* (of WFA), because it is easier to work with rigorously. We refer to the books [Sak09; BR11] for more background. In this context, a **linear representation** is a triple  $\mathcal{A} = (u, \mu, v)$  with a vector  $u \in K^{1 \times d}$  of **initial weights**, a vector  $v \in K^{d \times 1}$  of **terminal weights** and a semigroup homomorphism  $\mu: X^* \rightarrow K^{d \times d}$  (describing the **transitions**). Thus, to each letter  $x \in X$  we assign a matrix  $\mu(x)$ . We say that  $\mathcal{A}$  is **invertible** if  $\mu(x)$  is an invertible matrix for every  $x \in X$  (the inverse need not be contained in  $\mu(X^*)$ ).

To  $\mathcal{A}$  we assign a function  $\llbracket \mathcal{A} \rrbracket: X^* \rightarrow K$  by defining  $\llbracket \mathcal{A} \rrbracket(w) := u\mu(w)v$ . The function  $\llbracket \mathcal{A} \rrbracket$  is the **behavior** of  $\mathcal{A}$  and is typically represented as a noncommutative (formal) series

$$\llbracket \mathcal{A} \rrbracket = \sum_{w \in X^*} \llbracket \mathcal{A} \rrbracket(w) w \in K\langle\langle X \rangle\rangle.$$

By the fundamental Kleene–Schützenberger Theorem the series representable in this way are precisely the noncommutative rational series [BR11, Chapter 1.7]. Any given rational series can be represented by many linear representations. However, the minimal linear representations, that is, those of minimal dimension, are unique up to conjugation by an invertible matrix [BR11, Theorem 2.4]. Two linear representations  $\mathcal{A}, \mathcal{A}'$  are **equivalent** if  $\llbracket \mathcal{A} \rrbracket = \llbracket \mathcal{A}' \rrbracket$ .

From a computational perspective, we may think of a linear representation  $\mathcal{A}$  of dimension  $d$  as a **weighted finite automaton** (WFA) [BR11, Chapter 1.6] on the set of states  $Q = [d]$  as follows: A state  $p \in Q$  is assigned an initial weight  $u_p$  and terminal weight  $v_p$  (where the subscripts denote the corresponding coordinates of the vectors). For any two states  $p, q \in Q$  (not necessarily distinct) and every letter  $x \in X$  we have a directed edge (transition)  $p \rightarrow q$  with weight  $\mu(x)_{pq}$  and label  $x$  if  $\mu(x)_{pq} \neq 0$ ; and no edge if  $\mu(x)_{pq} = 0$ .

Given a word  $w = x_1 \cdots x_l \in X^*$ , a **run** for  $w$  is then a finite sequence  $(p_0, x_1, p_1), (p_1, x_2, p_2), \dots, (p_{l-1}, x_l, p_l)$  with  $p_i \in Q$ . The **weight** (or **output**) of a run is the product

$$u_{p_0} \mu(x_0)_{p_0 p_1} \mu(x_1)_{p_1 p_2} \cdots \mu(x_l)_{p_{l-1} p_l} v_{p_l},$$

and the run is **successful** if its output is nonzero. The **output** of the word  $w$  is then the sum of the output of all runs for  $w$ .

Thus, a WFA is represented as a directed graph (with loops), where each edge has a label (a letter) and a weight (an element of  $K$ ). The initial and terminal weights are represented by weighted edges that are only attached to the digraph by their sinks,

respectively, sources. A successful run for  $w \in X^*$  is then a walk in the digraph, from an initial state to a terminal state, in which the edges are labelled by the letters of  $w$ .

From the way matrix multiplication is defined, it is easily seen that the output for  $w$  is just  $\llbracket A \rrbracket(w)$ . Indeed, every successful run corresponds to a nonzero summand when we fully expand the matrix product  $u\mu(x_1)\cdots\mu(x_l)v$ . We think of the WFA as *computing* the function  $\llbracket A \rrbracket: X^* \rightarrow K$ , that is, inputting a word  $w$ , the value  $\llbracket A \rrbracket(w)$  is computed.

We say that  $\mathcal{A}$  is trim if every state lies on a successful run.

The correspondence between WFA and linear representations is bijective up to a relabelling of the states (that is, conjugation by a permutation matrix on the linear representation side). Since such a relabelling will not matter for any properties we consider, we shall tacitly identify WFA and their linear representations in the following.

**Definition 9.1.** A WFA  $\mathcal{A}$  is

- (1) *deterministic (or sequential)* if there is at most one initial state (that is, at most one  $p$  with  $u_p \neq 0$ ), and for every  $p \in Q$  and  $x \in X$  there exists at most one outgoing transition  $p \rightarrow q$  labelled by  $x$  (that is, the  $p$ -th row of  $\mu(x)$  has at most one nonzero entry),
- (2)  *$M$ -ambiguous* ( $M \geq 0$ ) if every word has at most  $M$  successful runs,
- (3) *unambiguous* if it is 1-ambiguous,
- (4) *finitely ambiguous* if it is  $M$ -ambiguous for some  $M$ ,
- (5) *polynomially ambiguous* if there is a constant  $C$  and an exponent  $k \geq 0$  such that every word has at most  $C|w|^k$  successful runs, where  $|w|$  is the length of the word.

For a finitely ambiguous  $\mathcal{A}$ , the **ambiguity degree** is the minimal  $M$  for which  $\mathcal{A}$  is  $M$ -ambiguous. Finitely ambiguous, respectively, polynomially ambiguous automata can be characterized by the absence of certain characteristic features of the underlying graph [WS91, Theorems 3.1 and 4.1]. As a consequence, every WFA that is not polynomially ambiguous is **exponentially ambiguous**, that is, there are words  $u, v, w$  such that  $uw^n v$  has at least  $2^n$  successful runs.

*Remark 9.2.* More specifically, for a WFA  $\mathcal{A}$ , let  $g(n)$  be the maximal number of successful runs for a word of length  $n$ . Then trivially  $g(n) \in O(d^n)$  where  $d$  is the number of states. The automaton  $\mathcal{A}$  is finitely ambiguous if  $g(n) \in O(1)$  and polynomially ambiguous if  $g(n) \in O(n^k)$  for some  $k \geq 1$ . The characterizations in [WS91] imply that there are *gaps* in the possible growths of the function  $g(n)$ : we always have  $g(n) \in O(d^n)$ , and if  $g(n) \in o(c^n)$  for all  $c \in \mathbb{R}_{>1}$ , then already  $g(n) \in O(n^k)$  for some  $k \geq 1$ . So any automaton with subexponential ambiguity is already polynomially ambiguous. Similarly, if  $g(n) \in o(n)$  then  $g(n) \in O(1)$ , so any automaton with sublinear ambiguity is already

finitely ambiguous. From this point of view, the classes considered in Definition 9.1 give a natural hierarchy.

A trim WFA that is polynomially ambiguous is also cycle-unambiguous, that is, for every  $p \in Q$  and every word  $w$ , there exists at most one walk from  $p$  to itself that is labeled by  $w$ . This is easy to see: otherwise, there would be some  $q \in [d]$  with two distinct directed cycles, say of lengths  $k_1$  and  $k_2$  based off it; then  $w^{\text{lcm}(k_1, k_2)n}$  has at least  $2^n$  closed directed walks based at  $q$ ; using trimness, this contradicts polynomial ambiguity. Conversely, cycle-unambiguous WFA are polynomially ambiguous [WS91, Theorem 4.1], but this is harder to prove, and we will not need it.

We are now ready to prove the easier direction of the arithmetic characterization of ambiguity classes. Here we do not need to assume invertibility of the WFA.

**Proposition 9.3.** *Let  $\mathcal{A}^\circ = (u^\circ, \mu^\circ, v^\circ)$  be a trim linear representation and let  $\mathcal{A} = (u, \mu, v)$  be a minimal linear representation with  $\llbracket \mathcal{A} \rrbracket = \llbracket \mathcal{A}^\circ \rrbracket$ .*

- (1) *If  $\mathcal{A}^\circ$  is  $M$ -ambiguous, then there exists a finitely generated  $\Gamma \leq K^\times$  such that  $\llbracket \mathcal{A} \rrbracket(w) \in M \cdot \Gamma_0$  for all  $w \in X^*$ .*
- (2) *If  $\llbracket \mathcal{A}^\circ \rrbracket(X^*)$  is Bézivin, then so is  $\mu(X^*)$ . In particular, if  $\mathcal{A}^\circ$  is finitely ambiguous, then  $\mu(X^*)$  is Bézivin.*
- (3) *If  $\mathcal{A}^\circ$  is polynomially ambiguous, then  $\mu(X^*)$  has finitely generated spectrum<sup>5</sup> and  $K$  is uniformly power-splitting for  $\mu(X^*)$ .*

*Proof.* (1) Let  $\Gamma$  be the group generated by all weights of  $\mathcal{A}$ , that is, by the entries of  $u$ ,  $v$  and  $\mu(x)$  for  $x \in X$ . The product of weights along a walk in  $\mathcal{A}$  gives an element of  $\Gamma$ , and since every word has at most  $M$  successful runs, the claim follows.

(2) Let  $M \geq 0$  and  $\Gamma \leq K^\times$  be finitely generated such that  $u\mu(w)v \in M \cdot \Gamma_0$  for all  $w \in X^*$ . Because  $\mathcal{A}$  is minimal, there exists a basis of  $K^{1 \times d}$  of the form  $u\mu(w_1), \dots, u\mu(w_d)$  with  $w_i \in X^*$ , and a basis of  $K^{d \times 1}$  of the form  $\mu(w'_1)v, \dots, \mu(w'_d)v$  with  $w'_i \in X^*$  [BR11, Proposition 2.2.1]. Let  $e_1, \dots, e_d$  denote the standard basis of  $K^{d \times 1}$  and  $e_1^*, \dots, e_d^*$  its dual basis, so that the  $ij$ -entry of a matrix  $A \in K^{d \times d}$  is just  $e_i^* A e_j$ . We express  $e_i = \sum_{k=1}^d t_{ik} \mu(w'_k)v$  and  $e_j^* = \sum_{l=1}^d s_{lj} u\mu(w_l)$  with  $t_{ik}, s_{lj} \in K$ . For every word  $w \in X^*$ , then

$$e_i^* \mu(w) e_j = \sum_{k,l=1}^d t_{ik} s_{lj} u\mu(w_l) \mu(w) \mu(w'_k)v = \sum_{k,l=1}^d t_{ik} s_{lj} u\mu(w_l w w'_k)v.$$

Since  $u\mu(w_l w w'_k)v \in M \cdot \Gamma_0$ , therefore all entries of  $\mu(w)$  are contained in  $d^2 M \cdot \Gamma'_0$  with  $\Gamma'$  generated by  $\Gamma$ ,  $t_{ik}$ , and  $s_{lj}$ .

<sup>5</sup>The nonzero eigenvalues are contained in a finitely generated subgroup of  $\overline{K}^\times$ .

Finally, if  $\mathcal{A}^\circ$  is finitely ambiguous, then by (1) we get  $M \geq 0$  and a finitely generated  $\Gamma \leq K^\times$  such that  $u\mu(w)v \in M \cdot \Gamma_0$  for all  $w \in X^*$ . Hence, the semigroup  $\mu(X^*)$  is Bézivin by what we just showed.

(3) Since  $\mathcal{A}^\circ$  is trim, it is cycle-unambiguous. Let  $d$  be the dimension of  $\mathcal{A}^\circ$ . After conjugation by a suitable invertible matrix, we can assume that  $\mu^\circ(X^*)$  has block-triangular shape with one of the diagonal blocks being  $\mu(X^*)$  [BR11, Corollary 2.2.5]. It is therefore sufficient to establish that  $\mu^\circ(X^*)$  has finitely generated spectrum and that  $K$  is uniformly power-splitting for  $\mu^\circ(X^*)$ . Let  $\Gamma' \leq K^\times$  be the subgroup generated by the weights of  $\mathcal{A}^\circ$ .

Fix a word  $w$  and consider the following directed graph (with loops)  $G(w)$ : the vertex set is  $[d]$  and there is an edge  $i \rightarrow j$  if there is a directed walk in  $\mathcal{A}^\circ$  from  $i$  to  $j$  that is labeled by  $w$ . Since  $\mathcal{A}^\circ$  is cycle-unambiguous, the non-trivial strongly connected components of  $G(w)$  must be directed cycles<sup>6</sup>, each of length at most  $d$ . It follows that the only non-trivial strongly connected components in  $G(w^{d!})$  are loops. Therefore, the states  $[d]$  can be totally ordered in such a way that  $i < j$  implies that there is no path from  $j$  to  $i$  in  $G(w^{d!})$ .

After a corresponding permutation of the standard basis vectors, the matrix  $\mu^\circ(w^{d!}) = \mu^\circ(w)^{d!}$  is upper-triangular. Therefore, its diagonal entries are precisely its eigenvalues. The  $p$ -th diagonal entry of  $\mu^\circ(w^{d!})$  is obtained as follows: for each directed walk from  $p$  to  $p$  in  $\mathcal{A}^\circ$ , one first takes the product over all weights along the walk, and then sums these products over all directed walks from  $p$  to  $p$ . However, since  $\mathcal{A}^\circ$  is cycle-unambiguous, there exists at most one directed walk from  $p$  to  $p$ . Thus, the nonzero diagonal entries of  $\mu^\circ(w^{d!})$  are products of weights of  $\mathcal{A}^\circ$ , and hence are contained in  $\Gamma'$ .

Consequently, if  $\lambda \in \overline{K}$  is an eigenvalue of  $\mu(w)$ , then  $\lambda^{d!} \in \Gamma'$ . Since  $\Gamma = \{\gamma \in \overline{K} : \gamma^{d!} \in \Gamma'\}$  is finitely generated by Lemma 8.1, the claim follows.  $\square$

**Proposition 9.4.** *Let  $\mathcal{A} = (u, \mu, v)$  be an invertible linear representation.*

- (1) *If  $\mu(X^*)$  is Bézivin, then there exists a linear representation  $\widehat{\mathcal{A}} = (\widehat{u}, \widehat{\mu}, \widehat{v})$  with  $\llbracket \widehat{\mathcal{A}} \rrbracket = \llbracket \mathcal{A} \rrbracket$  and  $\widehat{\mu}$  monomial. In particular, the automaton  $\widehat{\mathcal{A}}$  is finitely ambiguous.*
- (2) *If  $\mu(X^*)$  has finitely generated spectrum and  $K$  is power-splitting for  $\mu(X^*)$ , then there exists a linear representation  $\widehat{\mathcal{A}}$  with  $\llbracket \widehat{\mathcal{A}} \rrbracket = \llbracket \mathcal{A} \rrbracket$  and  $\widehat{\mathcal{A}}$  is block-triangular with monomial diagonal blocks. In particular, the automaton  $\widehat{\mathcal{A}}$  is polynomially ambiguous.*

*Proof.* Let  $d$  be the dimension of  $\mathcal{A}$ . We first make a general observation that will be useful for both claims. Suppose  $\widehat{\mu}: X^* \rightarrow K^{n \times n}$  is a semigroup homomorphism, and  $T \in K^{d \times n}$

<sup>6</sup>A directed cycle is a closed directed walk in which all vertices except the first and the last are pairwise distinct.

has rank  $d$ , and is such that  $T\widehat{\mu}(w) = \mu(w)T$  for all  $w \in X^*$ . Since  $T: K^{n \times 1} \rightarrow K^{d \times 1}$  is surjective, there exists  $\widehat{v} \in K^{n \times 1}$  such that  $v = T\widehat{v}$ . Let  $\widehat{u} = uT$ . Then  $\widehat{\mathcal{A}} = (\widehat{u}, \widehat{\mu}, \widehat{v})$  is equivalent to  $\mathcal{A}$ , because

$$\widehat{u}\widehat{\mu}(w)\widehat{v} = uT\widehat{\mu}(w)\widehat{v} = u\mu(w)T\widehat{v} = u\mu(w)v \quad \text{for all } w \in X^*.$$

(1) By Corollary 3.2 there exists a monomial semigroup representation  $\widehat{\mu}: X^* \rightarrow \text{GL}_n(K)$  and an  $X^*$ -invariant epimorphism  $\widehat{\mu} \rightarrow \mu$ . Since  $\widehat{\mu}(X^*)$  is monomial, the resulting linear representation  $(\widehat{u}, \widehat{\mu}, \widehat{v})$  is finitely ambiguous.

(2) By Corollary 3.5 we can now find a linear representation  $(\widehat{u}, \widehat{\mu}, \widehat{v})$  with  $\widehat{\mu}$  block-triangular with monomial diagonal blocks. Such a representation is polynomially ambiguous: corresponding to the block-triangular structure, we obtain a partition of the states  $Q_1 \cup \dots \cup Q_r$  for which it is only possible to transition from a state in  $Q_i$  to one in  $Q_j$  if  $j > i$ . Furthermore, due to the monomial diagonals, transitions from states in  $Q_i$  to states in the same set  $Q_i$  are unambiguous.  $\square$

The following shows that the minimal ambiguity degree achievable for a WFA computing a particular series is at the same time the minimal  $M$  such that all outputs are contained in  $M \cdot \Gamma_0$  for some finitely generated  $\Gamma \leq K^\times$ . It is obtained by taking a slightly closer look at the construction resulting from Proposition 9.4 (and choosing a suitable basis).

**Corollary 9.5.** *Let  $\mathcal{A} = (u, \mu, v)$  be an invertible linear representation. If there exists  $M \geq 0$  and a finitely generated  $\Gamma \leq K^\times$  with  $[\![\mathcal{A}]\!](w) \in M \cdot \Gamma_0$  for all  $w \in X^*$ , then  $\mathcal{A}$  is equivalent to an  $M$ -ambiguous WFA.*

*Proof.* We may assume that  $\mathcal{A} = (u, \mu, v)$  is a minimal linear representation (using [BR11, Corollary 2.2.5]). By (2) of Proposition 9.3, then  $\mu(X^*) \subseteq \text{GL}_d(K)$  is Bézivin. Let  $G := \langle \mu(X^*) \rangle$ . By Theorem 7.5 the finitely generated group  $G \leq \text{GL}_d(K)$  is epimonomial. Let  $K^{d \times 1} = V_1 \oplus \dots \oplus V_r$  be an epimonomial decomposition for  $G$ , with  $D \leq G$  the corresponding diagonal. Decompose  $v = v_1 + \dots + v_r$  with  $v_j \in V_j$ , and for each  $V_j$  choose a basis  $e_{j1}, \dots, e_{js_j}$ , with  $e_{j1} = v_j$  if  $v_j \neq 0$ . After changing to this basis, we can assume that  $D$  consists of diagonal matrices and that each  $v_j$  has at most one nonzero entry.

Let  $A_1 = I, A_2, \dots, A_n$  be a set of representatives for  $G/D$  and consider the  $dn$ -dimensional representation  $\rho: G \rightarrow \text{GL}(W)$  induced from  $D \hookrightarrow \text{GL}_d(K)$ , with

$$W = \bigoplus_{i=1}^n A_i K^{d \times 1} = \bigoplus_{i=1}^n \bigoplus_{j=1}^r A_i V_j$$

(as in Lemma 7.11), and set  $W_{ij} := A_i V_j$ . We choose as basis of  $W_{ij}$  the vectors  $f_{ijk} := A_i e_{jk}$  (where  $k \in [s_j]$ ), and identify  $W = K^{dn \times 1}$ . Then  $\rho$  is monomial; explicitly, let  $A \in G$



and let  $i \in [n]$ ,  $j \in [r]$ ,  $k \in [s_j]$ . If  $i' \in [n]$  and  $C \in D$  are such that  $AA_i = A_{i'}C$  and  $C|_{V_j} = \lambda_j \text{id}$  with  $\lambda_j \in K$ , then we have

$$\rho(A)f_{ijk} = \lambda_j f_{i'jk}.$$

The  $G$ -equivariant epimorphism

$$\psi: W \rightarrow K^{d \times 1}, \quad f_{ijk} \mapsto A_i e_{jk},$$

is split by

$$\iota: K^{d \times 1} \rightarrow W, \quad e_{jk} \mapsto f_{1jk}.$$

Let  $\widehat{v} = \iota(v)$  and let  $\widehat{u} \in K^{1 \times dn}$  such that  $\widehat{u}f_{ijk} = uA_i e_{jk}$ . With  $\widehat{\mu} = \rho\mu$ , we obtain a monomial WFA  $\widehat{\mathcal{A}} = (\widehat{u}, \widehat{\mu}, \widehat{v})$  that is equivalent to  $\mathcal{A}$  (as in Proposition 9.4).

Using Lemma 7.10, choose  $A \in D$  steady and such that the eigenvalues on all  $V_j$  are pairwise distinct. Consider  $\rho(A)$ . Each  $A_i$  permutes the eigenspaces of  $A$ , say  $A_i V_j = V_{\pi_i(j)}$  with  $\pi_i \in \mathfrak{S}_r$  a permutation. Then  $A_i^{-1}AA_i$  is again diagonal with the same eigenspaces as  $A$  itself, but the eigenvalues permuted: if  $A$  has eigenvalue  $\lambda_j$  on  $V_j$ , then  $A_i^{-1}AA_i$  has the eigenvalue  $\lambda_{\pi_i(j)}$  on  $V_j$ . Setting, for  $j \in [r]$ ,

$$W_j := \bigoplus_{i=1}^n A_i V_{\pi_i^{-1}(j)} \subseteq W,$$

we see that the  $W_j$  are precisely the eigenspaces of the steady element  $\rho(A)$ . Since  $\rho(G)$  is monomial, it is certainly Bézivin, and we can apply Proposition 6.4 to the block structure  $W = W_1 \oplus \dots \oplus W_r$ . Thus, for all  $x_1, \dots, x_l \in X$  (using square brackets to denote the blocks), we have that

$$\widehat{u}[\mu(x_1)]_{i_0 i_1} [\mu(x_2)]_{i_1 i_2} \dots [\mu(x_l)]_{i_{l-1} i_l} [\widehat{v}]_{i_l} \neq 0, \quad (7)$$

for at most  $M$  choices of  $(i_0, i_1, \dots, i_l) \in [r]^{l+1}$ . We shall argue that each such nonzero block product corresponds to at most one successful run of  $\widehat{\mathcal{A}}$ .

For  $x \in X$  let  $\widehat{\mu}_{pq}(x)$  denote the  $pq$  entry of the matrix  $\widehat{\mu}(x)$ . A successful run for a word  $x_1 \dots x_l \in X^*$  is simply a tuple  $(p_0, p_1, \dots, p_l) \in [dn]$  such that

$$\widehat{u}_{p_0} \widehat{\mu}_{p_0 p_1}(x_1) \widehat{\mu}_{p_1 p_2}(x_2) \dots \widehat{\mu}_{p_{l-1} p_l}(x_l) \widehat{v}_{p_l} \neq 0.$$

This relates to the block structure in Eq. (7) as follows: Let  $V := V_1 \oplus \dots \oplus V_r = K^{d \times 1}$ . The terminal vector  $\widehat{v}$  is contained in the summand  $IV$  of  $W$ , and projected onto each  $IV_j$ , it has at most one nonzero component by the choice of the basis (namely, the component is a multiple of  $f_{1j1}$ ). By construction of the induced representation, the matrices  $\mu(x)$  permute the spaces  $A_1 V, \dots, A_n V \subseteq W$ . Thus, for each  $k \in [1, l]$  there is a unique  $i$  with  $\mu(x_k \dots x_l) \widehat{v} \in A_i V$ . Fix  $k$ . Then, for  $j \in [r]$ , there is a unique  $t \in [r]$  with  $\mu(x_k \dots x_l) f_{1j1} \in A_i V_t$ . But then in fact  $j$  uniquely determines the space  $W_{t'}$  for which  $\mu(x_k \dots x_l) f_{1j1} \in W_{t'}$  (namely  $t' = \pi_i(t)$ ). It follows that each successful run (each one

necessary starting at a distinct  $f_{1j1}$ , by monomiality) contributes to a distinct block product of the form Eq. (7). Hence, the maximal number of nonzero such block products, namely  $M$ , bounds the maximal number of successful runs for a given word. Thus, the linear representation  $\widehat{\mathcal{A}}$  is at most  $M$ -ambiguous.  $\square$

## 10. DECIDABILITY

For finitely generated semigroups we now show decidability of the properties under discussion (and that the monomial, respectively block-triangular representations with monomial blocks are indeed computable, if they exist). Obviously this requires some restrictions on the field  $K$ , in particular, we need to be able to perform exact arithmetic and equality tests, so that we can do exact linear algebra, use Gröbner bases techniques, and compute in finite field extensions of  $K$  (as splitting fields of suitable polynomials). We shall call such a field *computable*.

Most importantly, number fields (finite-dimensional extensions of  $\mathbb{Q}$ ) are computable. In general, a finitely generated semigroup  $S$  will always be contained in  $K^{d \times d}$  with  $K$  a field that is finitely generated over its prime field  $K_0$ . Then  $K$  is the field of fractions of an affine  $K_0$ -algebra  $R$ . If  $K$  can be specified by giving explicit generators and relations of  $R$  as a  $K_0$ -algebra, then  $K$  is computable. For example, for the transcendental extension  $K = \mathbb{Q}(X)$  with  $X$  a nonempty finite set we trivially know such a representation, but for  $K = \mathbb{Q}(e, \pi)$  we do not have such a representation, since we do not know whether there are algebraic relations between  $e$  and  $\pi$ . *For the remainder of the section, fix a computable field  $K$  and a vector space  $V$ .*

The linear Zariski topology on  $V$  is the topology having as closed sets the finite unions of vector subspaces (including the empty set as the empty union) [BS21; BS23b]. It is a noetherian topology. If  $K$  is infinite, the irreducible sets are precisely the subspaces of  $V$ . If  $K$  is finite, then the irreducible sets are the subspaces of dimension at most 1.

The Zariski closure and the linear Zariski closure of a finitely generated group [DJK05b; Nos+22], and even a finitely generated semigroup [Hru+18; Hru+23; BS23b], of matrices is computable.

The following observation will reduce everything to the computation of the linear Zariski closure. The argument is similar to the one in the proof of [BS23b, Theorem 10].

**Proposition 10.1.** *Let  $G \subseteq \mathrm{GL}(V)$  be a locally Bézivin group, let  $V = V_1 \oplus \cdots \oplus V_r$  be the decomposition into joint-eigenspaces of steady elements, and let  $D \trianglelefteq G$  be the corresponding diagonal. Let  $P_i \in \mathrm{End}(V)$  be such that  $P_i|_{V_i} = \mathrm{id}$  and  $P_i|_{V_j} = 0$  for  $j \neq i$ . Then the linear Zariski closure  $\overline{D}$  of  $D$  in  $\mathrm{End}(V)$  is*

$$\overline{D} = \mathrm{span}\{P_1, \dots, P_r\}.$$

In particular, if  $G/D$  is finite with set of representatives  $A_1, \dots, A_n$ , then the irreducible components of  $\overline{G}$  are

$$A_1 \text{span}\{P_1, \dots, P_r\}, \dots, A_n \text{span}\{P_1, \dots, P_r\}.$$

*Proof.* Identify  $\text{GL}(V) = \text{GL}_d(K)$  using a basis that refines the decomposition  $V = V_1 \oplus \dots \oplus V_r$ . Let  $\overline{D} = Z_0 \cup \dots \cup Z_l$  with  $Z_k \subseteq V$  the irreducible components of  $\overline{D}$ . Because  $D \subseteq \text{span}\{P_1, \dots, P_r\}$ , also  $Z_0 \cup \dots \cup Z_l \subseteq \text{span}\{P_1, \dots, P_r\}$ . Note that  $\overline{D} \cap \text{GL}_d(K)$  is the set of  $K$ -rational points of a linear algebraic group. There exists a unique component, say  $Z_0$ , containing the identity, and  $Z_0 \cap \text{GL}_d(K)$  is a subgroup of  $G$  [BS23b, Lemma 9].

By Lemma 7.10 there exists a steady  $A \in G$  that has distinct eigenvalues on the  $V_i$ , say  $A = \lambda_1 P_1 + \dots + \lambda_r P_r$  with  $\lambda_i \in K$  such that  $\lambda_i/\lambda_j$  is not a root of unity for  $i \neq j$ . Since multiplication by  $A$  permutes the components  $Z_0, \dots, Z_l$ , there exists an  $N \geq 1$  such that  $A^N \in Z_0$ . Then  $A^{Nn} = \lambda_1^{Nn} P_1 + \dots + \lambda_r^{Nn} P_r \in Z_0$  for all  $n \in \mathbb{Z}$ , and, since  $Z_0$  is a vector space, we have  $W := \text{span}\{A^{Nn} : n \in \mathbb{Z}\} \subseteq Z_0$ . Keeping in mind that  $\lambda_i^N \neq \lambda_j^N$  for all  $i \neq j \in [r]$ , a Vandermonde determinant argument shows  $\dim W = r$ . Because of  $W \subseteq Z_0 \subseteq \text{span}\{P_1, \dots, P_r\}$  we must have equality throughout, hence  $\overline{D} = \text{span}\{P_1, \dots, P_r\}$ .  $\square$

**Proposition 10.2.** *For a finitely generated semigroup  $S \subseteq \text{GL}_d(K)$ , given by generators  $B_1, \dots, B_m \in S$ , it is decidable whether  $S$*

- (1) *is Bézivin;*
- (2) *has finitely generated spectrum and  $K$  is power-splitting for  $S$ .*

*One can moreover compute a representation  $\rho: S \rightarrow \text{GL}_{d'}(K)$  and an epimorphism of representations  $\rho \twoheadrightarrow (S \hookrightarrow \text{GL}_d(K))$  such that  $\rho$  is monomial in the first case, and such that  $\rho$  is block-triangular with monomial diagonal blocks in the second case.*

*Proof.* The linear Zariski closure  $\overline{S}$  is computable; more explicitly, one can compute a basis of the irreducible component  $Z_0$  containing the identity, and  $A_0 = I, A_1, \dots, A_n \in S$  such that  $\overline{S} = \bigcup_{i=0}^n A_i Z_0$ , with  $A_i Z_0 \neq A_j Z_0$  for  $j \neq i$ . The closure  $\overline{S}$  contains the identity matrix and  $A^{-1}$  for every  $A \in S$ , so that in fact  $\overline{S}$  is also the closure of the group  $G$  generated by  $S$  [BS23b, Lemma 9].

(1) In light of Proposition 10.1 and Theorem 7.5, it is enough to check whether  $Z_0$  is simultaneously diagonalizable over  $K$ . This is easily done iteratively: suppose the component  $Z_0$  is not simultaneously diagonal in the given basis. We can then compute a non-diagonal  $A \in Z_0 \cap \text{GL}_d(K)$  using [BS23b, Lemma 17] and the fact that  $Z_0 \cap \text{GL}_d(K) \supseteq Z_0 \cap S$  is dense in  $Z_0$  in the linear Zariski topology. If  $A$  is not diagonalizable over  $K$ , then  $S$  is not Bézivin and we abort. Otherwise, we change the basis to one in which  $A$  is

diagonal. We now check if  $Z_0$  has block-diagonal structure with respect to the eigenspaces of  $A$  (it is sufficient to check this on the basis vectors): if not, then  $S$  is not Bézivin, and we abort. If  $Z_0$  has the required block-diagonal structure, but is not diagonal, then we can pick a non-diagonal  $A' \in Z_0 \cap \text{GL}_d(K)$  (again using [BS23b, Lemma 17]), and we attempt to diagonalize it. This will either result in a finer-grained block-diagonal structure, or show that  $S$  is not Bézivin. Repeating this finitely many times, we will either have diagonalized all of  $Z_0$  (showing that  $S$  is Bézivin), or discovered that  $S$  is not Bézivin.

(2) We first compute a finite field extension  $L/K$  and a chain of  $G$ -invariant  $L$ -subspaces  $0 = W_1 \subsetneq W_2 \subsetneq \dots \subsetneq W_k = L^{d \times 1}$  such that each  $W_i/W_{i-1}$  is absolutely irreducible as a  $G$ -representation (that is, there exists no  $G$ -invariant  $\overline{K}$ -subspace  $\overline{K} \otimes_K W_{i-1} \subsetneq W' \subsetneq \overline{K} \otimes_K W_i$ ). This is possible, because the  $\overline{K}$ -invariant subspaces of a  $K$ -vector space  $V$  can be described as the  $\overline{K}$ -points on a projective variety that is defined over  $K$ , by using exterior powers and Plücker coordinates, see [AP04]. Testing whether this variety is empty can be done using Gröbner basis techniques, and if it is not empty, one can compute a finite field extension  $L/K$  and a point defined over  $L$ ,<sup>7</sup> giving rise to an invariant subspace. Iterating this procedure finitely many times, we find the required decomposition.

Viewing  $L^{d \times 1}$  as a  $K$ -vector space, we obtain a  $K$ -representation  $\rho$  of  $G$ . If  $G$  has locally finitely generated spectrum and  $K$  is power-splitting for  $G$ , then the resulting diagonal blocks of  $\rho$  are epimonomial (using Proposition 8.4 and Theorem 7.5), and we can check this using (1). So, if one of the diagonal blocks is not epimonomial, then  $G$  did not have locally finitely generated spectrum or  $K$  is not power-splitting. If they are all epimonomial, then  $S \rightarrow \text{GL}_d(K)$  factors through a representation with monomial diagonal blocks by Corollary 3.5.

Computing the claimed representations is now a matter of computing suitable induced representations. To compute the induced representation, we need a set of representatives for  $G/(G \cap Z_0)$ . However, such a set is given by the elements  $A_0, \dots, A_n$  computed at the beginning.  $\square$

**Corollary 10.3.** *Let  $\mathcal{A}$  be a linear representation of a WFA.*

- (1) *It is decidable whether  $\mathcal{A}$  is equivalent to a polynomially ambiguous, respectively, a finitely ambiguous automaton.*
- (2) *If  $\mathcal{A}$  is equivalent to a finitely ambiguous automaton, it is also decidable which is the minimal  $M \geq 0$  such that  $\mathcal{A}$  is  $M$ -ambiguous*

<sup>7</sup>It suffices to treat the affine case. One possibility is outlined in [Hru+23, Section 3.3], based on the fact that one can compute projections of *constructible* sets [Sch07, Corollary 4.12] [CLO15, Chapter 4.7].

*Proof.* (1) From a linear representation one can compute a minimal linear representation [BR11, Chapter 2.3]. So without restriction assume that  $\mathcal{A} = (u, \mu, v)$  is minimal. By Propositions 9.3 and 9.4 it is now sufficient to decide whether  $\mu(X^*)$  is Bézivin, respectively, whether  $\mu(X^*)$  has finitely generated spectrum and  $K$  is power-splitting for  $\mu(X^*)$ . These properties are decidable by Proposition 10.2.

(2) The construction described in Corollary 9.5 produces an automaton  $(u, \mu, v)$  of minimal ambiguity degree. Hence, it is a matter of computing its ambiguity.

The matrices in  $\mu(X^*)$  are all monomial matrices, so that any application of a matrix  $\mu(x)$  permutes the entries in the vector and then possibly multiplies them by (different) weights. Thus, for every word  $w \in X^*$  the vector  $\mu(w)v$  will have the same number of nonzero entries as the vector  $v$ , but possibly in different places.

The number of successful runs for  $w$  is then simply the number of  $i \in [d]$  for which the  $i$ -th entry of both  $u$  and  $\mu(w)v$  is nonzero. Deciding the degree of ambiguity therefore boils down to finding the possibly patterns of nonzero entries in  $\mu(w)v$  as  $w$  ranges through  $X^*$ . This is easily done, because mapping each  $\mu(w)$  to the corresponding permutation matrix  $\overline{\mu(w)}$  (replacing every nonzero entry by 1) is a group homomorphism to a finite group, and so we can compute all the permutations  $\{\overline{\mu(w)} : w \in X^*\}$ .  $\square$

*Remark 10.4.* The algorithm also produces a polynomially ambiguous, respectively, finitely ambiguous automaton. Since Corollary 9.5 is constructive, we can also find a finitely ambiguous automaton of minimal ambiguity degree  $M \geq 0$  (if there exists one). Because of the monomial structure, the only ambiguity arises from the multiple initial and terminal states. The constructed automaton can therefore be easily transformed into a sum of  $M$  deterministic automata. (It is known that every finitely ambiguous automaton is a sum of unambiguous automata.)

*Remark 10.5 (Nonefficiency).* The algorithms presented here are conceptually simple, but not efficient. For computing the linear Zariski closure of a semigroup, the only known upper bounds are *double-exponential* in the dimension, as recently shown by Benalioua, Lhote, and Reynier [BLR24]. The linear Zariski closure can have *super-exponentially* many irreducible components [BS23b, Remark 7], and so we cannot expect efficient decidability algorithms that first compute the entire (linear) Zariski closure.

- (1) In case we are willing to decide existence of a block-triangular representation with monomial diagonal blocks over a finite field extension (so that the power-splitting condition plays no role), checking virtual solvability, that is, deciding the Tits' alternative, is sufficient by Corollary 1.4.

This algorithmic problem is well-studied and efficient and practical algorithms exist. For instance, Beals [Bea99; Bea01] proved that, over a number field, virtual

solvability is decidable in *polynomial time*. Further algorithms can be found in the works of Detinko, Flannery, and O’Brien [DFO11; DFO11]. For subgroups of  $\mathrm{GL}_d(\mathbb{Z})$ , virtual solvability is equivalent to virtual polycyclicity, which can be decided by work of Baumslag, Cannonito, Robinson, and Segal [Bau+91] and Ostheimer [Ost99].

- (2) Very recently, Jecker, Mazowiecki, and Purser [JMP24] showed (over  $K = \mathbb{Q}$ ): if one is given a polynomially ambiguous WFA, then it is possible to decide whether it is equivalent to a deterministic or an unambiguous one in *polynomial space* ( $PSPACE$ ).

## 11. CONCLUSION AND OPEN QUESTIONS

In the present paper we comprehensively resolve the problem of characterizing, by arithmetic properties, the representability of a linear *group* by monomial respectively block-triangular matrices with monomial diagonal blocks over any field. For WFA *with invertible transition matrices* we used this to relate ambiguity classes of rational series to arithmetic properties, and resolve the arising decidability questions. In particular, it is possibly to decide whether an invertible WFA is equivalent to a finitely ambiguous, respectively, a polynomial ambiguous one.

- (1) Given the applications to WFA, the most pressing issue is whether analogous results to the main theorem can be established for finitely generated semigroups  $S \subseteq \mathrm{End}(V)$  of possibly non-invertible matrices, that is, to find sufficiently nice representations of Bézivin semigroups and semigroups with finitely generated spectrum.

Not every such semigroup can have a monomial representation. For instance, if  $\lambda \in \mathbb{Q} \setminus \{\pm 1\}$  the semigroup generated by

$$\begin{pmatrix} \lambda & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$$

has infinitely many idempotents  $\begin{pmatrix} 0 & \lambda^n \\ 0 & 1 \end{pmatrix}$ , which is impossible in an epimorphic image of a monomial semigroup. However, there are some positive results.

- (i) In the case that all outputs of the WFA are contained in  $\Gamma_0$  (Pólya property), it is known that the linear representation of the WFA factors through a *semimonomial* representation [BS21]. Here the matrices have a block structure: each row of blocks contains at most one nonzero blocks, and each column of each block contains at most one nonzero entry.
- (ii) There are several proofs showing that finitely generated torsion matrix semigroups are finite, that is, Burnside–Schur extends to semigroups (see [Ste12] for many references).

- (iii) Suppose that  $S$  contains an invertible steady  $A$  that has  $d$  distinct eigenvalues. Then applying Corollary 6.5 immediately shows that  $S$  is finitely ambiguous, and so we get the following.

**Corollary 11.1.** *Let  $\mathcal{A} = (u, \mu, v)$  be a  $d$ -dimensional linear representation of a WFA over the field  $K$ . If  $\mu(\Sigma^*)$  is a Bézivin semigroup, and it contains at least one steady invertible matrix that has  $d$  pairwise distinct eigenvalues, then a change of basis transforms  $\mathcal{A}$  into a finitely ambiguous linear representation.*

The algebraic structure of finitely generated matrix semigroups and their Zariski closure is well-understood based on work of Putcha and Okniński [Put88; Okn98; Ren05]. In particular, if one considers the (linear) Zariski closure, there is a finite chain of semigroup ideals, the successive Rees quotients of which are either completely 0-simple or nilpotent. This structure has already played a key role in the computation of the (linear) Zariski closure [Hru+18; Hru+23; BS23b]. While the semigroup case makes use of the group case, in these instances, it turned out that the step from groups to arbitrary semigroups is the most complex one.

- (2) From the point of view of theoretical computer science, it would be interesting to establish which complexity classes the decidability problems arising in Section 10 actually fall into. This is particularly interesting because the algorithms based on the (linear) Zariski closure are necessarily quite inefficient, while more efficient algorithms are already known for certain subproblems and the decidability of virtual solvability (Remark 10.5).

Further, there is a trade-off between ambiguity and the number of states. From this point of view it is interesting to what degree our constructions are optimal (in minimizing the number of states in the given complexity class) and what the minimum number of states for a WFA in a given complexity class, as a function of the number of states of a minimal WFA, is.

Little appears to be known here. For instance, it is well-known that the maximal subgroups of  $\mathrm{GL}_d(\mathbb{Q})$  have cardinality  $2^d d!$  for  $d$  sufficiently large (and can be realized by signed permutation matrices). For  $n$ -generated subsemigroups of  $\mathbb{Q}^{d \times d}$ , the only known upper bounds appear to be double-exponential in  $d$ .

## REFERENCES

- [AB12] B. Adamczewski and J. P. Bell. On vanishing coefficients of algebraic power series over fields of positive characteristic. *Invent. Math.* 187.2 (2012), 343–393. DOI: [10.1007/s00222-011-0337-4](https://doi.org/10.1007/s00222-011-0337-4).
- [ABS23] B. Adamczewski, J. Bell, and D. Smertnig. A height gap theorem for coefficients of Mahler functions. *J. Eur. Math. Soc. (JEMS)* 25.7 (2023), 2525–2571. DOI: [10.4171/jems/1244](https://doi.org/10.4171/jems/1244).



- [ALP03] M. Abért, A. Lubotzky, and L. Pyber. Bounded generation and linear groups. *Internat. J. Algebra Comput.* 13.4 (2003), 401–413. DOI: [10.1142/S0218196703001493](https://doi.org/10.1142/S0218196703001493).
- [AP04] D. Arapura and C. Peterson. The common invariant subspace problem: an approach via Gröbner bases. *Linear Algebra Appl.* 384 (2004), 1–7. DOI: [10.1016/j.laa.2003.03.001](https://doi.org/10.1016/j.laa.2003.03.001).
- [AS03] J.-P. Allouche and J. Shallit. *Automatic sequences*. Theory, applications, generalizations. Cambridge University Press, Cambridge, 2003, pp. xvi+571. DOI: [10.1017/CB09780511546563](https://doi.org/10.1017/CB09780511546563).
- [Bau+91] G. Baumslag, F. B. Cannonito, D. J. Robinson, and D. Segal. The algorithmic theory of polycyclic-by-finite groups. *J. Algebra* 142.1 (1991), 118–149. DOI: [10.1016/0021-8693\(91\)90221-S](https://doi.org/10.1016/0021-8693(91)90221-S).
- [BC17] J. P. Bell and S. Chen. Power series with coefficients from a finite set. *J. Combin. Theory Ser. A* 151 (2017), 241–253. DOI: [10.1016/j.jcta.2017.05.002](https://doi.org/10.1016/j.jcta.2017.05.002).
- [BCH16] J. P. Bell, M. Coons, and K. G. Hare. Growth degree classification for finitely generated semigroups of integer matrices. *Semigroup Forum* 92.1 (2016), 23–44. DOI: [10.1007/s00233-015-9725-1](https://doi.org/10.1007/s00233-015-9725-1).
- [Bea01] R. Beals. Improved algorithms for the Tits alternative. *Groups and computation, III (Columbus, OH, 1999)*. Vol. 8. Ohio State Univ. Math. Res. Inst. Publ. de Gruyter, Berlin, 2001, 63–77.
- [Bea99] R. Beals. Algorithms for matrix groups and the Tits alternative. Vol. 58. 2. *36th IEEE Symposium on the Foundations of Computer Science (Milwaukee, WI, 1995)*. 1999, 260–279. DOI: [10.1006/jcss.1998.1614](https://doi.org/10.1006/jcss.1998.1614).
- [Bel+24] J. P. Bell, S. Chen, K. D. Nguyen, and U. Zannier. D-finiteness, rationality, and height III: multivariate Pólya-Carlson dichotomy. *Math. Z.* 306.4 (2024), Paper No. 70, 13. DOI: [10.1007/s00209-024-03470-3](https://doi.org/10.1007/s00209-024-03470-3).
- [Ben70] B. Benzaghou. Algèbres de Hadamard. *Bull. Soc. Math. France* 98 (1970), 209–252.
- [Ber05] J. Bernik. On groups and semigroups of matrices with spectra in a finitely generated field. *Linear Multilinear Algebra* 53.4 (2005), 259–267. DOI: [10.1080/03081080500094022](https://doi.org/10.1080/03081080500094022).
- [Ber07] J. Bernik. The eigenvalue field is a splitting field. *Arch. Math. (Basel)* 88.6 (2007), 481–490. DOI: [10.1007/s00013-007-1970-4](https://doi.org/10.1007/s00013-007-1970-4).
- [Béz86] J.-P. Bézivin. Sur un théorème de G. Pólya. *J. Reine Angew. Math.* 364 (1986), 60–68. DOI: [10.1515/crll.1986.364.60](https://doi.org/10.1515/crll.1986.364.60).
- [Béz87] J.-P. Bézivin. Suites récurrentes linéaires en caractéristique non nulle. *Bull. Soc. Math. France* 115.2 (1987), 227–239.
- [BG06] E. Bombieri and W. Gubler. *Heights in Diophantine geometry*. Vol. 4. New Mathematical Monographs. Cambridge University Press, Cambridge, 2006, pp. xvi+652. DOI: [10.1017/CB09780511542879](https://doi.org/10.1017/CB09780511542879).
- [BLR24] Y. I. Benalioua, N. Lhote, and P.-A. Reynier. Minimizing Cost Register Automata over a Field. *49th International Symposium on Mathematical Foundations of Computer Science, MFCS 2024, August 26 to 30, 2024, Bratislava, Slovakia*. LIPIcs. To appear. 2024. arXiv: [2307.13505](https://arxiv.org/abs/2307.13505).
- [BMR11] J. Bernik, M. Mastnak, and H. Radjavi. Positivity and matrix semigroups. *Linear Algebra Appl.* 434.3 (2011), 801–812. DOI: [10.1016/j.laa.2010.09.045](https://doi.org/10.1016/j.laa.2010.09.045).
- [BNZ20] J. P. Bell, K. D. Nguyen, and U. Zannier. D-finiteness, rationality, and height. *Trans. Amer. Math. Soc.* 373.7 (2020), 4889–4906. DOI: [10.1090/tran/8046](https://doi.org/10.1090/tran/8046).
- [BNZ23] J. P. Bell, K. D. Nguyen, and U. Zannier. D-finiteness, rationality, and height II: Lower bounds over a set of positive density. *Adv. Math.* 414 (2023), Paper No. 108859, 24. DOI: [10.1016/j.aim.2023.108859](https://doi.org/10.1016/j.aim.2023.108859).
- [Bor91] A. Borel. *Linear algebraic groups*. Second. Vol. 126. Graduate Texts in Mathematics. Springer-Verlag, New York, 1991, pp. xii+288. DOI: [10.1007/978-1-4612-0941-6](https://doi.org/10.1007/978-1-4612-0941-6).
- [BR11] J. Berstel and C. Reutenauer. *Noncommutative rational series with applications*. Vol. 137. Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge, 2011, pp. xiv+248.
- [BS21] J. Bell and D. Smertnig. Noncommutative rational Pólya series. *Selecta Math. (N.S.)* 27.3 (2021), Paper No. 34, 34. DOI: [10.1007/s00029-021-00629-2](https://doi.org/10.1007/s00029-021-00629-2).

- [BS23a] J. Bell and D. Smertnig.  $D$ -finite multivariate series with arithmetic restrictions on their coefficients. *Canad. J. Math.* 75.6 (2023), 1745–1779. DOI: [10.4153/s0008414x22000517](https://doi.org/10.4153/s0008414x22000517).
- [BS23b] J. P. Bell and D. Smertnig. Computing the linear hull: deciding deterministic? And unambiguous? For weighted automata over fields. *2023 38th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*. IEEE Comput. Soc. Press, Los Alamitos, CA, 2023, 13.
- [CD13] G. Cigler and R. Drnovšek. On semigroups of matrices with nonnegative diagonals. *Linear Algebra Appl.* 438.1 (2013), 626–633. DOI: [10.1016/j.laa.2012.08.035](https://doi.org/10.1016/j.laa.2012.08.035).
- [CLO15] D. A. Cox, J. Little, and D. O’Shea. *Ideals, varieties, and algorithms*. Fourth. Undergraduate Texts in Mathematics. An introduction to computational algebraic geometry and commutative algebra. Springer, Cham, 2015, pp. xvi+646. DOI: [10.1007/978-3-319-16721-3](https://doi.org/10.1007/978-3-319-16721-3).
- [Col15] T. Colcombet. Unambiguity in automata theory. *Descriptive complexity of formal systems*. Vol. 9118. Lecture Notes in Comput. Sci. Springer, Cham, 2015, 3–18. DOI: [10.1007/978-3-319-19225-3\\_1](https://doi.org/10.1007/978-3-319-19225-3_1).
- [Cor+22a] P. Corvaja, J. L. Demeio, A. S. Rapinchuk, J. Ren, and U. M. Zannier. Bounded generation by semi-simple elements: quantitative results. *C. R. Math. Acad. Sci. Paris* 360 (2022), 1249–1255. DOI: [10.5802/crmath.376](https://doi.org/10.5802/crmath.376).
- [Cor+22b] P. Corvaja, A. S. Rapinchuk, J. Ren, and U. M. Zannier. Non-virtually abelian anisotropic linear groups are not boundedly generated. *Invent. Math.* 227.1 (2022), 1–26. DOI: [10.1007/s00222-021-01064-y](https://doi.org/10.1007/s00222-021-01064-y).
- [Cor+23] P. Corvaja, J. Demeio, A. Rapinchuk, J. Ren, and U. Zannier. Purely exponential parametrizations and their group-theoretic applications (2023). *Preprint*. arXiv: [2308.14013](https://arxiv.org/abs/2308.14013).
- [Cor24] P. Corvaja. *E-Mail communication from 2024-10-12 with J. L. Demeio, A. Puch, A. S. Rapinchuk, J. Ren, D. Smertnig, U. M. Zannier*. 2024.
- [CP17] T. Colcombet and D. Petrişan. Automata in the category of glued vector spaces. *42nd International Symposium on Mathematical Foundations of Computer Science*. Vol. 83. LIPIcs. Leibniz Int. Proc. Inform. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2017, Art. No. 52, 14.
- [CP20] T. Colcombet and D. Petrişan. Automata minimization: a functorial approach. *Log. Methods Comput. Sci.* 16.1 (2020), Paper No. 32, 28. DOI: [10.23638/LMCS-16\(1:32\)2020](https://doi.org/10.23638/LMCS-16(1:32)2020).
- [CR90] C. W. Curtis and I. Reiner. *Methods of representation theory. Vol. I*. Wiley Classics Library. With applications to finite groups and orders, Reprint of the 1981 original, A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1990, pp. xxiv+819.
- [Dav+17] L. Daviaud, I. Jecker, P.-A. Reynier, and D. Villevalois. Degree of sequentiality of weighted automata. *Foundations of software science and computation structures*. Vol. 10203. Lecture Notes in Comput. Sci. Springer, Berlin, 2017, 215–230. DOI: [10.1007/978-3-662-54458-7\\_13](https://doi.org/10.1007/978-3-662-54458-7_13).
- [DFO11] A. S. Detinko, D. L. Flannery, and E. A. O’Brien. Algorithms for the Tits alternative and related problems. *J. Algebra* 344 (2011), 397–406. DOI: [10.1016/j.jalgebra.2011.06.036](https://doi.org/10.1016/j.jalgebra.2011.06.036).
- [DJK05a] H. Derksen, E. Jeandel, and P. Koiran. Quantum automata and algebraic groups. *J. Symbolic Comput.* 39.3-4 (2005), 357–371. DOI: [10.1016/j.jsc.2004.11.008](https://doi.org/10.1016/j.jsc.2004.11.008).
- [DJK05b] H. Derksen, E. Jeandel, and P. Koiran. Quantum automata and algebraic groups. *J. Symbolic Comput.* 39.3-4 (2005), 357–371. DOI: [10.1016/j.jsc.2004.11.008](https://doi.org/10.1016/j.jsc.2004.11.008).
- [DKV09] M. Droste, W. Kuich, and H. Vogler, eds. *Handbook of weighted automata*. Monographs in Theoretical Computer Science. An EATCS Series. Springer-Verlag, Berlin, 2009, pp. xviii+608. DOI: [10.1007/978-3-642-01492-5](https://doi.org/10.1007/978-3-642-01492-5).
- [DM12] H. Derksen and D. Masser. Linear equations over multiplicative groups, recurrences, and mixing I. *Proc. Lond. Math. Soc. (3)* 104.5 (2012), 1045–1083. DOI: [10.1112/plms/pdr040](https://doi.org/10.1112/plms/pdr040).
- [EG15] J.-H. Evertse and K. Györy. *Unit equations in Diophantine number theory*. Vol. 146. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2015, pp. xv+363. DOI: [10.1017/CBO9781316160749](https://doi.org/10.1017/CBO9781316160749).
- [Eve84] J.-H. Evertse. On sums of  $S$ -units and linear recurrences. *Compositio Math.* 53.2 (1984), 225–244.

- [Gru+12] L. Grunenfelder, T. Košir, M. Omladič, and H. Radjavi. Finite groups with submultiplicative spectra. *J. Pure Appl. Algebra* 216.5 (2012), 1196–1206. DOI: [10.1016/j.jpaa.2011.10.004](https://doi.org/10.1016/j.jpaa.2011.10.004).
- [Hru+18] E. Hrushovski, J. Ouaknine, A. Pouly, and J. Worrell. Polynomial invariants for affine programs. *LICS '18—33rd Annual ACM/IEEE Symposium on Logic in Computer Science*. ACM, New York, 2018, 10. DOI: [10.1145/3209108.3209142](https://doi.org/10.1145/3209108.3209142).
- [Hru+23] E. Hrushovski, J. Ouaknine, A. Pouly, and J. Worrell. On strongest algebraic program invariants. *J. ACM* 70.5 (2023), Art. 29, 22. DOI: [10.1145/3614319](https://doi.org/10.1145/3614319).
- [IK22] M. S. Im and M. Khovanov. *Topological theories and automata*. Preprint. 2022. arXiv: [2202.13398](https://arxiv.org/abs/2202.13398) [[math.QA](https://arxiv.org/archive/math)].
- [IK24] M. S. Im and M. Khovanov. One-dimensional topological theories with defects: the linear case. *Algebraic and topological aspects of representation theory*. Vol. 791. Contemp. Math. Amer. Math. Soc., [Providence], RI, 2024, 105–146. DOI: [10.1090/comm/791/15874](https://doi.org/10.1090/comm/791/15874).
- [JMP24] I. Jecker, F. Mazowiecki, and D. Purser. Determinisation and Unambiguisation of Polynomially-Ambiguous Rational Weighted Automata. *Proceedings of the 39th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2024, Tallinn, Estonia, July 8-11, 2024*. Ed. by P. Sobocinski, U. D. Lago, and J. Esparza. ACM, 2024, 46:1–46:13. DOI: [10.1145/3661814.3662073](https://doi.org/10.1145/3661814.3662073).
- [Kir12] D. Kirsten. Decidability, undecidability, and PSPACE-completeness of the twins property in the tropical semiring. *Theoret. Comput. Sci.* 420 (2012), 56–63. DOI: [10.1016/j.tcs.2011.11.006](https://doi.org/10.1016/j.tcs.2011.11.006).
- [KL09] D. Kirsten and S. Lombardy. Deciding unambiguity and sequentiality of polynomially ambiguous min-plus automata. *STACS 2009: 26th International Symposium on Theoretical Aspects of Computer Science*. Vol. 3. LIPIcs. Leibniz Int. Proc. Inform. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2009, 589–600.
- [Kos22] P. Kostolányi. Determinisability of unary weighted automata over the rational numbers. *Theoret. Comput. Sci.* 898 (2022), 110–131. DOI: [10.1016/j.tcs.2021.11.002](https://doi.org/10.1016/j.tcs.2021.11.002).
- [Kos24] P. Kostolányi. Finitely ambiguous and finitely sequential weighted automata over fields. *Theoret. Comput. Sci.* 1012 (2024), Paper No. 114725. DOI: [10.1016/j.tcs.2024.114725](https://doi.org/10.1016/j.tcs.2024.114725).
- [Kra05] M. Kramar. The structure of irreducible matrix groups with submultiplicative spectrum. *Linear Multilinear Algebra* 53.1 (2005), 13–25. DOI: [10.1080/03081080410001709765](https://doi.org/10.1080/03081080410001709765).
- [Lam91] T. Y. Lam. *A first course in noncommutative rings*. Vol. 131. Graduate Texts in Mathematics. Springer-Verlag, New York, 1991, pp. xvi+397. DOI: [10.1007/978-1-4684-0406-7](https://doi.org/10.1007/978-1-4684-0406-7).
- [LS06] S. Lombardy and J. Sakarovitch. Sequential? *Theoret. Comput. Sci.* 356.1-2 (2006), 224–244. DOI: [10.1016/j.tcs.2006.01.028](https://doi.org/10.1016/j.tcs.2006.01.028).
- [MR17] M. Mohri and M. D. Riley. A disambiguation algorithm for weighted automata. *Theoret. Comput. Sci.* 679 (2017), 53–68. DOI: [10.1016/j.tcs.2016.08.019](https://doi.org/10.1016/j.tcs.2016.08.019).
- [Neu54] B. H. Neumann. Groups covered by permutable subsets. *J. London Math. Soc.* 29 (1954), 236–248. DOI: [10.1112/jlms/s1-29.2.236](https://doi.org/10.1112/jlms/s1-29.2.236).
- [Nos+22] K. Nosan, A. Pouly, S. Schmitz, M. Shirmohammadi, and J. Worrell. On the computation of the Zariski closure of finitely generated groups of matrices. *ISSAC '22—Proceedings of the 2022 International Symposium on Symbolic and Algebraic Computation*. ACM, New York, 2022, 129–138.
- [Okn98] J. Okniński. *Semigroups of matrices*. Vol. 6. Series in Algebra. World Scientific Publishing Co., Inc., River Edge, NJ, 1998, pp. xiv+311. DOI: [10.1142/9789812816290](https://doi.org/10.1142/9789812816290).
- [Ost99] G. Ostheimer. Practical algorithms for polycyclic matrix groups. *J. Symbolic Comput.* 28.3 (1999), 361–379. DOI: [10.1006/jsco.1999.0287](https://doi.org/10.1006/jsco.1999.0287).
- [Pól21] G. Pólya. Arithmetische Eigenschaften der Reihenentwicklungen rationaler Funktionen. *J. Reine Angew. Math.* 151 (1921), 1–31. DOI: [10.1515/crll.1921.151.1](https://doi.org/10.1515/crll.1921.151.1).
- [PR03] G. Prasad and A. S. Rapinchuk. Existence of irreducible  $\mathbb{R}$ -regular elements in Zariski-dense subgroups. *Math. Res. Lett.* 10.1 (2003), 21–32. DOI: [10.4310/MRL.2003.v10.n1.a3](https://doi.org/10.4310/MRL.2003.v10.n1.a3).

- [PR14] G. Prasad and A. S. Rapinchuk. Generic elements in Zariski-dense subgroups and isospectral locally symmetric spaces. *Thin groups and superstrong approximation*. Vol. 61. Math. Sci. Res. Inst. Publ. Cambridge Univ. Press, Cambridge, 2014, 211–252.
- [PR17] G. Prasad and A. S. Rapinchuk. Generic elements of a Zariski-dense subgroup form an open subset. *Trans. Moscow Math. Soc.* 78 (2017), 299–314. DOI: [10.1090/mosc/274](https://doi.org/10.1090/mosc/274).
- [PS82] A. J. van der Poorten and H. P. Schlickewei. The growth condition for recurrence sequences. *Macquarie University Math. Rep.* 82–0041 (1982).
- [PS91] A. J. van der Poorten and H. P. Schlickewei. Additive relations in fields. *J. Austral. Math. Soc. Ser. A* 51.1 (1991), 154–170.
- [PS96] A. J. van der Poorten and I. E. Shparlinski. On linear recurrence sequences with polynomial coefficients. *Glasgow Math. J.* 38.2 (1996), 147–155. DOI: [10.1017/S0017089500031372](https://doi.org/10.1017/S0017089500031372).
- [Put88] M. S. Putcha. *Linear algebraic monoids*. Vol. 133. London Mathematical Society Lecture Note Series. Cambridge University Press, Cambridge, 1988, pp. x+171. DOI: [10.1017/CB09780511600661](https://doi.org/10.1017/CB09780511600661).
- [Rad00] H. Radjavi. Sublinearity and other spectral conditions on a semigroup. *Canad. J. Math.* 52.1 (2000), 197–224. DOI: [10.4153/CJM-2000-009-5](https://doi.org/10.4153/CJM-2000-009-5).
- [Ren05] L. E. Renner. *Linear algebraic monoids*. Vol. 134. Encyclopaedia of Mathematical Sciences. Invariant Theory and Algebraic Transformation Groups, V. Springer-Verlag, Berlin, 2005, pp. xii+246.
- [Reu79] C. Reutenauer. On Pólya series in noncommuting variables. *Fundamentals of computation theory (Proc. Conf. Algebraic, Arith. and Categorical Methods in Comput. Theory, Berlin/Wendisch-Rietz, 1979)*. Vol. 2. Math. Res. Akademie-Verlag, Berlin, 1979, 391–396.
- [Rob96] D. J. S. Robinson. *A course in the theory of groups*. Second. Vol. 80. Graduate Texts in Mathematics. Springer-Verlag, New York, 1996, pp. xviii+499. DOI: [10.1007/978-1-4419-8594-1](https://doi.org/10.1007/978-1-4419-8594-1).
- [RY19] H. Radjavi and B. R. Yahaghi. Burnside’s theorem in the setting of general fields. *Arch. Math. (Basel)* 113.1 (2019), 53–57. DOI: [10.1007/s00013-019-01301-0](https://doi.org/10.1007/s00013-019-01301-0).
- [Sak09] J. Sakarovitch. *Elements of automata theory*. Translated by Reuben Thomas. Cambridge University Press, Cambridge, 2009, pp. xxiv+758. DOI: [10.1017/CB09781139195218](https://doi.org/10.1017/CB09781139195218).
- [Sch07] P. Schauenburg. A Gröbner-based treatment of elimination theory for affine varieties. *J. Symbolic Comput.* 42.9 (2007), 859–870. DOI: [10.1016/j.jsc.2007.06.003](https://doi.org/10.1016/j.jsc.2007.06.003).
- [Ste12] B. Steinberg. Yet another solution to the Burnside problem for matrix semigroups. *Canad. Math. Bull.* 55.1 (2012), 188–192. DOI: [10.4153/CMB-2011-030-1](https://doi.org/10.4153/CMB-2011-030-1).
- [Tao14] T. Tao. *Hilbert’s fifth problem and related topics*. Vol. 153. Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2014, pp. xiv+338. DOI: [10.1090/gsm/153](https://doi.org/10.1090/gsm/153).
- [Weh73] B. A. F. Wehrfritz. *Infinite linear groups. An account of the group-theoretic properties of infinite groups of matrices*. Vol. Band 76. Ergebnisse der Mathematik und ihrer Grenzgebiete [Results in Mathematics and Related Areas]. Springer-Verlag, New York-Heidelberg, 1973, pp. xiv+229.
- [WS91] A. Weber and H. Seidl. On the degree of ambiguity of finite automata. *Theoret. Comput. Sci.* 88.2 (1991), 325–349. DOI: [10.1016/0304-3975\(91\)90381-B](https://doi.org/10.1016/0304-3975(91)90381-B).

FACULTY OF MATHEMATICS, INFORMATICS AND MECHANICS (MIMUW), UNIVERSITY OF WARSAW,  
 BANACHA 2, 02-097 WARSAW, POLAND  
*Email address:* [a.puch@student.uw.edu.pl](mailto:a.puch@student.uw.edu.pl)

FACULTY OF MATHEMATICS AND PHYSICS (FMF), UNIVERSITY OF LJUBLJANA AND INSTITUTE OF  
 MATHEMATICS, PHYSICS AND MECHANICS (IMFM), JADRANSKA ULICA 21, 1000 LJUBLJANA, SLOVENIA  
*Email address:* [daniel.smertnig@fmf.uni-lj.si](mailto:daniel.smertnig@fmf.uni-lj.si)