Distinguishing Elements in Semigroups

Markus Lohrey
Universität Siegen, Germany

Alexander Thumm
Universität Siegen, Germany

Julio Xochitemol
Universität Siegen, Germany

Abstract

We investigate randomized streaming algorithms for word problems in finitely generated semigroups. For this we use the notion of a distinguisher: a randomized streaming algorithm that processes two input words in parallel and, with high probability, reaches identical memory states if the words represent the same element, and distinct states otherwise. We construct such distinguishers with space complexity $\mathcal{O}(\log \log n)$ for finitely generated commutative semigroups. Moreover, we show a transfer result for semilattice decompositions that allows to construct a distinguisher for a finitely generated semigroup from distinguishers for the components of its semilattice decomposition. Thereby the space complexity and the error probability of the distinguisher increase only by a constant factor. We use this result to obtain distinguishers with space complexity $\mathcal{O}(\log n)$ for free Clifford semigroups and distinguishers with space complexity $\mathcal{O}(\log \log n)$ for finitely generated regular nilpotent semigroups. We complement these upper bounds with lower bounds demonstrating that certain well-known semigroups do not admit distinguishers with sublinear space complexity. This includes, for example, free inverse monoids of rank at least two and polycyclic semigroups.

2012 ACM Subject Classification Theory of computation → Streaming models

Keywords and phrases Streaming algorithms, semigroups, word problem, space complexity

Digital Object Identifier 10.4230/LIPIcs.MFCS.2026.30

Related Version *Full Version*: `https://arxiv.org/abs/2202.04060` [11]

1 Introduction

The word problem for finitely generated (f.g.) groups and, more generally, f.g. semigroups is a central algorithmic decision problem for infinite algebraic structures. In the semigroup setting, it asks whether two words over a fixed generating set represent the same element, whereas for groups it suffices to test whether a single word represents the identity. A vast body of work has established decidability and complexity results for the word problem across a wide range of algebraic classes [5, 9, 14].

In recent work [10, 12], randomized streaming algorithms for word problems in groups were introduced. In this model, the input word is processed sequentially, without random access, and the algorithm must decide – with high probability – whether the input evaluates to the identity. The primary complexity measure is space, in line with the standard streaming literature. It was shown that for many classes of groups, randomized streaming algorithms with space complexity $\mathcal{O}(\log n)$ or even $\mathcal{O}(\log \log n)$ exist. A key technical tool in these results is a stronger model of streaming algorithms for groups called *distinguishers*.

An ϵ-*distinguisher*, where $0 \leq \epsilon < 1/2$, for a group G with finite generating set Σ is a randomized streaming algorithm $\mathcal{R}$ such that for all words $u, v \in \Sigma^*$ of length at most n the following holds: (i) if u and v represent the same element of G, then with probability at least $1 - \epsilon$ they lead to the same memory state of $\mathcal{R}$; and (ii) if they represent different elements, then with probability at least $1 - \epsilon$ they lead to different memory states. Formally,

such a distinguisher is given in a non-uniform way by a sequence $(\mathcal{A}_n)_{n\geq 0}$ of probabilistic finite automata, where $\mathcal{A}_n$ satisfies (i) and (ii) for all inputs of length at most n. The space complexity is then simply defined as the mapping $n \mapsto \log_2(\text{number of states of } \mathcal{A}_n)$, that is, the number of bits required to encode the states of $\mathcal{A}_n$.

Every ϵ-distinguisher for a group G can be transformed into a (non-uniform) randomized streaming algorithm for testing whether a given input stream evaluates to the group identity of G (in other words, it solves the word problem for G). The error probability of the latter algorithm is still bounded by ϵ and its space complexity is at most twice the space complexity of the ϵ-distinguisher for G [10].

The following results were established in [10, 12]; see also [11]:

(a) Every f.g. linear group admits an ϵ-distinguisher with space complexity $\mathcal{O}(\log n)$. Here, the error probability ϵ can be chosen as any inverse polynomial in n, i.e., $\epsilon(n) = 1/n^c$.

(b) Every f.g. nilpotent group admits an ϵ-distinguisher with space complexity $\mathcal{O}(\log\log n)$, where ϵ can be chosen as any inverse polylogarithmic function in n, i.e., $\epsilon(n) = 1/(\log n)^c$.

(c) For every f.g. abelian group A and every f.g. group G with an ϵ-distinguisher of space complexity $s(n)$, the corresponding wreath product $A \wr G$ admits an ϵ'-distinguisher with error probability $\epsilon'(n) = \Theta(\epsilon(n) \cdot n^2)$ and space complexity $s'(n) = \mathcal{O}(s(n) + \log n)$.

(d) For all f.g. groups $G_1, \ldots, G_k$ with ϵ-distinguishers of space complexity $s(n)$, all graph products of these groups (in particular, free and direct products) admit an ϵ'-distinguisher with error probability $\epsilon'(n) = \Theta(\epsilon(n) \cdot n^2)$ and space complexity $s'(n) = \mathcal{O}(s(n) + \log n)$.

A notable advantage of the distinguisher model is its direct applicability beyond groups. In fact, distinguishers serve as a natural notion of a streaming algorithm for the word problem of semigroups.[1] More generally, ϵ-distinguishers can be defined for any infinite deterministic automaton $\mathcal{A}$, where a collection of finite probabilistic automata $(\mathcal{A}_n)_{n\geq 0}$ provides an approximation that preserves state equality and separation up to error ϵ. While we plan to investigate this aspect in future work, the present paper focuses on the space complexity of distinguishers for semigroups.

Our results. The first main result of this paper is Theorem 4.2. It shows that every f.g. commutative semigroup admits an ϵ-distinguisher with space complexity $\mathcal{O}(\log\log n)$, where the error probability ϵ can be chosen as any inverse polylogarithmic function in n. Moreover, this bound is asymptotically optimal: a simple communication complexity argument shows that an ϵ-distinguisher for any f.g. infinite semigroup for some constant $\epsilon < 1/2$ must have space complexity $\Omega(\log\log n)$; see the end of Section 3.

The second main result, which is Theorem 5.4, concerns semilattice decompositions. Such a decomposition partitions a semigroup S into pairwise disjoint subsemigroups S_α with a semilattice structure on the set of these subsemigroups. This is a central decomposition technique in semigroup theory; see e.g. [3]. We show that if every constituent subsemigroup S_α is a monoid and has an $\epsilon(n)$-distinguisher with space complexity $s(n)$, then the semigroup S has a $c \cdot \epsilon(dn)$-distinguisher with space complexity $s'(n) = c \cdot s(dn)$ for some constants c, d.

We present two applications of Theorem 5.4: Every f.g. regular nilpotent semigroup has an ϵ-distinguisher with space complexity $\mathcal{O}(\log\log n)$ (Corollary 5.6), and every f.g. free Clifford semigroup has an ϵ-distinguisher with space complexity $\mathcal{O}(\log n)$ (Corollary 5.5).

In the final section of this paper, we establish lower bounds on the space complexity of distinguishers for inverse monoids, which form an important class of monoids situated between monoids and groups. Specifically, we show that every ϵ-distinguisher for a free

[1] The word problem of a semigroup S with the generating set Σ is usually defined as the set of pairs (u, v) where u and v are non-empty words over the alphabet Σ that evaluate to the same element in S.

inverse monoid on at least two generators for a constant $\epsilon < 1/2$ must have space complexity $\Omega(n)$ (Theorem 6.1). Note that this is the worst possible, as every f.g. monoid has a 0-distinguisher (i.e., a deterministic distinguisher) of space complexity $\mathcal{O}(n)$. In contrast, for a single generator the optimal space complexity is $\Theta(\log n)$, which is also achieved by a deterministic distinguisher (Theorem 6.2). Analogous bounds hold for polycyclic monoids (Theorems 6.3 and 6.4), which are closely connected to context-free languages [21].

Missing proofs can be found in the full version [11].

Further results. Let us briefly summarize some additional results from the full version [11] that extend the above-mentioned results for groups but are not treated here in detail.

- The result in (a) for f.g. linear groups generalizes directly to f.g. linear semigroups, i.e., semigroups of $(d \times d)$-matrices over a field. For every constant c, every such semigroup admits an ϵ-distinguisher with error probability $\epsilon(n) = 1/n^c$ and space complexity $\mathcal{O}(\log n)$.
- The wreath product construction in (c) extends to the case where A is a f.g. cancellative commutative monoid (hence, embeds into a f.g. abelian group). In contrast, a sharp lower bound holds beyond this case: Suppose that G is a f.g. infinite group and M is a f.g. monoid that is non-cancellative or non-commutative. Then every ϵ-distinguisher for the wreath product $M \wr G$ and some constant $\epsilon < 1/2$ has space complexity $\Omega(n)$.
- The upper bound for graph products in (d) extends from groups to left-cancellative monoids. It also holds for semigroup-theoretic graph products or, equivalently, for graph products of monoids $M_1, \ldots, M_k$ where the non-neutral elements form an ideal in each M_i. This is complemented by a lower bound: Suppose that M is a f.g. monoid in which $M \setminus \{1\}$ is not an ideal, and let B be the bicyclic monoid. Then every ϵ-distinguisher for the free product $M * B$ and some constant $\epsilon < 1/2$ has space complexity $\Omega(n)$.

2 Preliminaries

We use standard notation for words. For a finite alphabet Σ, we write Σ^* for the set of all words over Σ. The empty word is denoted by ε, and the set of non-empty words by $\Sigma^+ = \Sigma^* \setminus \{\epsilon\}$. The length of a word w is denoted by $|w|$. We define $\Sigma^{\leq n} := \{w \in \Sigma^* \mid |w| \leq n\}$.

For integers $a \leq b$ we write $[a, b]$ to denote the corresponding interval $\{a, a+1, \ldots, b-1, b\}$.

2.1 Semigroups

We assume that the reader has some basic knowledge in semigroup theory. In this section we briefly recall some elementary definitions and fix our notation. For further details, as well as any terms left undefined, the reader may consult e.g. the monographs [2, 3].

Let S be a *semigroup*; that is, a set equipped with an associative binary operation. An element $e \in S$ is a *neutral* (resp. *zero*) element of S, and usually denoted by 1 (resp. 0), if $ea = a = ae$ (resp. $ea = e = ae$) holds for all $a \in S$. Both are examples of *idempotents*, that is, elements e with $e^2 = e$. A semigroup with a neutral element is a *monoid*. Every semigroup S has a smallest monoid containing it, which we denote by S^1. If S is a monoid, then $S = S^1$, otherwise $S^1 = S \uplus \{1\}$, where $s1 = 1s$ for all $s \in S$.

A *subsemigroup* of S is a subset $T \subseteq S$ closed under multiplication, that is, with $T^2 \subseteq T$. An *ideal* of S is a subset $I \subseteq S$ with $S^1 I S^1 \subseteq I$. Every subset $\Sigma \subseteq S$ is contained in a smallest subsemigroup $\langle \Sigma \rangle \subseteq S$. If $S = \langle \Sigma \rangle$, then Σ is a *generating set* of S. The semigroup S is said to be *finitely generated*, or *f.g.* for short, if it has a finite generating set.

Let $\Sigma \subseteq S$. Every $w \in \Sigma^+$ evaluates to an element in S called the *S-value* of w. Conversely, if $\Sigma \subseteq S$ is a generating set, then every element of S can be represented by a word $w \in \Sigma^+$ in this way. For $u, v \in \Sigma^+$, we write $u \equiv_S v$ if u and v have the same S-value. We frequently identify a semigroup homomorphism $h\colon S \to T$ with a homomorphism $\tilde{h}\colon \Sigma^+ \to \Gamma^+$ such that $S = \langle \Sigma \rangle$, $T = \langle \Gamma \rangle$ and for every $a \in \Sigma$, the T-value of $\tilde{h}(a) \in \Gamma^+$ is $h(a) \in T$.

The notational conventions discussed above apply, *mutatis mutandis*, to monoids as well. For this, the set of non-empty words Σ^+ has to be replaced by Σ^*.

Commutative semigroups. A semigroup S is *commutative* if $ab = ba$ holds for all $a, b \in S$. In a commutative semigroup S, Green's $\mathcal{H}$-, $\mathcal{L}$- and $\mathcal{R}$-preorders coincide and can be defined by $a \leq b$ if and only if $a = bc$ for some $c \in S^1$. The corresponding equivalence relation $\mathcal{H}$, with $a \mathcal{H} b$ if and only if $a \leq b \leq a$, is thus a congruence on S. The quotient $S/\mathcal{H}$ is partially ordered by $\leq$, hence *group-free* (that is, it contains no nontrivial group as a subsemigroup).

Later in this article, we will encounter two special classes of commutative semigroups as such quotients: *commutative nilsemigroups* and *commutative bands*, also known as *semilattices*. In general, a *nilsemigroup* is a semigroup X with zero where every $\alpha \in X$ has a zero power, that is, $\alpha^n = 0$ for some $n \geq 1$ and a *band* is a semigroup Y where every $\alpha \in Y$ is idempotent. Importantly, every f.g. commutative nilsemigroup and every f.g. semilattice is finite.

Nilpotent semigroups. A group is *nilpotent* if its lower central series reaches the trivial group in finitely many steps. By expressing nilpotency for groups via semigroup identities, this concept was generalized to semigroups by Mal'cev [13] and by Neumann and Taylor [17]. Their two sets of identities differ slightly and, in this article, we adopt the latter. Specifically, let X_i and Y_i be the words over formal variables $x, y, z_2, \ldots, z_i$ defined inductively by

$$X_1 = xy,\ Y_1 = yx \quad \text{and} \quad X_i = X_{i-1} z_i Y_{i-1},\ Y_i = Y_{i-1} z_i X_{i-1}\ (i \geq 2).$$

We say that a semigroup S is *i-step nilpotent* if it satisfies $X_i \approx Y_i$ identically (that is, for all values of S substituted into the variables), and *nilpotent* if it is i-step nilpotent for some i.

Be aware that a very different notion of nilpotent semigroup, originating in ring theory, is frequently encountered in the literature. However, this notion will not be used here.

Regular and inverse semigroups. Consider a semigroup S. An element $a \in S$ is *regular* if there is an element $b \in S$ such that $aba = a$. If $aba = a$ and $bab = b$ both hold, then b is called an *inverse* of a and vice versa. A *regular semigroup* is a semigroup where every element is regular or, equivalently, where every element has at least one inverse.

An *inverse semigroup* is a regular semigroup with unique inverses or, equivalently, with commuting idempotents [19, Thm II.1.2]. In such a semigroup, we write a^{-1} for the unique inverse of an element a. Be aware that $aa^{-1} = 1$ and $a^{-1}a = 1$ need not hold; in fact, there need not even be a neutral element. By the Vagner-Preston representation theorem [24, 20], a semigroup is inverse if and only if it is isomorphic to an inverse-closed semigroup of partial bijections on some set (where multiplication is the ordinary composition of partial mappings). Therefore, inverse semigroups can be used to model partial symmetries.

2.2 Automata

Throughout this work, we will only consider automata without final states. Hence, we define a *deterministic automaton* as a tuple $\mathcal{A} = (Q, \Sigma, q_0, \delta)$, where Q is a (possibly infinite) set of states, Σ is a finite alphabet, $q_0 \in Q$ is an initial state, and $\delta\colon Q \times \Sigma \to Q$ is a transition function. The transition function $\delta\colon Q \times \Sigma \to Q$ is extended to a mapping $\delta\colon Q \times \Sigma^* \to Q$ in the usual way; that is, $\delta(q, \varepsilon) = q$ and $\delta(q, wa) = \delta(\delta(q, w), a)$ for all $w \in \Sigma^*$ and $a \in \Sigma$.

The *growth function* $\gamma_{\mathcal{A}} \colon \mathbb{N} \to \mathbb{N}$ of $\mathcal{A}$ is defined by

$$\gamma_{\mathcal{A}}(n) = |\{\delta(q_0, w) \in Q \mid w \in \Sigma^{\leq n}\}|.$$

Thus, $\gamma_{\mathcal{A}}(n)$ is the number of states reachable from the initial state in at most n steps.

Finite automata. A *deterministic finite automaton*, or DFA for short, is a deterministic automaton $\mathcal{A} = (Q, \Sigma, q_0, \delta)$ with finitely many states. A *semi-probabilistic finite automaton*, or semiPFA, is defined analogously as a tuple $\mathcal{A} = (Q, \Sigma, \iota, \delta)$ where the single initial state is replaced by a probability distribution $\iota \colon Q \to \{p \in \mathbb{R} \mid 0 \leq p \leq 1\}$. Thus, the semiPFA $\mathcal{A}$ can also be viewed as a probability distribution on DFAs, where (Q, Σ, q_0, δ) has probability $\iota(q_0)$.[2] We write $q \sim \iota$ if the state $q \in Q$ is randomly chosen according to the distribution ι.

Cayley graphs. Let M be a monoid with finite generating set $\Sigma \subseteq M$. We can then construct a deterministic automaton $\mathcal{A}_{M,\Sigma} = (M, \Sigma, 1, \delta)$ where the transition function $\delta \colon M \times \Sigma \to M$ is given by multiplication in M; that is, $\delta(x, a) = xa$ for all $x \in M$ and $a \in \Sigma$. Viewed as a directed edge-labeled graph (and forgetting about the special role of $1 \in M$), this automaton is known as the *Cayley graph* of M with respect to the generating set Σ.

Using this construction, we can transfer concepts and results from deterministic automata to f.g. monoids (and semigroups). For instance, the *growth function* of M with respect to Σ, which we denote by $\gamma_{M,\Sigma} \colon \mathbb{N} \to \mathbb{N}$, is the growth function of the corresponding Cayley graph.

2.3 Communication complexity

Our lower bounds on the space complexity of distinguishers are based on randomized communication complexity. In this section, we recall the necessary background; see [7] for a comprehensive introduction.

Let $f \colon X \times Y \to \{0, 1\}$ be a function, where X and Y are finite sets. A *randomized (communication) protocol* P for f involves two parties, Alice and Bob. Alice receives an input $x \in X$ and a random choice $r \in R$, while Bob receives an input $y \in Y$ and a random choice $s \in S$. Here, R and S are finite sets equipped with probability distributions ι and ζ, respectively. The goal of Alice and Bob is to compute $f(x, y)$. To this end, they communicate in a finite number of rounds. In each round, Alice sends a bit string to Bob or vice versa, and the protocol P determines which communication direction is chosen. A *one-way protocol* consists of a single communication round in which Alice sends a bit string to Bob. At the end of the communication, Bob outputs a single bit $P(x, r, y, s)$.

We say that the protocol P *computes* the function f if, for every $(x, y) \in X \times Y$,

$$\operatorname*{Prob}_{r \sim \iota, s \sim \zeta}[P(x, r, y, s) \neq f(x, y)] \leq \frac{1}{3}. \tag{1}$$

The cost of the protocol P is the maximum number of transmitted bits, where the maximum is taken over all $(x, r, y, s) \in X \times R \times Y \times S$. The *randomized (one-way) communication complexity* of f is the minimum cost among all (one-way) randomized protocols computing f. No restriction is imposed on the sizes of the sets R and S.

[2] In general probabilistic finite automata (PFA), also the transition that is taken in each step is randomly chosen. For our purpose, semiPFA are sufficient.

The constant 1/3 in (1) is arbitrary: replacing it by any constant $\epsilon < 1/2$ changes the communication complexity only by a constant factor depending on ϵ; see [7, p. 30]. We only consider the private version of randomized communication protocols, where Alice and Bob make private and independent random choices from the sets R and S, respectively, and (in contrast to the public version) their choices are not known to the other party.

We will use the following communication problems for our lower bounds. Their randomized one-way communication complexities are summarized in Theorem 2.1 below.

- GT_n (greater than): The inputs for Alice and Bob are numbers $x, y \in [1, n]$, respectively. In this problem, Bob should output 1 if and only if $x < y$.
- IDX_n (index): Alice receives a bit string $x \in \{0,1\}^n$, and Bob receives an index $i \in [1, n]$. In this problem, Bob should output 1 if and only if $x[i] = 1$.

We will also use the following variant of the index problem. It is a *promise communication problem*, that is, a problem defined by a function $f\colon A \to \{0,1\}$, where $A \subseteq X \times Y$. In this setting, condition (1) is only required to hold for inputs $(x, y) \in A$.

- augIDX_n (augmented index): Alice receives a bit string $x \in \{0,1\}^n$, while Bob receives an index $i \in [1, n]$ and a bit string $y \in \{0,1\}^{n-i}$ with the promise that y is a suffix of x. In this problem, Bob should output 1 if and only if $x[i] = 1$.

▶ **Theorem 2.1.** *The following hold.*

- *The randomized one-way communication complexity of* GT_n *is* $\Theta(\log n)$. [15, Thm. 19]
- *The randomized one-way communication complexity of* IDX_n *is* $\Theta(n)$. [6, Thm. 3.7]
- *The randomized one-way communication complexity of* augIDX_n *is* $\Theta(n)$. [1, Thm. 5.1]

3 Distinguishers

The following notion, introduced under the name "ϵ-injective streaming algorithm" in [10], serves as a convenient abstraction of randomized streaming algorithms for word problems.

Let $\mathcal{A} = (Q, \Sigma, q_0, \delta)$ be a (possibly infinite) deterministic automaton. Further, let $\epsilon_0, \epsilon_1 : \mathbb{N} \to \{p \in \mathbb{R} \mid 0 \le p \le 1\}$ be monotonically decreasing functions. An (ϵ_0, ϵ_1)-*distinguisher* for $\mathcal{A}$ is a collection of semiPFAs $\mathcal{R} = (\mathcal{A}_n)_{n \ge 0}$ with $\mathcal{A}_n = (Q_n, \Sigma, \iota_n, \delta_n)$ such that for all sufficiently large n and all words $u, v \in \Sigma^{\le n}$ the following properties hold.

- If $\delta(q_0, u) = \delta(q_0, v)$, then $\mathsf{Prob}_{q \sim \iota_n}[\delta_n(q, u) = \delta_n(q, v)] \ge 1 - \epsilon_1(n)$. In other words, $\mathcal{A}_n$ reaches the same state after reading u and v with probability at least $1 - \epsilon_1(n)$.
- If $\delta(q_0, u) \ne \delta(q_0, v)$, then $\mathsf{Prob}_{q \sim \iota_n}[\delta_n(q, u) \ne \delta_n(q, v)] \ge 1 - \epsilon_0(n)$. In other words, $\mathcal{A}_n$ reaches different states after reading u and v with probability at least $1 - \epsilon_0(n)$.

The *space complexity* of an (ϵ_0, ϵ_1)-distinguisher $\mathcal{R} = (\mathcal{A}_n)_{n \ge 0}$ is the function $s(\mathcal{R}, \cdot)\colon \mathbb{N} \to \mathbb{N}$ defined by $s(\mathcal{R}, n) = \lceil \log_2 |Q_n| \rceil$ where Q_n is the set of states of $\mathcal{A}_n$. The motivation is that states of $\mathcal{A}_n$ can be encoded by bit strings of length at most $s(\mathcal{R}, n)$. Throughout, we assume that $s(\mathcal{R}, n)$ is monotone (for $n < m$ with $s(\mathcal{R}, n) > s(\mathcal{R}, m)$, we can replace $\mathcal{A}_n$ by $\mathcal{A}_m$).

If the functions ϵ_0 and ϵ_1 coincide, then we speak of an ϵ_0-*distinguisher*. A 0-distinguisher is called a *deterministic distinguisher*. In this case, every semiPFA $\mathcal{A}_n$ can be converted into a DFA by replacing the distribution ι_n by any fixed initial state $q_{0,n} \in Q_n$ with $\iota_n(q_{0,n}) > 0$.

Every deterministic automaton $\mathcal{A} = (Q, \Sigma, q_0, \delta)$ has a deterministic distinguisher with space complexity $\lceil \log_2 \gamma_{\mathcal{A}}(n) \rceil$ where $\gamma_{\mathcal{A}}(n)$ is the growth function of $\mathcal{A}$ defined above – for every $n \ge 0$ one simply restricts $\mathcal{A}$ to states in $\{\delta(q_0, w) \mid w \in \Sigma^{\le n}\}$ and complete the transitions arbitrarily. A simple fooling argument shows that this yields a space-optimal deterministic distinguisher. Therefore, deterministic distinguishers are of limited theoretical interest, and our focus will be on ϵ-distinguishers for $\epsilon > 0$. Then, the growth $\gamma_{\mathcal{A}}$ and error probability ϵ give rise to the following lower bounds for space; for details, see [11].

▶ **Theorem 3.1.** *Let $\mathcal{A}$ be a deterministic automaton, and let $\mathcal{R}$ be an ϵ-distinguisher for $\mathcal{A}$.*
- *If $\epsilon \leq 1/2 - \epsilon'$ for a constant $\epsilon' > 0$, then $s(\mathcal{R}, n) \geq \Omega(\log\log \gamma_{\mathcal{A}}(n))$.* [11, Thm. 13]
- *For all sufficiently large n, $s(\mathcal{R}, n) \geq \min\{\log_2 \gamma_{\mathcal{A}}(n), \log_2(1/\epsilon(n))\}$.* [11, Thm. 14]

An ϵ-distinguisher for a deterministic automaton $\mathcal{A}$ can be seen as a non-uniform probabilistic approximation of $\mathcal{A}$ by finite systems. Proving results in this general setting seems to be difficult. For the remainder of this article, we therefore focus on distinguishers for monoids (and semigroups) and derive space bounds from their algebraic structure.

Let M be a monoid with finite generating set Σ. An (ϵ_0, ϵ_1)-*distinguisher* for M with respect to Σ is an (ϵ_0, ϵ_1)-distinguisher, as above, for the Cayley graph $\mathcal{A}_{M,\Sigma}$. In other words, an (ϵ_0, ϵ_1)-distinguisher for M is a collection of semiPFAs $\mathcal{R} = (\mathcal{A}_n)_{n\geq 0}$ such that, for all sufficiently large n and all words $u, v \in \Sigma^{\leq n}$, the automaton $\mathcal{A}_n$ reaches the same state after reading u and v with probability at least $1 - \epsilon_1(n)$ whenever $u \equiv_M v$, and it reaches different states with probability at least $1 - \epsilon_0(n)$ whenever $u \not\equiv_M v$.

Since every f.g. infinite monoid M has growth $\Omega(n)$, Theorem 3.1 implies that every ϵ-distinguisher for M and a constant $\epsilon < 1/2$ has space complexity $\Omega(\log\log n)$.

In the following, we are mainly interested in distinguishers for monoids. Nevertheless, many of the constructions that we will encounter are more adequately formulated in the language of semigroups. There are two canonical ways to extend the above to a semigroup S: by restricting to non-empty words, or by working with the smallest monoid S^1 that contains S. The former approach is more conceptual, while the latter allows for a uniform treatment; however, the differences between them are largely negligible. We adopt the latter approach in this work – thus, by a distinguisher for a semigroup S, we mean a distinguisher for S^1.

4 Commutative semigroups

By Ore's theorem [18] (see also [4, Prop. 3.2]), every cancellative commutative semigroup S embeds into an abelian group, namely the group of fractions $S^{-1}S$. Clearly, if S is finitely generated, then so is $S^{-1}S$. Since f.g. abelian groups have $((\log n)^{-c}, 0)$-distinguishers with space complexity $\mathcal{O}(\log\log n)$ for every $c > 0$ by [11, Thm. 25],[3] we obtain the following.

▶ **Lemma 4.1.** *For every f.g. cancellative commutative semigroup S and every constant $c > 0$, there exists a $((\log n)^{-c}, 0)$-distinguisher with space complexity $\mathcal{O}(\log\log n)$.*

The goal of this section is to generalize Lemma 4.1 to all finitely generated commutative semigroups, that is, to prove the following result.

▶ **Theorem 4.2.** *For every f.g. commutative semigroup S and every constant $c > 0$, there exists a $((\log n)^{-c}, 0)$-distinguisher with space complexity $\mathcal{O}(\log\log n)$.*

Note that, by Theorem 3.1, the inverse polylogarithmic error probability in Theorem 4.2 cannot be improved to, e.g., an inverse polynomial error probability.

Our proof of Theorem 4.2 relies on a structure theorem for f.g. commutative semigroups, which we will now briefly explain. It expresses each such semigroup S as a *subdirect product*, meaning that S is a subsemigroup of a direct product $\prod_{i\in I} S_i$ of semigroups S_i such that, moreover, $\pi_i(S) = S_i$ for all $i \in I$ where $\pi_j \colon \prod_{i\in I} S_i \to S_j$ denotes the projection onto S_j.

[3] Since this result is central for us, we give a short proof sketch: Consider a f.g. abelian group G. It is a direct product of cyclic groups $Z_1, \ldots, Z_k$. Moreover, one can assume that $Z_i \cong \mathbb{Z}$ for every i as elements of a finite cyclic group need space $\mathcal{O}(1)$. By storing an element of G as a k-tuple of binary encoded integers, one obtains a deterministic distinguisher for G with space complexity $\mathcal{O}(\log n)$. Storing each integer modulo a randomly chosen prime of bit length $\mathcal{O}(\log\log n)$, yields a $((\log n)^{-c}, 0)$-distinguisher with space complexity $\mathcal{O}(\log\log n)$ instead.

More precisely, the structure theorem [4, Thm. VI.2.2] states that every f.g. commutative semigroup S is a subdirect product of *finitely many* semigroups $S_1, \dots, S_k$ where each S_i is

- a f.g. cancellative commutative semigroup (hence, covered by Lemma 4.1),
- a f.g. commutative nilsemigroup (hence, necessarily finite), or
- a f.g. subelementary semigroup.

Herein, a *subelementary* semigroup S is a commutative semigroup S which can be written as the disjoint union $S = C \uplus N$ where C is a subsemigroup which is cancellative in S (that is, $ac = bc$ implies $a = b$ for all $a, b \in S$ and $c \in C$) and N is a nilsemigroup and an ideal in S.

There is also a related notion of an *elementary* semigroup, that is, a subelementary semigroup $S = C \uplus N$, as above, where C is an abelian group. It is easy to see that such an S is a monoid with neutral element $1 \in C$. A variant of Ore's theorem implies that every subelementary semigroup $S = C \uplus N$ embeds into an elementary semigroup, namely the semigroup of fractions $C^{-1}S$; see [4, Prop. VI.3.1]. Moreover, if S is finitely generated, then so is $C^{-1}S$; see [4, Prop. VI.3.3]. Hence, for the purpose of proving Theorem 4.2 it suffices to show the following result.

▶ **Proposition 4.3.** *For every f.g. elementary semigroup S and every constant $c > 0$, there exists a $((\log n)^{-c}, 0)$-distinguisher with space complexity $\mathcal{O}(\log \log n)$.*

Proof. Let $S = G \uplus N$ be a f.g. elementary semigroup where G is an abelian subgroup of S and N is an ideal of S and a nilsemigroup. Further, let $\Sigma \subseteq S$ be a finite generating set for S. We write Σ as a disjoint union $\Sigma = \Sigma_G \uplus \Sigma_N$ with $\Sigma_G = \Sigma \cap G$ and $\Sigma_N = \Sigma \cap N$. Then $\Sigma_G \subseteq G$ is a generating set of G, as the complement of G in S is an ideal. In particular, G is finitely generated. Be aware, however, that N need not be finitely generated.

Let us now consider Green's $\mathcal{H}$-preorder on S, which we simply denote by $\leq$. Since S is a commutative semigroup, it can be defined by $a \leq b$ if and only if $a = bc$ for some $c \in S^1$, and the corresponding equivalence relation $\mathcal{H}$ (with $a \mathrel{\mathcal{H}} b$ if and only if $a \leq b \leq a$) is a congruence. Let $\pi \colon S \to X$ with $X = S/\mathcal{H}$ be the corresponding quotient homomorphism.

The monoid X consists of the $\mathcal{H}$-classes of S. It is a commutative *nilmonoid* [4, Cor. V.1.8], i.e., a monoid obtained from a commutative nilsemigroup by adjoining a neutral element. The neutral element $1 \in X$ corresponds to $\pi^{-1}(1) = G$, which forms a single $\mathcal{H}$-class of S. Likewise, the zero element $0 \in X$ corresponds to $\pi^{-1}(0) = \{0\} \subseteq N$. Moreover, Green's $\mathcal{H}$-preorder on X, also denoted by $\leq$, is a partial order with greatest element $1 \in X$ and least element $0 \in X$. The monoid X is generated by $\pi(\Sigma)$ as a semigroup (and by $\pi(\Sigma_N)$ as a monoid). Since every f.g. commutative nilmonoid is finite, so is the monoid X.

Let us now consider the action of the group G on S by right multiplication. For every $a \in S$ and $g \in G$, we have $a = agg^{-1} \leq ag \leq a$ and, hence, $a \mathrel{\mathcal{H}} ag$. Thus, the action of G preserves the $\mathcal{H}$-classes of S. In fact, the $\mathcal{H}$-classes of S are precisely the orbits under the action of G; see [4, Prop. IV.5.1]. Therefore, we can also write X as the quotient $X = S/G$.

Since G is an abelian group, the stabilizer $\mathrm{Stab}_G(a) = \{g \in G \mid ag = a\}$ of any $a \in S$ depends only on the orbit $aG \subseteq S$ of a, that is, on the $\mathcal{H}$-class $\pi(a) \in X$ of a. For $\alpha \in X$, we thus obtain a well-defined quotient group $G_\alpha \coloneqq G/\mathrm{Stab}_G(a)$ with $a \in \pi^{-1}(\alpha)$ arbitrary. We denote the corresponding quotient homomorphism by $r_\alpha \colon G \to G_\alpha$. It is easy to see that these maps are compatible with one another in the following sense.

▷ Claim 4.4. For all $g, h \in G$ and $\alpha, \beta \in X$ with $\alpha \leq \beta$, $r_\beta(g) = r_\beta(h)$ implies $r_\alpha(g) = r_\alpha(h)$.

Proof. Let $a, b, c \in S$ with $a = bc$, $\pi(a) = \alpha$, and $\pi(b) = \beta$. By definition of the mapping r_β, we have $r_\beta(g) = r_\beta(h)$ if and only if $gh^{-1} \in \mathrm{Stab}_G(b)$ if and only if $bg = bh$. The latter implies $ag = bcg = bch = ah$ and, therefore, $r_\alpha(g) = r_\alpha(h)$. ◁

Algorithm 1 A $((\log n)^{-c}, 0)$-distinguisher for elementary semigroups.

global variables: $\chi \in X$ and $q_\alpha \in Q_{\alpha,n}$ $(\alpha \in X)$

initialization:

1 **for** $\alpha \in X$ **do**
2 | choose q_α according to the distribution $\iota_{\alpha,n}$
3 **end**
4 $\chi \coloneqq 1$

next input letter: $a \in \Sigma$

5 **for** $\alpha \in X$ **do**
6 | **if** $\alpha \leq \chi \cdot \pi(a)$ **then**
7 | | $q_\alpha \coloneqq \delta_{\alpha,n}(q_\alpha, r_\alpha(g_{\chi,a}))$
8 | **else**
9 | | $q_\alpha \coloneqq q^*_\alpha$
10 | **end**
11 **end**
12 $\chi \coloneqq \chi \cdot \pi(a)$

Let us now choose for each $\alpha \in X$ a canonical representative $b_\alpha \in \pi^{-1}(\alpha) \subseteq S$ where $b_1 = 1$. At this point, every $a \in S$ can be written as $a = b_{\pi(a)}g$ for some $g \in G$, since a and $b_{\pi(a)}$ are contained in the same orbit under the action of G. The expression $b_{\pi(a)}g$ is not unique, but for $\alpha \in X$ and $g, h \in G$ we have $b_\alpha g = b_\alpha h$ if and only if $r_\alpha(g) = r_\alpha(h)$. Moreover, let us choose for each $\alpha \in X$ and $a \in \Sigma$ an element $g_{\alpha,a} \in G$ such that $b_\alpha a = b_{\alpha\cdot\pi(a)}g_{\alpha,a}$ holds in S. For $a \in \Sigma_G$ we can choose $g_{\alpha,a} = a$ since $b_\alpha a = b_{\alpha\cdot\pi(a)}g_{\alpha,a} = b_{\alpha\cdot 1}g_{\alpha,a} = b_\alpha g_{\alpha,a}$. Thus, for $a_1, a_2, \ldots, a_n \in \Sigma$, we have

$$a_1 a_2 \cdots a_n = b_{\chi_n} g_{\chi_0,a_1} g_{\chi_1,a_2} \cdots g_{\chi_{n-1},a_n}$$

in S where $\chi_0 = 1 \in X$ and $\chi_i = \pi(a_1 a_2 \cdots a_i) \in X$ for $1 \leq i \leq n$. Finally, let

$$\Gamma = \{g_{\alpha,a} \mid \alpha \in X, a \in \Sigma\},$$

which finitely generates G since $\Sigma_G \subseteq \Gamma$.

With all prerequisites in place, let us now construct a distinguisher for the semigroup S. By [11, Thm. 25], we can fix for each $\alpha \in X$ a $((\log n)^{-c}, 0)$-distinguisher $\mathcal{R}_\alpha = (\mathcal{A}_{\alpha,n})_{n\geq 0}$ for the f.g. abelian group G_α with respect to the generating set $\Gamma_\alpha \coloneqq r_\alpha(\Gamma)$ such that the space complexity of $\mathcal{R}_\alpha$ is $\mathcal{O}(\log\log n)$. Now fix the input length n. For our construction, we use the semiPFAs $\mathcal{A}_{\alpha,n} = (Q_{\alpha,n}, \Gamma_\alpha, \iota_{\alpha,n}, \delta_{\alpha,n})$ and (arbitrary) default states $q^*_\alpha \in Q_{\alpha,n}$. The semiPFA $\mathcal{A}_n$ for our distinguisher $\mathcal{R} = (\mathcal{A}_n)_{n\geq 0}$ for S is implicitly defined by Algorithm 1. From its global variables we see that the space complexity of the distinguisher is

$$s(\mathcal{R}, n) = \log_2|X| + \sum_{\alpha\in X} s(\mathcal{R}_\alpha, n) \leq \mathcal{O}(\log\log n).$$

Let $w \in \Sigma^{\leq n}$ be an input word with S-value s_w. Then Algorithm 1 incrementally computes $\beta \coloneqq \pi(s_w) \in X$, which is stored in the program variable χ, in lines 4 and 12. It also implicitly computes a word $\tilde{w} \in \Gamma^*$ as a concatenation of elements $g_{\alpha,a} \in \Gamma$ in line 7. This word satisfies $w \equiv_S b_\beta \tilde{w}$ by construction; see the discussion following Claim 4.4. Thus, the S-value s_w of the word w is uniquely determined by β and the G_β-value g_β of the word $r_\beta(\tilde{w}) \in \Gamma^*_\beta$.

Note that Algorithm 1 simulates the semiPFA $\mathcal{A}_{\alpha,n}$ on input $\tilde{w}_\alpha$ for each $\alpha \leq \beta$ in lines 2 and 7. For $\alpha \in X$ with $\alpha \not\leq \beta$ such a simulation is performed initially, but the state q_α is then eventually set to the default state q^*_α in line 9.

It is now straightforward to analyze the error probability of $\mathcal{A}_n = (Q_n, \Sigma, \iota_n, \delta_n)$ for two input words $u, v \in \Sigma^{\leq n}$. Let us first assume that $u \not\equiv_S v$. If $\pi(u) \not\equiv_X \pi(v)$, then the error probability is zero, since the variable χ will have different values after reading u and v. If $\pi(u)$ and $\pi(v)$ have the same X-value β, then we must have $r_\beta(\tilde{u}) \not\equiv_{G_\beta} r_\beta(\tilde{v})$ where $\tilde{u}, \tilde{v} \in \Gamma^*$ are the implicitly computed words on inputs u and v, respectively (otherwise, $u \equiv_S v$ would hold). The algorithm inputs $r_\beta(\tilde{u})$ and $r_\beta(\tilde{v})$ into the semiPFA $\mathcal{A}_{\beta,n}$ and both words have length at most n; hence, the resulting states of $\mathcal{A}_{\beta,n}$ differ with probability at least $1 - (\log n)^{-c}$.

On the other hand, for $u, v \in \Sigma^{\leq n}$ with $u \equiv_S v$ the error probability is equal to zero: The variable χ holds the X-value β of $\pi(u) \equiv_X \pi(v)$ after reading u and v, respectively. Moreover, since $r_\beta(\tilde{u}) \equiv_{G_\beta} r_\beta(\tilde{v})$ (and hence $r_\alpha(\tilde{u}) \equiv_{G_\alpha} r_\beta(\tilde{v})$ for all $\alpha \leq \beta$ by Claim 4.4) *all* semiPFAs $\mathcal{A}_{\alpha,n}$ will be in the same state with probability one after reading u and v. If $\alpha \not\leq \beta$ this will be the default state q^*_α. ◀

By [11, Thm. 25], every f.g. nilpotent group also has a $((\log n)^{-c}, 0)$-distinguisher with space complexity $\mathcal{O}(\log \log n)$ for every $c > 0$. Thus, it is natural to ask whether the statement in Theorem 4.2 extends beyond commutative semigroups to nilpotent semigroups; see Section 2.1. For f.g. cancellative nilpotent semigroups this is the case: such a semigroup S again embeds into the group of fractions $S^{-1}S$, and the latter is a f.g. nilpotent group [13, 17].

▶ **Corollary 4.5.** *For every f.g. cancellative nilpotent semigroup S and every constant $c > 0$, there exists a $((\log n)^{-c}, 0)$-distinguisher with space complexity $\mathcal{O}(\log \log n)$.*

We will also prove a complementary result for f.g. regular nilpotent semigroups (Corollary 5.6) in the next section. In general, however, the answer to the above question is negative due to the existence of f.g. nilpotent semigroups of exponential growth [23, p. 512]. In this case, every ϵ-distinguisher for a constant $\epsilon < 1/2$ has space complexity $\Omega(\log n)$ by Theorem 3.1. An interesting open problem in this context is whether every f.g. nilpotent semigroup has an ϵ-distinguisher with $\epsilon \leq 1/2 - \epsilon'$ for some constant $\epsilon' > 0$ and space complexity $\mathcal{O}(\log n)$.

5 Semilattice decompositions

Our second main result for distinguishers concerns an important decomposition technique for semigroups. A *semilattice decomposition* of a semigroup S is an epimorphism (that is, a surjective homomorphism) $\pi \colon S \to Y$ onto a semilattice Y. Equivalently, it consists of a partition of $S = \bigcup_{\alpha \in Y} S_\alpha$ indexed by elements α of a semilattice Y such that $S_\alpha S_\beta \subseteq S_{\alpha\beta}$ holds for all $\alpha, \beta \in Y$. In particular, each S_α is a subsemigroup of S, since $\alpha^2 = \alpha$ holds for every $\alpha \in Y$. The $\mathcal{H}$-preorder $\leq$ on Y is a partial order.

In the following, we will work with the new notion of a *retractive* semilattice decomposition. For this, we require the existence of semigroup homomorphisms $r_\alpha \colon S^\uparrow_\alpha \to S_\alpha$, where $S^\uparrow_\alpha \subseteq S$ is the union of all S_β with $\alpha \leq \beta$,[4] such that each r_α restricts to the identity mapping on S_α. Thus, r_α is a *retraction* of $S^\uparrow_\alpha$ onto S_α for each $\alpha \in Y$.

▶ Remark 5.1. The reader may already be familiar with *strong* semilattice decompositions, which play an important role in the literature. In this case, one requires the existence of homomorphisms $h^\beta_\alpha \colon S_\beta \to S_\alpha$ for all $\alpha, \beta \in Y$ with $\alpha \leq \beta$ with the following properties.

[4] Note that $S^\uparrow_\alpha$ is indeed a subsemigroup of S, since $\alpha \leq \beta_1$ and $\alpha \leq \beta_2$ together imply $\alpha \leq \beta_1\beta_2$ in Y.

(i) The mapping h_α^α is the identity on S_α; and if $\alpha \leq \beta \leq \gamma$, then $h_\alpha^\beta \circ h_\beta^\gamma = h_\alpha^\gamma$.
(ii) For all $a_1 \in S_{\beta_1}$ and $a_2 \in S_{\beta_2}$, we have $a_1 a_2 = h_\alpha^{\beta_1}(a_1) h_\alpha^{\beta_2}(a_2)$ in S where $\alpha = \beta_1\beta_2$.

Roughly speaking, the first property ensures that the maps h_α^β are compatible with the structure of Y, while the second ensures compatibility with the structure of S.

We note that every strong semilattice decomposition is retractive: one can simply take the retraction $r_\alpha \colon S_\alpha^\uparrow \to S_\alpha$ to be the union of all $h_\alpha^\beta \colon S_\beta \to S_\alpha$ with $\alpha \leq \beta$; see [11, Sec. 5.6]. Moreover, if all S_α in a semilattice decomposition $S = \bigcup_{\alpha \in Y} S_\alpha$ are groups, then it is a strong semilattice decomposition [19, Thm. II.2.6]. This result does not generalize to the case where all S_α are monoids (see [11, Ex. 54] for an example), but the following holds.

▶ **Lemma 5.2.** *Suppose that $S = \bigcup_{\alpha \in Y} S_\alpha$ is a semilattice decomposition in which every subsemigroup S_α is a monoid. Then the semilattice decomposition $S = \bigcup_{\alpha \in Y} S_\alpha$ is retractive.*

Proof. Let $r_\alpha \colon S_\alpha^\uparrow \to S_\alpha$ be given by $r_\alpha(a) = ae_\alpha$ where $e_\alpha \in S_\alpha$ is the neutral element. Then r_α restricts to the identity mapping on S_α. It is a homomorphism since

$$r_\alpha(a_1a_2) = a_1a_2e_\alpha = a_1e_\alpha a_2e_\alpha = r_\alpha(a_1)r_\alpha(a_2)$$

for all $a_1, a_2 \in S_\alpha^\uparrow$, where the second equality uses $a_2e_\alpha \in S_\alpha$ and, hence, $a_2e_\alpha = e_\alpha a_2 e_\alpha$. ◀

▶ **Lemma 5.3.** *Suppose that $S = \bigcup_{\alpha \in Y} S_\alpha$ is a retractive semilattice decomposition. If the semigroup S is generated by $\Sigma \subseteq S$, then the subsemigroups S_α are generated by $r_\alpha(\Sigma \cap S_\alpha^\uparrow)$.*

Proof. Every element $a \in S_\alpha$ can be written as a product $a = a_1 \cdots a_n$ with $a_1, \ldots, a_n \in \Sigma$. If $a_i \in S_{\beta_i}$ for $1 \leq i \leq n$, then $\alpha = \beta_1 \cdots \beta_n$ since $a = a_1 \cdots a_n$. Hence, $\alpha \leq \beta_i$ for all i, which shows that $a_1, \ldots, a_n \in \Sigma \cap S_\alpha^\uparrow$. In turn, this yields $a = r_\alpha(a) = r_\alpha(a_1) \cdots r_\alpha(a_n)$. Since $a \in S_\alpha$ was arbitrary, this shows that S_α is indeed generated $r_\alpha(\Sigma \cap S_\alpha^\uparrow)$. ◀

Note that if $S = \bigcup_{\alpha \in Y} S_\alpha$ is a semilattice decomposition of a f.g. semigroup S, then the semilattice Y is, as a homomorphic image of S, finitely generated as well, and hence finite as remarked in Section 2.1. By Lemma 5.3, each subsemigroup S_α is also a f.g. semigroup if the semilattice decomposition is retractive.

Let us now turn to the main result of this section. Informally speaking, it asserts that for a retractive semilattice decomposition $S = \bigcup_{\alpha \in Y} S_\alpha$ distinguishers for the subsemigroups S_α can be efficiently combined into a distinguisher for S.

▶ **Theorem 5.4.** *Suppose that $S = \bigcup_{\alpha \in Y} S_\alpha$ is a retractive semilattice decomposition and that S is finitely generated. Further, suppose that every S_α has an $(\epsilon_0(n), \epsilon_1(n))$-distinguisher with space complexity bounded by some function $s(n)$. Then there is a constant d such that S has an $(\epsilon_0(dn), |Y| \cdot \epsilon_1(dn))$-distinguisher with space complexity bounded by $|Y| \cdot s(dn) + \mathcal{O}(1)$.*

Proof sketch. Let $\Sigma \subseteq S$ be a finite generating set. By Lemma 5.3, each subsemigroup S_α is generated by the finite set $\Sigma_\alpha := r_\alpha(\Sigma \cap S_\alpha^\uparrow)$. By assumption, we can fix for each $\alpha \in Y$, and some suitable constant d, an $(\epsilon_0(dn), \epsilon_1(dn))$-distinguisher $\mathcal{R}_\alpha = (\mathcal{A}_{\alpha,n})_{n \geq 0}$ for S_α with respect to the generating set Σ_α such that the space complexity of $\mathcal{R}_\alpha$ is $s(dn)$. Here, the constant factor d is required due to dependence on the specific choice of generating sets.

Now fix the input length n. Our construction is similar to the one given in the proof of Proposition 4.3: For any input word $w \in \Sigma^{\leq n}$ with S-value s_w, we incrementally compute the element $\beta := \pi(s_w) \in Y$; thus, $s_w \in S_\beta$. For each $\alpha \leq \beta$, we additionally simulate the semiPFA $\mathcal{A}_{\alpha,n}$ on the word $r_\alpha(w) \in \Sigma_\alpha^{\leq n}$. As in Algorithm 1, the simulation is initially performed for all $\alpha \in Y$, but eventually aborted using a default state if $\alpha \not\leq \beta$.

The analysis of this construction is straightforward and closely mimics the corresponding parts in the proof of Proposition 4.3. We therefore omit the details; see the full version [11]. ◀

We conclude this section with two applications of Theorem 5.4. Both of these are based on the notion of *Clifford semigroups*, meaning semilattices of groups. Equivalently, a Clifford semigroup S is an inverse semigroup where $ea = ae$ holds for all elements $a \in S$ and all idempotents $e \in S$; see e.g. [19, Thm. II.2.6]. Each Clifford semigroup S has a (unique) semilattice decomposition $S = \bigcup_{\alpha \in Y} S_\alpha$ in which all S_α are groups.[5] This semilattice decomposition is retractive by Lemma 5.2. We can therefore apply Theorem 5.4 to obtain a distinguisher for S from distinguishers for its constituent subgroups S_α.

Our first application concerns the Clifford semigroup freely generated by a finite set Σ, that is, the *free Clifford semigroup* $\mathsf{FCS}(\Sigma)$.[6] As is well known, see e.g. [19, p. 369], the semigroup $\mathsf{FCS}(\Sigma)$ is a semilattice of f.g. free groups. Hence, by combining Theorem 5.4 with the corresponding result for free groups [11, Thm. 24], we obtain the following.

▶ **Corollary 5.5.** *Every f.g. free Clifford semigroup* $\mathsf{FCS}(\Sigma)$ *has a* $(1/n^c, 0)$*-distinguisher with space complexity* $\mathcal{O}(\log n)$ *for every constant* $c > 0$*.*

Our second application concerns f.g. nilpotent semigroups that are regular (see Section 2.1). By a result of Lallement [8], such semigroups are semilattices of f.g. nilpotent groups. Hence, we obtain the following from Theorem 5.4 and [11, Thm. 25].

▶ **Corollary 5.6.** *Every f.g. regular nilpotent semigroup has a* $((\log n)^{-c}, 0)$*-distinguisher with space complexity* $\mathcal{O}(\log \log n)$ *for every constant* $c > 0$*.*

6 Lower bounds

In this section, we present lower bounds on the space complexity for distinguishers for certain inverse monoids, namely free inverse monoids and polycyclic monoids. These lower bounds will be proved using randomized communication complexity; see Section 2.3.

Recall from Section 2.1 that an inverse monoid is a regular monoid with unique inverses or, equivalently, commuting idempotents. For every set Γ, there is a *free inverse monoid* $\mathsf{FIM}(\Gamma)$ generated by Γ (as an inverse monoid). It can be constructed as follows.

First, consider the disjoint union $\Sigma = \Gamma \cup \Gamma^{-1}$ where $\Gamma^{-1} = \{a^{-1} \mid a \in \Gamma\}$ is a set of formal inverses. The mapping $a \mapsto a^{-1}$ extends to an involution on Σ with $(a^{-1})^{-1} = a$ for all $a \in \Sigma$, and an involution on Σ^* with $(a_1 a_2 \cdots a_n)^{-1} = a_n^{-1} \cdots a_2^{-1} a_1^{-1}$. Then $\mathsf{FIM}(\Gamma)$ can be defined as the quotient of Σ^* by the smallest congruence containing the *Vagner relations*

$$uu^{-1}u = u \quad (u \in \Sigma^*), \qquad uu^{-1}vv^{-1} = vv^{-1}uu^{-1} \quad (u, v \in \Sigma^*).$$

This description of $\mathsf{FIM}(\Gamma)$ is, however, of limited use for algorithmic purposes (such as solving the word problem). A more practical description was given by Munn [16]:

Consider the free group $\mathsf{FG}(\Gamma)$. Its elements can be identified with *reduced* words over the alphabet Σ, meaning words that contain no factor aa^{-1} with $a \in \Sigma$. Let $\mathsf{IRR}(\Sigma)$ denote the set of such words. Further, let $\mathsf{red} \colon \Sigma^* \to \mathsf{IRR}(\Sigma)$ denote the mapping where $\mathsf{red}(w)$ is the unique reduced word obtained from w by successively deleting all factors aa^{-1} with $a \in \Sigma$.

The *Munn tree* of $w \in \Sigma^*$ is the prefix-closed set $\mathsf{MT}(w) = \{\mathsf{red}(u) \mid u \text{ is a prefix of } w\}$. Clearly, $\mathsf{MT}(w) \subseteq \mathsf{IRR}(\Sigma)$. Under the identification of $\mathsf{IRR}(\Sigma)$ with $\mathsf{FG}(\Gamma)$, the Munn tree $\mathsf{MT}(w)$ induces a finite tree in the Cayley graph of $\mathsf{FG}(\Gamma)$: it is the trace of the path obtained

[5] The epimorphism $\pi \colon S \to Y$ mapping each $S_\alpha \subseteq S$ to $\alpha \in Y$ is the quotient by Green's $\mathcal{H}$-relation.

[6] Free semigroups can be defined for every variety of semigroups using the classical universal property. The details of this are not needed here and the interested reader may consult e.g. [22].

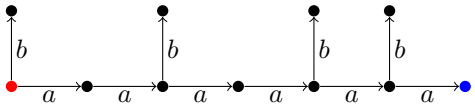

Figure 1 The Munn tree $\mathsf{MT}(w_0)$ of the word $w_0 = bb^{-1}aabb^{-1}aabb^{-1}abb^{-1}a$ that results from Alice's bit string 1010110 in the proof of Theorem 6.1. Every edge labelled with $x \in \{a, b\}$ has an inverse edge labelled with x^{-1}, which is not shown. The red vertex is the identity element of the free group and the blue vertex is $\mathsf{red}(w_0) = a^6$.

by reading w from left to right. Munn [16] proved that the pair $(\mathsf{MT}(w), \mathsf{red}(w))$ is a complete invariant of the image of w in $\mathsf{FIM}(\Gamma)$: two words $u, v \in \Sigma^*$ have identical images in $\mathsf{FIM}(\Gamma)$ if and only if they have identical Munn trees and identical images in $\mathsf{FG}(\Gamma)$.

In the following, we simply write FIM_k for $\mathsf{FIM}(\Gamma)$ where Γ is any finite set of size k.

▶ **Theorem 6.1.** *Let $k \geq 2$. Then, for every constant $\epsilon < 1/2$, every ϵ-distinguisher for the free inverse monoid FIM_k has space complexity $\Omega(n)$.*

Proof. It suffices to prove the lower bound for $k = 2$. For this, we make a reduction from the randomized one-way communication complexity of the *index problem* IDX_n; see Section 2.3.

Let $\mathcal{R} = (\mathcal{A}_n)_{n\geq 0}$ with $\mathcal{A}_n = (Q_n, \{a, b, a^{-1}, b^{-1}\}, \iota_n, \delta_n)$ be an ϵ-distinguisher with $\epsilon < 1/2$ for the free inverse monoid $\mathsf{FIM}_2 = \mathsf{FIM}(\{a, b\})$. We construct a randomized one-way communication protocol for IDX_n as follows. Let $x_1x_2\cdots x_n \in \{0,1\}^n$ be Alice's input, and let $i \in [1, n]$ be Bob's input. The protocol is as follows, where $m = 5n - 1$ and $t = bb^{-1}$.

- Alice chooses an initial state q_0 of $\mathcal{A}_m$ according to the distribution ι_m, computes the state $q_1 = \delta_m(q_0, w_0)$ where $w_0 = t^{x_1}at^{x_2}a\cdots t^{x_{n-1}}at^{x_n}$, and then sends q_1 to Bob.
- Bob uses the state q_1 to compute the state $q_2 = \delta_m(q_1, a^{i-n}ta^{n-i}) = \delta_m(q_0, w_0a^{i-n}ta^{n-i})$, and he then outputs 1 if $q_1 = q_2$ and 0 otherwise.

Now note that $w_0a^{i-n}ta^{n-i} \equiv_{\mathsf{FIM}_2} t^{x_1}a\cdots t^{x_{i-1}}a\,t\,a\,t^{x_{i+1}}a\cdots t^{x_{n-1}}a\,t^{x_n}$, which is equal to w_0 in FIM_2 if and only if $x_i = 1$. To see this, it suffices to consider the Munn trees of both words; see Figure 1. Alternatively, one can repeatedly use that, for $y \in \{0, 1\}$ and $j \geq 1$,

$$at^ya^{-j}ta^j \equiv_{\mathsf{FIM}_2} aa^{-1}at^ya^{-j}ta^j \equiv_{\mathsf{FIM}_2} aa^{-j}ta^ja^{-1}at^y \equiv_{\mathsf{FIM}_2} a^{-j+1}ta^{j-1}at^y,$$

where the first and third steps use the identity $a \equiv_{\mathsf{FIM}_2} aa^{-1}a$, and the second step uses that the words $a^{-1}a$, t^y, and $a^{-j}ta^j$ represent idempotents and, hence, commute.

Note that the length of the word $w_0a^{i-n}ta^{n-i}$ is bounded by $m = 5n-1$. Hence, if $x_i = 1$, then $q_1 = q_2$ with probability at least $1 - \epsilon$; and if $x_i = 0$, then $q_1 \neq q_2$ with probability at least $1 - \epsilon$. Thus, the protocol is correct. In our protocol, Alice sends at most $s(\mathcal{R}, 5n - 1)$ bits to Bob. Therefore, $s(\mathcal{R}, 5n - 1) \geq \Omega(n)$ by Theorem 2.1. Recalling that $s(\mathcal{R}, n)$ is monotone by our assumptions from Section 3, we obtain $s(\mathcal{R}, n) \geq \Omega(n)$. ◀

For the monoid FIM_1, which is also known as the *free monogenic inverse monoid*, the problem becomes easier. In this case, Munn trees are finite intervals $[a, b]$ in $\mathbb{Z}$ (note that the free group generated by a single element is $\mathbb{Z}$). Elements of FIM_1 can thus be represented by triples $[a, i, b] \in \mathbb{Z}^3$ with $a \leq b$ and $0, i \in [a, b]$. The multiplication becomes

$$[a, i, b] \cdot [c, j, d] = [\min\{a, i + c\}, i + j, \max\{b, i + d\}],$$

and inversion is given by $[a, i, b]^{-1} = [a - i, -i, b - i]$; see [19, Prop. IX.1.1 and Cor. IX.1.2]. In this representation, the monoid generators are $s = [0, +1, +1]$ and $s^{-1} = [-1, -1, 0]$.

▶ **Theorem 6.2.** *The free inverse monoid* FIM_1 *has a deterministic distinguisher with space complexity* $\mathcal{O}(\log n)$*. Moreover, for every constant* $\epsilon < 1/2$*, every* ϵ*-distinguisher for the free inverse monoid* FIM_1 *has space complexity* $\Omega(\log n)$*.*

Proof. The upper bound is obtained by storing all triples $[a, i, b] \in \mathbb{Z}^3$ in binary. The lower bound is shown by a reduction from the randomized one-way communication complexity of the *greater-than* problem GT_n; see Section 2.3.

Assume that $\mathcal{R} = (\mathcal{A}_n)_{n\geq 0}$ is an ϵ-distinguisher for FIM_1 with $\mathcal{A}_n = (Q_n, \{s, s^{-1}\}, \iota_n, \delta_n)$ and $\epsilon < 1/2$. We construct a one-way randomized communication protocol for GT_n as follows. Let x be the input of Alice and y be the input of Bob, where $1 \leq x, y \leq n$. Alice chooses an initial state q_0 of $\mathcal{A}_{4n}$ according to the distribution ι_{4n}, computes the state $q_1 = \delta_{4n}(q_0, s^x s^{-x})$, and sends it to Bob. Bob then computes the state $q_2 = \delta_{4n}(q_1, s^y s^{-y}) = \delta_{4n}(q_0, s^x s^{-x} s^y s^{-y})$ and outputs 1 if $q_1 \neq q_2$ and 0 otherwise. Note that $s^x s^{-x} s^y s^{-y}$ has length at most $4n$ and

$$s^x s^{-x} s^y s^{-y} \equiv_{\mathsf{FIM}_1} s^{\max\{x,y\}} s^{-\max\{x,y\}},$$

which equals $s^x s^{-x}$ if and only if $x \geq y$. Hence, if $x < y$, then $q_1 \neq q_2$ holds with probability at least $1 - \epsilon$, so that Bob outputs 1 with probability at least $1 - \epsilon$. On the other hand, if $x \geq y$, then Bob outputs 0 with probability at least $1 - \epsilon$. Hence, the protocol is correct.

In our protocol, Alice sends at most $s(\mathcal{R}, 4n)$ bits to Bob. Therefore, $s(\mathcal{R}, 4n) \geq \Omega(\log n)$ by Theorem 2.1, which implies $s(\mathcal{R}, n) \geq \Omega(\log n)$ by monotonicity. ◀

Similar lower bounds can be shown for polycyclic monoids. For any alphabet Σ, the *polycyclic monoid* $\mathsf{PM}(\Sigma)$ is the inverse monoid generated by the partial bijections

$$p_a\colon \ \Sigma^* \to \Sigma^* a \ \text{ with } \ p_a(w) = wa \tag{2}$$

$$p_a^{-1}\colon \ \Sigma^* a \to \Sigma^* \ \text{ with } \ p_a^{-1}(wa) = w \tag{3}$$

where $a \in \Sigma$. The neutral element of $\mathsf{PM}(\Sigma)$ is the identity mapping on Σ^*. As above, we simply write PM_k for $\mathsf{PM}(\Sigma)$ where Σ is any finite alphabet of size k.

The following result is shown in a similar way as Theorem 6.1, but here we reduce from the augmented index problem.

▶ **Theorem 6.3.** *Let* $k \geq 2$*. Then, for every constant* $\epsilon < 1/2$*, every* ϵ*-distinguisher for the polycyclic monoid* PM_k *has space complexity* $\Omega(n)$*.*

Proof. We prove the lower bound for $k = 2$ by a reduction from the randomized one-way communication complexity of the augmented index problem augIDX_n. Throughout, we let $\mathsf{PM}_2 = \mathsf{PM}(\{0, 1\})$. For $i = 0, 1$, we then write $a_i \coloneqq p_i^{-1}$ and $b_i \coloneqq p_i$ for the partial bijections generating PM_2, which were defined in (2) and (3). We use the standard order (right-to-left) for composition of mappings, so that $a_0 b_0$ and $a_1 b_1$ are both equal to the identity on $\{0, 1\}^*$. Also note that $a_0 b_1$ is the totally undefined mapping.

Let $\mathcal{R} = (\mathcal{A}_n)_{n\geq 0}$ be an ϵ-distinguisher for PM_2 with $\mathcal{A}_n = (Q_n, \{a_0, b_0, a_1, b_1\}, \iota_n, \delta_n)$ and a constant $\epsilon < 1/2$. We will construct a randomized one-way communication protocol for the problem augIDX_n. Let $x_1 x_2 \cdots x_n \in \{0, 1\}^n$ be the input of Alice. The input of Bob is $i \in [1, n]$ and the promised suffix $x_{i+1} \cdots x_n$ of Alice's input. Bob has to find out the bit x_i.

Let $m = 2n + 1$. Alice chooses an initial state q_0 of $\mathcal{A}_m$ according to the distribution ι_m, computes the state $q_1 = \delta_m(q_0, w_0)$ where $w_0 = a_{x_1} a_{x_2} \cdots a_{x_n}$, and sends it to Bob. Then Bob computes the states q_2 and q_3 given by

$$\begin{aligned} q_2 &= \delta_m(q_1, b_{x_n} \cdots b_{x_{i+1}}) &&= \delta_m(q_0, w_0 b_{x_n} \cdots b_{x_{i+1}}), \\ q_3 &= \delta_m(q_2, b_1 a_1) &&= \delta_m(q_0, w_0 b_{x_n} \cdots b_{x_{i+1}} b_1 a_1). \end{aligned}$$

He outputs 1 if $q_2 = q_3$, and 0 otherwise.

Note that $w_0 b_{x_n} \cdots b_{x_{i+1}} \equiv_{\mathsf{PM}_2} a_{x_1} a_{x_2} \cdots a_{x_i}$ and that $x_i = 1$ if and only if

$w_0 b_{x_n} \cdots b_{x_{i+1}} \equiv_{\mathsf{PM}_2} w_0 b_{x_n} \cdots b_{x_{i+1}} b_1 a_1$.

The length of the word $w_0 b_{x_n} \cdots b_{x_{i+1}} b_1 a_1$ is bounded by $m = 2n + 1$. Hence, if $x_i = 1$, then $q_2 = q_3$ with probability at least $1 - \epsilon$, and if $x_i = 0$, then $q_2 \neq q_3$ with probability at least $1 - \epsilon$. This shows that the protocol is indeed correct. By Theorem 2.1, Alice has to send $\Omega(n)$ bits to Bob in every randomized one-way protocol for augIDX_n. As Alice sends at most $s(\mathcal{R}, 2n + 1)$ bits in our protocol, $s(\mathcal{R}, 2n + 1) \geq \Omega(n)$ and hence $s(\mathcal{R}, n) \geq \Omega(n)$. ◀

Similar to Theorem 6.2, we also obtain complexity bounds for the polycyclic monoid PM_1, which is better known as the *bicyclic monoid*. It is known to be the quotient of the free monoid $\{a, b\}^*$ modulo the congruence generated by the identity $ab = \varepsilon$, which is usually expressed as $\mathsf{PM}_1 = \langle a, b \mid ab = 1 \rangle$. Note that the bicyclic monoid as well as the free monogenic inverse monoid have polynomial growth.

▶ **Theorem 6.4.** *The bicyclic monoid* PM_1 *has a deterministic distinguisher with space complexity* $\mathcal{O}(\log n)$*. Moreover, for every constant* $\epsilon < 1/2$*, every* ϵ*-distinguisher for the bicyclic monoid* PM_1 *has space complexity* $\Omega(\log n)$*.*

Proof. We start with the upper bound. Every element of the bicyclic monoid PM_1 can be uniquely written as $b^m a^n$ for $m, n \geq 0$ and the multiplication is defined by the following rule:

$$(b^k a^\ell)(b^m a^n) = \begin{cases} b^k a^{\ell - m + n} & \text{if } \ell \geq m, \\ b^{k+m-\ell} a^n & \text{if } \ell < m. \end{cases}$$

The upper bound follows from this description of PM_1 by storing the exponents of the generators a and b in binary notation.

For the lower bound we make a reduction from the randomized one-way communication complexity of the greater-than problem GT_n similar to the proof of Theorem 6.2. Assume that $\mathcal{R} = (\mathcal{A}_n)_{n \geq 0}$ is an ϵ-distinguisher for PM_1 with $\mathcal{A}_n = (Q_n, \{a, b\}, \iota_n, \delta_n)$ and a constant $\epsilon < 1/2$. We will construct a randomized one-way communication protocol for GT_n. Let $x \in [1, n]$ be the input of Alice and $y \in [1, n]$ be the input of Bob. Let $m = 3n$. Alice chooses an initial state q_0 of $\mathcal{A}_m$ according to the distribution ι_m, computes the state $q_1 = \delta_m(q_0, a^x)$ and sends it to Bob. Bob then computes the state $q_2 = \delta_m(q_1, b^y a^y) = \delta_m(q_0, a^x b^y a^y)$ and outputs 1 if $q_1 \neq q_2$ and 0 otherwise.

Note that the word $a^x b^y a^y$ has length at most $3n = m$ and that

$$a^x b^y a^y \equiv_{\mathsf{PM}_1} \begin{cases} a^{x-y} a^y \equiv_{\mathsf{PM}_1} a^x & \text{if } x \geq y, \\ b^{y-x} a^y \not\equiv_{\mathsf{PM}_1} a^x & \text{if } x < y. \end{cases}$$

Hence, if $x < y$, then $q_1 \neq q_2$ holds with probability at least $1 - \epsilon$; thus, Bob outputs 1 with probability at least $1 - \epsilon$. On the other hand, if $x \geq y$, then Bob outputs 0 with probability at least $1 - \epsilon$. This shows that our protocol is correct. Alice sends $s(\mathcal{R}, 3n)$ bits to Bob in our protocol, which implies $s(\mathcal{R}, 3n) \geq \Omega(\log n)$ by Theorem 2.1. Hence, $s(\mathcal{R}, n) \geq \Omega(\log n)$. ◀

7 Conclusion

We started the investigation of randomized streaming algorithms for word problems in finitely generated semigroups using the concept of a distinguisher. For finitely generated commutative semigroups we constructed distinguishers with space complexity $\mathcal{O}(\log \log n)$

and inverse polylogarithmic error probability. The same holds for finitely generated nilpotent semigroups that are cancellative or regular. It would be interesting to see further examples of semigroups having distinguishers with space complexity $\mathcal{O}(\log \log n)$, which is asymptotically optimal due to the $\Omega(\log \log n)$-bound valid for every finitely generated infinite semigroup.

An open question regarding the distinguisher model in general is whether it allows for efficient error reduction. The standard technique is to run k independent copies. This reduces the error in the negative case, where inputs should be distiguished. In the positive case, however, this results in an error of at most $1 - (1 - \epsilon_1(n))^k$, which converges to 1 for $k \to \infty$.

References

1 Khanh Do Ba, Piotr Indyk, Eric Price, and David P. Woodruff. Lower bounds for sparse recovery. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2010*, pages 1190–1197. SIAM, 2010. doi:10.1137/1.9781611973075.95.

2 Alfred H. Clifford and Gordon B. Preston. *The Algebraic Theory of Semigroups.* Number 7 in Mathematical Surveys. AMS, 1961.

3 P. A. Grillet. *Semigroups: An Introduction to the Structure Theory.* Taylor & Francis, 1995. doi:10.4324/9780203739938.

4 P. A. Grillet. *Commutative Semigroups.* Springer, 2001. doi:10.1007/978-1-4757-3389-1.

5 Olga Kharlampovich and Mark V. Sapir. Algorithmic problems in varieties. *International Journal of Algebra and Computation*, 5(4-5):379–602, 1995. doi:10.1142/S0218196795000227.

6 Ilan Kremer, Noam Nisan, and Dana Ron. On randomized one-round communication complexity. *Computational Complexity*, 8(1):21–49, 1999. doi:10.1007/s000370050018.

7 Eyal Kushilevitz and Noam Nisan. *Communication Complexity.* Cambridge University Press, 1997. doi:10.1017/CBO9780511574948.

8 Gérard Lallement. On nilpotency and residual finiteness in semigroups. *Pacific Journal of Mathematics*, 42(3):693–700, 1972. doi:10.2140/pjm.1972.42.693.

9 Markus Lohrey. Parallel complexity in group theory. In Benjamin Steinberg, editor, *Languages and Automata, GAGTA BOOK 3*, chapter 5, pages 183–264. De Gruyter, Berlin, Boston, 2024. doi:10.1515/9783110984323-005.

10 Markus Lohrey and Lukas Lück. Streaming word problems. In *Proceedings of the 47th International Symposium on Mathematical Foundations of Computer Science, MFCS 2022*, volume 241 of *LIPIcs*, pages 72:1–72:15. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022. doi:10.4230/LIPIcs.MFCS.2022.72.

11 Markus Lohrey, Lukas Lück, Alexander Thumm, and Julio Xochitemol. Streaming algorithms for groups and semigroups, 2026. arXiv:2202.04060.

12 Markus Lohrey and Julio Xochitemol. Streaming in graph products. In *Proceedings of the 49th International Symposium on Mathematical Foundations of Computer Science, MFCS 2024*, volume 306 of *LIPIcs*, pages 71:1–71:17. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024. doi:10.4230/LIPIcs.MFCS.2024.71.

13 Anatoli I. Mal'cev. Nilpotent semigroups. *Fluchen.–Zap. Ivanovsk. Ped.Inst*, 4:107–111, 1953.

14 Charles F Miller III. Decision problems for groups – survey and reflections. In G. Baumslag and Charles F Miller III, editors, *Algorithms and classification in combinatorial group theory*, pages 1–59. Springer, 1992. doi:10.1007/978-1-4613-9730-4_1.

15 Peter Bro Miltersen, Noam Nisan, Shmuel Safra, and Avi Wigderson. On data structures and asymmetric communication complexity. *Journal of Computer and System Sciences*, 57(1):37–49, 1998. doi:10.1006/JCSS.1998.1577.

16 Walter D. Munn. Free inverse semigroups. *Proceedings of the London Mathematical Society*, s3-29(3):385–404, 1974. doi:10.1112/plms/s3-29.3.385.

17 Bernhard Hermann Neumann and Tekla Taylor. Subsemigroups of nilpotent groups. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 274(1356):1–4, 1963. doi:10.1098/rspa.1963.0110.

18 Oystein Ore. Linear equations in non-commutative fields. *Annals of Mathematics. Second Series*, 32(3):463–477, 1931. doi:10.2307/1968245.

19 Mario Petrich. *Inverse semigroups*. Wiley, 1984.

20 Gordon B. Preston. Representations of inverse semi-groups. *Journal of the London Mathematical Society*, 29:411–419, 1954. doi:10.1112/jlms/s1-29.4.411.

21 Elaine Render and Mark Kambites. Rational subsets of polycyclic monoids and valence automata. *Information and Computation*, 207(11):1329–1339, 2009. doi:10.1016/j.ic.2009.02.012.

22 Mark V. Sapir. *Combinatorial Algebra: Syntax and Semantics*. Springer, 2014. doi:10.1007/978-3-319-08031-4.

23 Lev M. Shneerson. Relatively free semigroups of intermediate growth. *Journal of Algebra*, 235(2):484–546, 2001. doi:10.1006/jabr.2000.8503.

24 Viktor V. Vagner. Generalized groups. *Doklady Akad. Nauk SSSR (N.S.)*, 84:1119–1122, 1952.