DECIDABILITY OF MEMBERSHIP PROBLEMS FOR FLAT RATIONAL SUBSETS OF $\mathrm{GL}(2,\mathbb{Q})$ AND SINGULAR MATRICES*

VOLKER DIEKERT†, IGOR POTAPOV‡, AND PAVEL SEMUKHIN§

Abstract. We consider membership problems for rational subsets of the semigroup of 2×2 matrices over $\mathbb{Q}$. For a semigroup M, the rational subsets $\mathrm{Rat}(M)$ are defined as the sets accepted by NFAs whose transitions are labeled by elements of M. In general, it is undecidable on inputs $m \in M$ and $R \in \mathrm{Rat}(M)$ whether m belongs to R. Therefore, we restrict our attention to the family $\mathrm{FRat}(M,S)$ of flat rational subsets of M over S, where S is a subsemigroup of M. It consists of finite unions of the form $g_0 L_1 g_1 \cdots L_t g_t$, where $L_i \in \mathrm{Rat}(S)$ and $g_i \in M$. Assuming that the membership for $\mathrm{Rat}(S)$ is decidable, we prove various results when the membership for $\mathrm{FRat}(M,S)$ is decidable.

If H is a subgroup of a group G, then we provide a rather general condition when $\mathrm{FRat}(G,H)$ is an (effective) relative Boolean algebra. This leads to one of our main results that the emptiness problem for Boolean combinations of sets in $\mathrm{FRat}(\mathrm{GL}(2,\mathbb{Q}),\mathrm{GL}(2,\mathbb{Z}))$ is decidable. It is possible that such a strong decidability result cannot be pushed any further for groups sitting between $\mathrm{GL}(2,\mathbb{Z})$ and $\mathrm{GL}(2,\mathbb{Q})$. To support this possibility, we prove the following dichotomy: if G is a finitely generated group such that $\mathrm{GL}(2,\mathbb{Z}) < G < \mathrm{GL}(2,\mathbb{Q})$, then either $G \cong \mathrm{GL}(2,\mathbb{Z}) \times \mathbb{Z}^k$ or G contains an extension of the Baumslag-Solitar group $\mathrm{BS}(1,q)$ of infinite index. It is open whether the membership for rational subsets is decidable in the latter case. For singular matrices, we will show that the membership problem for $\mathrm{FRat}(\mathbb{Q}^{2\times 2},S)$ is decidable in doubly exponential time, where S is the monoid generated by $\mathrm{GL}(2,\mathbb{Z}) \cup \{r \in \mathbb{Q} \mid r > 1\} \cup \{0, \left(\begin{smallmatrix}1 & 0\\ 0 & 0\end{smallmatrix}\right)\}$.

Key words. membership problem, finite automata, (flat) rational sets, general linear group, special linear group

MSC codes. 68Q45, 68W30

sec:intro

1. Introduction.

Many computational problems in matrix theory are inherently difficult to solve even for 2×2 matrices, and most them are undecidable in a higher dimension. One of these problems is the *semigroup membership problem* over some fixed commutative ring R: given a sequence $A, A_1, \ldots, A_m$ in $R^{n\times n}$, determine whether A belongs to the semigroup generated by the A_i's. In other words, determine whether there exist an integer $k \geq 1$ and $i_1, \ldots, i_k \in \{1, \ldots, m\}$ such that $A = A_{i_1} \cdots A_{i_k}$. Here and in the following $R^{n\times n}$ denotes the multiplicative monoid of $n \times n$ matrices with coefficients in R, and $\mathrm{GL}(n,R)$ denotes its group of units which consists of the matrices that are invertible in $R^{n\times n}$. We also use $\mathrm{SL}(n,R)$ to denote the subgroup of $\mathrm{GL}(n,R)$ of matrices with determinant one. The semigroup membership problem has been intensively studied since 1947 when Markov showed in [50] that this problem is undecidable for matrices in $\mathbb{Z}^{6\times 6}$. A special case is the *mortality problem* where the target matrix A is the the zero matrix. The mortality problem is undecidable for $\mathbb{Z}^{3\times 3}$ by Paterson [59]. For $\mathbb{Z}^{2\times 2}$ it is unknown whether the mortality problem is decidable. If A and the A_i's are in $\mathrm{GL}(n,R)$, then the *subgroup membership problem* asks whether A belongs to the matrix semigroup which is generated by the A_i's and A_i^{-1}'s. The subgroup membership problem is undecidable for $\mathrm{GL}(4,\mathbb{Z})$ by Mihailova [53]. It is unknown whether the subgroup membership is decidable for $\mathrm{GL}(3,\mathbb{Z})$. Even significantly restricted cases of these membership problems turn out to be undecidable for

Markov47 Paterson70 Mihailova58

*This manuscript substantially extends the following three conference papers [22, 61, 62]. DiekertPS20,PS_SODA,PS_MFCS2017

†Formale Methoden der Informatik, Universität Stuttgart, Germany (diekert@fmi.uni-stuttgart.de).

‡Department of Computer Science, Ashton Building, Ashton Street, University of Liverpool, UK (potapov@liverpool.ac.uk).

§Department of Computer Science, James Parsons Building, Liverpool John Moores University, UK (p.semukhin@ljmu.ac.uk).

high dimensional matrices over the integers [7, 43], and very few cases are known to be decidable, see [4, 8, 16]. The decidability of many of these problems remains open even for 2×2 matrices over integers [15, 18, 36, 42, 60].

A natural and important generalization of the semigroup membership problem is the *membership problem for rational subsets* of a semigroup M: given an element $a \in M$ and a rational subset $L \subseteq M$, decide whether a belongs to L. The family of *rational subsets* of M is denoted by $\mathrm{Rat}(M)$, and it has various equivalent definitions: homomorphic images of regular subsets of f.g. free semigroups, regular expressions over M, or acceptance by M-NFAs. An M-NFA is a non-deterministic finite automaton $\mathcal{A}$ whose transitions are labeled by elements in M. The label of a directed path is the directed product over its labels, and the accepted language is the set $L(\mathcal{A}) \subseteq M$ of labels of directed paths from initial to final states. Using M-NFAs allows for a graphical representation and is typically a more concise notation than using regular expressions. Thus, M-NFAs are our preferred way of defining sets in $\mathrm{Rat}(M)$.

It is well-known that the group $\mathrm{SL}(2,\mathbb{Z})$ has a free subgroup of rank 2 of index 12 by [55]. Hence, both $\mathrm{GL}(2,\mathbb{Z})$ and $\mathrm{SL}(2,\mathbb{Z})$ are finitely generated virtually free groups, and the families of their rational subsets form effective Boolean algebras [72, 74]. In particular, the membership problem for rational subsets in $\mathrm{GL}(2,\mathbb{Z})$ and in $\mathrm{SL}(2,\mathbb{Z})$ is decidable. This is no longer the case in higher dimensions. For example, in dimension four, $\mathrm{Rat}(\mathrm{SL}(4,\mathbb{Z}))$ is not even closed under finite intersections, and therefore it is not a Boolean algebra. However, this is still open for $\mathrm{SL}(3,\mathbb{Z})$, see Remark 3.7.

Two previous results that extended the decidability of the semigroup membership problem beyond $\mathrm{GL}(2,\mathbb{Z})$ are [61, 62]. The present paper pushes the frontier of decidability even further. First of all, we consider membership problems for 2×2 matrices over the rationals, whereas [61, 62] only dealt with integer matrices. Since the rational subset membership problem is known to be decidable for $\mathrm{GL}(2,\mathbb{Z})$, we focus on finitely generated subgroups G of $\mathrm{GL}(2,\mathbb{Q})$ which contain $\mathrm{GL}(2,\mathbb{Z})$. Also, in contrast to [61, 62], we give concrete complexity bounds: all complexities are in deterministic doubly exponential time (or better) for a natural binary encoding of the inputs.

In order to provide an essentially self-contained exposition of the main results, we combine a number of auxiliary results in Sections 2, 3, and 4. In Section 2, we characterize recognizable and rational sets in semigroups and highlight essential properties of (relative) Boolean algebras. In Section 3, we show so-called *Fatou property* for groups. It states that if G is a group and H is its subgroup, then $L \subseteq H$ and $L \in \mathrm{Rat}(G)$ implies that $L \in \mathrm{Rat}(H)$ (see Theorem 3.8 and Corollary 3.9). We also provide techniques for transferring results for rational subsets in group extensions of finite index (Corollary 3.12). In Section 4, we describe a cubic procedure for computing a *Smith normal form* of a non-zero matrix in $\mathbb{Q}^{2\times 2}$, and we discuss properties of *commensurators*, a notion borrowed from geometric group theory.

In Section 5, we prove our first main result which is Theorem 5.4. It states a dichotomy for a finitely generated (*f.g.* for short) subgroup G sitting strictly between $\mathrm{GL}(2,\mathbb{Z})$ and $\mathrm{GL}(2,\mathbb{Q})$. In the first case of the dichotomy, G is generated by $\mathrm{GL}(2,\mathbb{Z})$ and *finitely many* nonsingular central matrices $\left(\begin{smallmatrix} r & 0 \\ 0 & r \end{smallmatrix}\right)$. In this case, G is isomorphic to $\mathrm{GL}(2,\mathbb{Z}) \times \mathbb{Z}^k$ for $k \geq 1$, in which the membership problem for rational subsets is known to be decidable.

This is the best we can hope for groups sitting strictly between $\mathrm{GL}(2,\mathbb{Z})$ and $\mathrm{GL}(2,\mathbb{Q})$ in the general case. Indeed, our dichotomy states that if such a f.g. group G is not isomorphic to $\mathrm{GL}(2,\mathbb{Z}) \times \mathbb{Z}^k$, then G contains an extension of infinite index of a Baumslag-Solitar group $\mathrm{BS}(1,q)$ for some $q \geq 2$. The Baumslag-Solitar groups

BS(p,q) are defined by two generators a and t with the defining relation $ta^pt^{-1}=a^q$. They were introduced in [5] and have been widely studied since then. As we see in the proof of Theorem 5.4, BS$(1,q)$ cannot appear as a subgroup in $\mathrm{GL}(2,\mathbb{Z})\times\mathbb{Z}^k$, which implies that the two cases of the dichotomy are mutually exclusive. The group BS$(1,q)$ is metabelian, and subgroup membership is decidable for f.g. metabelian groups by [65]. Actually, a stronger result is known for Baumslag-Solitar groups: the membership problem for rational subsets of BS$(1,q)$ is decidable for all $q\geq 2$ by Cadilhac, Chistikov, and Zetzsche [14]. However, it is not clear how to generalize this result to extensions of BS$(1,q)$ of infinite index.

Motivated by the above results and observations, we introduce in Section 6 the notion of *flat rational sets* of a semigroup M over its subsemigroup S. We denote them by $\mathrm{FRat}(M,S)$. In the terminology of Schützenberger [71], $\mathrm{FRat}(M,S)$ is the *polynomial closure* of $\mathrm{Rat}(S)$ in M. More precisely, a subset $L\subseteq M$ is a flat rational set if and only if it can be written as a finite union of languages $L_0m_1L_1\cdots m_kL_k$, where the m_i's belong to M and $L_i\in\mathrm{Rat}(S)$ for $1\leq i\leq k$.

We are mainly interested in the study of $\mathrm{FRat}(\mathrm{GL}(2,\mathbb{Q}),S)$. Since $\mathrm{GL}(2,\mathbb{Q})$ is not finitely generated, the family $\mathrm{FRat}(\mathrm{GL}(2,\mathbb{Q}),S)$ is never a Boolean algebra because $\mathrm{GL}(2,\mathbb{Q})\notin\mathrm{FRat}(\mathrm{GL}(2,\mathbb{Q}),S)$. One of our main results about flat rational sets (Theorem 6.6) shows that under some natural assumptions on a group G and its subgroup H, the family $\mathrm{FRat}(G,H)$ forms an effective relative Boolean algebra (see Definition 2.12). As an application of this result, we will show that we can decide the emptiness of finite Boolean combinations of flat rational sets of $\mathrm{GL}(2,\mathbb{Q})$ over $\mathrm{GL}(2,\mathbb{Z})$ (Corollary 6.7). In Theorem 6.4, we provide an alternative intrinsic description of flat rational sets. In the rest of Section 6, we show a reduction of the membership problem for $\mathrm{FRat}(M,G)$ to that of $\mathrm{FRat}(M,H)$, where M is a monoid, G is a subgroup of its group of units, and H is a finite index subgroup of G (Theorem 6.9 and Corollary 6.10).

In the remaining three sections, we prove new decidability results for flat rational sets that contain matrices from $\mathbb{Q}^{2\times 2}$. In Section 7, we show that the membership problem for flat rational sets of $\mathrm{GL}(2,\mathbb{Q})$ over $\mathrm{GL}(2,\mathbb{Z})$ is decidable in exponential time (Theorem 7.1). We then prove various generalizations of this result, although with a worse complexity bound. For example, we show that the membership problem for $\mathrm{FRat}(\mathrm{GL}(2,\mathbb{Q}),S)$ is decidable in doubly exponential time, where $S=\mathrm{GL}(2,\mathbb{Z})\cup\{g\in\mathrm{GL}(2,\mathbb{Q})\mid|\det(g)|>1\}$ (Theorem 7.2).

If the target is a non-zero singular matrix, then we show in Section 8 that the membership problem for $\mathrm{FRat}(\mathbb{Q}^{2\times 2},S)$ is decidable in doubly exponential time for the monoid S which is generated by $\mathrm{GL}(2,\mathbb{Z})\cup\{r\in\mathbb{Q}\mid r>1\}\cup\{\left(\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right)\}$ (Theorem 8.2). However, we prove a better complexity bound for the *mortality problem*. Namely, we show that mortality for $\mathrm{FRat}(\mathbb{Q}^{2\times 2},S)$ is decidable in exponential time for the monoid S which is generated by $\mathrm{GL}(2,\mathbb{Z})\cup\mathbb{Q}\cup\{\left(\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right)\}$ (Theorem 8.1). In Section 9, we discuss potential directions for future research and list several open problems in this field.

2. Notation and preliminaries.

An *involution* of a set S is a mapping $x\mapsto\overline{x}$ such that $\overline{\overline{x}}=x$ for all $x\in S$. An involution of a semigroup $(S,\cdot)$ is an involution $^-$ of S such that $\overline{x\cdot y}=\overline{y}\cdot\overline{x}$. A monoid M is a semigroup $(M,\cdot)$ with a neutral element 1. Typically, we write commutative monoids like $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Z}/n\mathbb{Z}$ or $\mathbb{Q}$ with an additive notation. If we use a multiplicative notation, then 1 denotes the neutral element of a monoid. In particular, the empty word in free monoids is denoted by 1. It is also custom to write xy instead of $x\cdot y$. A *zero* in a semigroup $(M,\cdot)$ is an element 0 such

that $x \cdot 0 = 0 \cdot x = 0$ for all $x \in M$. If $(M, \cdot)$ is a semigroup with involution and with a zero 0, then $\overline{0} = 0$. If $(M, \cdot)$ is a monoid with involution, then $\overline{1} = 1$. If Γ is a set with an involution $^-$, then Γ^+ (resp., Γ^*) is a free semigroup (resp., free monoid) with involution, where the involution is defined on by Γ^* by extending it from Γ by using the law $\overline{uv} = \overline{v}\,\overline{u}$.

If G is a group, then it is a monoid with an involution $^-$ defined by $\overline{g} = g^{-1}$ for all $g \in G$. The identity mapping is an involution for commutative semigroups.

In commutative monoids without a zero-element, we might use an additive operation $+$, and then the neutral element is denoted as 0. There will be no risk of confusion.

For a subset $L \subseteq M$ of a semigroup M, the set L^+ denotes the subsemigroup of M generated by L. If M is a monoid, then the submonoid generated by L is $L^* = L^+ \cup \{1\}$. It is called the *Kleene-star* of L. We also use "f.g." as an abbreviation for "finitely generated". Hence, a semigroup (resp., monoid) is f.g. if it is a homomorphic image of a f.g. free semigroup (resp., monoid).

The group of *units* of M is the submonoid of invertible elements, denoted henceforth by $U(M)$. It is the set consisting of all $x \in M$ such that there is some $\overline{x} \in M$ with $\overline{x}x = x\overline{x} = 1$. If $x \in U(M)$, then we also write x^{-1} instead of $\overline{x}$. If x is a unit of M, then $x^{\mathbb{Z}}$ denotes the set $\{x^n \mid n \in \mathbb{Z}\}$, which is the subgroup generated by x. By $Z(M)$ we denote the *center* of M, that is, the set of elements which commute with all elements in M. We write $S \leq M$ if S is a subsemigroup of M, and $S < M$ if $S \leq M$ but $S \neq M$.

A subsemigroup I of a monoid M is an *ideal* if $M\,I\,M \subseteq I$. The empty set $\emptyset$ is an ideal. If M contains a zero 0, then $\{0\}$ is the least nonempty ideal. If an ideal I contains an element of $U(M)$, then $I = M$. Thus, if $I \neq M$, then I is contained in $M \setminus U(M)$. In general, $M \setminus U(M)$ is not an ideal (see an example in Remark 6.5).

A group G is called *virtually free* if it contains a free group of finite index. A group is finitely generated as a group if and only if it is finitely generated as a semigroup.

By $R^{n\times n}$ we denote the ring of $n \times n$ matrices over a commutative ring R, and we let $\det : R^{n\times n} \to R$ be the determinant function. The units of R are denoted by R^*. We view R as a subring of $R^{n\times n}$ by identifying $r \in R$ with the matrix $r = rI_n$, where I_n is the n-dimensional identity matrix. Hence, we may write $1 = I_n$ and $-1 = -I_n$. By $\mathrm{GL}(n, R)$ we mean the group of invertible matrices, that is, the matrices $g \in R^{n\times n}$ such that $\det(g) \in R^*$ is a unit. For $n \geq 2$, the center of $\mathrm{GL}(n, R)$ is $R^* = \{rI_n \mid r \in R^*\}$.

When we consider a matrix ring $R^{n\times n}$, a semigroup in $R^{n\times n}$ refers to a subsemigroup in the multiplicative monoid $(R^{n\times n}, \cdot)$.

By $\mathrm{SL}(n, R)$ we denote the *special linear group* $\det^{-1}(1) = \{g \in \mathrm{GL}(n, R) \mid \det(g) = 1\}$. It is a normal subgroup of $\mathrm{GL}(n, R)$. The structure of $\mathrm{SL}(2, \mathbb{Z})$ is well-understood.[1] The groups $\mathrm{SL}(2, \mathbb{Z})$ and $\mathrm{GL}(2, \mathbb{Z})$ are f.g. virtually free groups.

rem:glzv Newman62

Remark 2.1. It is shown in [55] that the *projective linear group* $\mathrm{PSL}(2, \mathbb{Z}) = \mathrm{SL}(2, \mathbb{Z})/\{\pm 1\}$ has a free subgroup of rank 2 and of index 6. Hence, $\mathrm{SL}(2, \mathbb{Z})$ has a free subgroup of rank 2 of index 12. Therefore, $\mathrm{GL}(2, \mathbb{Z})$ has a free subgroup of rank 2 which has index 24. Actually, the free subgroup of index 24 and rank 2 can be chosen to be the commutator subgroup of $\mathrm{SL}(2, \mathbb{Z})$. Possible generators for $\mathrm{SL}(2, \mathbb{Z})$ are the matrices $R = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 1 \end{smallmatrix}\right)$ of order 6 and $S = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$ of order 4. Possible generators for $\mathrm{GL}(2, \mathbb{Z})$ are S, R, and $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. In particular, since $\mathrm{SL}(2, \mathbb{Z})$ and $\mathrm{GL}(2, \mathbb{Z})$ are

[1] A discussion about $\mathrm{SL}(2, \mathbb{Z})$ including the computation of normal forms is, for example, in [21, Sec. 8.12]. edam16

generated by elements of finite order, none of the virtually free groups $\mathrm{PSL}(2,\mathbb{Z})$, $\mathrm{SL}(2,\mathbb{Z})$, or $\mathrm{GL}(2,\mathbb{Z})$ is free. ◇

sec:ratrec

2.1. Recognizable and rational sets in semigroups. Throughout this subsection $M = (M, \cdot)$ denotes a semigroup.[2] We recall some classical facts as they can be found with their proofs in the classical textbook of Eilenberg [26] or in the *q-Book* of Rhodes and Steinberg [64] as well as in [21].

def:rec

DEFINITION 2.2. *A subset $L \subseteq M$ belongs to the family of* recognizable sets $\mathrm{Rec}(M)$ *if there exists a homomorphism $\varphi : M \to N$ of M to a finite semigroup N such that $L = \varphi^{-1}(\varphi(L))$. We also say that φ (resp., N)* recognizes *L.*

Note that the canonical homomorphism of M to the trivial monoid $\{1\}$ recognizes $\emptyset$ and M.

prop:recbool

PROPOSITION 2.3. *If $\varphi_i : M \to N_i$ recognizes subsets $L_i \subseteq M$ for $i = 1, 2$, then the homomorphism $M \to N_1 \times N_2$, $m \mapsto \big(\varphi_1(m), \varphi_2(m)\big)$ recognizes $L_1 \cap L_2$ and $M \setminus L_i$ for $i = 1, 2$. In particular,* $\mathrm{Rec}(M)$ *is a Boolean algebra (in the sense of Definition 2.12 below).*

def:rat

DEFINITION 2.4. *The family of* rational sets $\mathrm{Rat}(M)$ *has the following definition using* regular *(aka* rational*) expressions. It is the least family such that:*

1. $|L| < \infty$, $L \subseteq M \implies L \in \mathrm{Rat}(M)$.
2. $L_1, L_2 \in \mathrm{Rat}(M) \implies L_1 \cup L_2$, $L_1 \cdot L_2$, *and* $L_1^+ \in \mathrm{Rat}(M)$.

Note that the definition of $\mathrm{Rat}(M)$ is intrinsic without reference to any generating set. Moreover, let M be a monoid and $L \in \mathrm{Rat}(M)$, then $L^* \in \mathrm{Rat}(M) \iff L^+ \in \mathrm{Rat}(M)$ because $L^* = L^+ \cup \{1\}$ and $L^+ = L \cdot L^*$.

rem:recipr

Remark 2.5. Let G be a group. Then $L \subseteq G$ is recognizable if and only if there is normal subgroup N of finite index and a finite subset $\{g_1, \ldots, g_k\} \subseteq G$ such that $L = \bigcup\{g_i N \mid 1 \le i \le k\}$. In particular, if G is infinite, then no finite subset of G is recognizable. A subgroup H belongs to $\mathrm{Rat}(G)$ if and only if H is f.g. by [2]. This does not hold for submonoids: the standard example is the additive group $\mathbb{Z} \times \mathbb{Z}$. It contains the submonoid $\{(m,n) \in \mathbb{N} \times \mathbb{N} \mid m = 0 \vee n \ge 1\} = \{(0,0)\} \cup \big((0,1) + \mathbb{N} \times \mathbb{N}\big)$ which is rational but not finitely generated, see [21, Sec. 8.9]. ◇

prop:freek

PROPOSITION 2.6. *Let $h : M \to M'$ be any homomorphism of semigroups. Then the following assertions hold.*

- *If $L' \in \mathrm{Rec}(M')$, then $h^{-1}(L') \in \mathrm{Rec}(M)$.*
- *If $L \in \mathrm{Rat}(M)$, then $h(L) \in \mathrm{Rat}(M')$.*
- *If $L \in \mathrm{Rec}(M)$ and $K \in \mathrm{Rat}(M)$, then $L \cap K \in \mathrm{Rat}(M)$.*
- Kleene's Theorem, [41]: *If M is either a f.g. free monoid or a f.g. free semigroup, then* $\mathrm{Rec}(M) = \mathrm{Rat}(M)$.
- *Let $L \in \mathrm{Rat}(M)$, then L is contained in a f.g. subsemigroup of M. In particular, $M \in \mathrm{Rat}(M)$ implies that M is finitely generated.*[3]
- *Let H be a subgroup of a group G. Then $H \in \mathrm{Rec}(G)$ if and only if the index $[G : H]$ is finite, see [2].*

For a f.g. free semigroup (or a f.g. free monoid) M, the family of *regular languages* $\mathrm{Reg}(M)$ is defined as $\mathrm{Reg}(M) = \mathrm{Rat}(M) = \mathrm{Rec}(M)$, where the last equality holds thanks to Kleene's Theorem stated in Proposition 2.6 above. Henceforth, if we use

[2] We call it M because in most of our cases the semigroup M is a monoid.

[3] McKnight's Theorem [52] is slightly more general. It states that M is finitely generated if and only if $M \in \mathrm{Rat}(M)$ if and only if $\mathrm{Rec}(M) \subseteq \mathrm{Rat}(M)$.

the term "regular language", then we always refer to a rational subset in some finitely generated free semigroup or monoid. For other monoids, we frequently have $\mathrm{Rec}(M) \neq \mathrm{Rat}(M)$. This happens, for example, if M is an infinite group. Moreover, if M contains a direct product $\{a,c\}^* \times \{b\}^*$, then, in contrast to $\mathrm{Rec}(M)$, the family $\mathrm{Rat}(M)$ is not closed under finite intersection, see for example [21, Ex. 7.21].

def:NFAm

DEFINITION 2.7. *Let M be a semigroup and $T \subseteq M$. A* nondeterministic finite automaton over T *(or a T-*NFA *for short) is a tuple $\mathcal{A} = (Q, \delta, I, F)$, where Q is a set of* states *with subsets $I, F \subseteq Q$ and $\delta \subseteq Q \times T \times Q$ is a finite set of* transitions*. The set I (resp., F) is called the set of* initial *(resp.,* final*) states. A transition $(p, s, q) \in \delta$ is also written as $p \xrightarrow{s} q$; and we say that $s \in T$ is its* label*. If M has a neutral element* 1*, then an* ε-transition[4] *is a transition with label $1 \in M$.*

A path (of transitions) of length $n \in \mathbb{N}$ is a sequence $q_0, a_1, q_1, \ldots, a_n, q_n$ such that $(q_{i-1}, a_i, q_i) \in \delta$ for all $1 \leq 1 \leq n$. Paths may also be depicted as

eq:trapath

$$q_0 \xrightarrow{a_1} q_1 \quad \cdots \quad \xrightarrow{a_{n-1}} q_{n-1} \xrightarrow{a_n} q_n. \tag{2.1}$$

If $q_0 \in I$ and $q_n \in F$, then we say that the path is accepting *for the element $a_1 \cdots a_n \in M$. The* accepted language $L(\mathcal{A})$ *is the set of all $m \in M$ for which there is a factorization $m = a_1 \cdots a_n$ that has an accepting path of length n.*

A subautomaton *of $\mathcal{A}$ is an NFA $\mathcal{A}' = (Q', \delta', I', F')$ such that $Q' \subseteq Q$ and $\delta' \subseteq \delta$, but there are no restrictions how to choose I' or F'.*

We follow the convention that if a path of length zero accepts $m \in M$, then M is a monoid and m is its neutral element. An NFA $\mathcal{A}$ is called *trim*, if every state belongs to some accepting path. Whenever convenient, we assume that $\mathcal{A}$ is trim. Note that $L(\mathcal{A})$ is contained in the subsemigroup of M which is generated by the finite set of labels of the transitions of $\mathcal{A}$. This holds whether or not $\mathcal{A}$ is trim.

prop:ratn

PROPOSITION 2.8. *Let $L \subseteq M$ be any subset. Then the following assertions are equivalent.*

- *The set L belongs to* $\mathrm{Rat}(M)$.
- *There is some M-NFA $\mathcal{A}$ such that $L = L(\mathcal{A})$.*
- *The set L is the image $\varphi(K)$ of a regular set $K \subseteq \Sigma^+$ under some homomorphism $\varphi : \Sigma^+ \to M$.*

The following lemma is used in the proof of Theorem 3.8 below. Its proof is straightforward.

lem:sil

LEMMA 2.9. *Let $\mathcal{A}$ be an M-NFA and $q_0 \xrightarrow{a_1} q_1 \quad \cdots \quad \xrightarrow{a_{n-1}} q_{n-1} \xrightarrow{a_n} q_n$ denote a path as in (2.1) with $n \geq 2$. Then adding (or removing) a transition $q_0 \xrightarrow{m} q_n$ with label $m = a_1 \cdots a_n$ does not change the accepted language.*

Note that adding transitions possibly makes accepting paths shorter, whereas removing transitions makes the size of the NFA smaller.

ec:inputsize

2.2. The input size of matrices and NFAs over matrices. We use the following notation. We let $\log(x) = \max\{1, \log_2(x)\}$. Let $f, g : \mathbb{N} \to \mathbb{R}_{\geq 0}$ be two functions with values in non-negative real numbers. As usual, we let $f \in \mathcal{O}(g)$ if there is some $k \in \mathbb{N}$ such that $f(n) \leq kg(n) + k$ for all $n \in \mathbb{N}$. Sometimes we measure complexities in *soft $\mathcal{O}$-notation* $\widetilde{\mathcal{O}}$. We write $f \in \widetilde{\mathcal{O}}(g)$ if $f \in \mathcal{O}(g \cdot \log^k(g))$ for some $k \in \mathbb{N}$. Thus, in soft $\mathcal{O}$-notation poly-logarithmic factors are neglected.

[4] The notation ε is used because it frequently appears in the literature on NFAs, where ε denotes the empty word.

The (bit-)complexity of an algorithm depends on the bit encoding of the input. When talking about complexity, we usually work with NFAs where the labels of transitions are $n \times n$ matrices over $\mathbb{Q}$, and therefore we define their size. There are two natural encodings: unary and binary. We will use both of them. For a matrix $A = (a_{ij})$ with integer entries $a_{ij} \in \mathbb{Z}$, we let $\| A \|_{\max} = \max\{|a_{ij}| \mid 1 \leq 1, j \leq n\}$. Given $m \in \mathbb{Q}^{n\times n}$, we assume that m is written as $m = p^{-1}A$ where p is the least positive integer such that $pm = A \in \mathbb{Z}^{n\times n}$. That is, p is 1 for the zero matrix, and otherwise p is the lcm of the denominators of non-zero entries in m. For such a representation $m = p^{-1}A$ as above we define its *unary size* as $\| m \|_{\max} = p\| A \|_{\max}$. It does not yield a matrix norm, however, for $n = 2$, we have:

{eq:Abs} (2.2) $$\| m_1 \cdots m_\ell \|_{\max} \leq 2^{\ell-1} \prod_{i=1}^{\ell} \| m_i \|_{\max}.$$

Since we are (mainly) interested in the bit complexity, we define *binary size* of m as $\| m \|_{\text{bin}} = \log(\| m \|_{\max})$. Hence, for $a, b, c, d \in \mathbb{Z}$ we have $\| \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \|_{\text{bin}} = \log_2(\max\{2, |a|, |b|, |c|, |d|\})$. In particular, $\| \left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right) \|_{\text{bin}} = \| \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right) \|_{\text{bin}} = \| \left(\begin{smallmatrix} 2 & 2 \\ 2 & 2 \end{smallmatrix}\right) \|_{\text{bin}} = 1$.

lem:bin LEMMA 2.10. *Let $m = m_1 \cdots m_\ell$ be a product of ℓ matrices in $\mathbb{Q}^{2\times 2}$ such that $\| m_i \|_{max} \leq 2^k$ for all $1 \leq i \leq \ell$. Then we have $\| m \|_{bin} \in \mathcal{O}(k\ell)$.*

Proof. This is a direct consequence of the inequality in (2.2). ∎

weightmatr DEFINITION 2.11. *Let $\mathcal{A} = (Q, \delta, I, F)$ be an $\mathbb{Q}^{2\times 2}$-NFA, that is, all labels of transitions of $\mathcal{A}$ are matrices in $\mathbb{Q}^{2\times 2}$. The* binary and unary sizes *$\| \mathcal{A} \|_{bin}$ and $\| \mathcal{A} \|_{max}$ of the NFA $\mathcal{A}$ are defined as follows:*

{eq:wnfa} (2.3) $$\| \mathcal{A} \|_{bin} = 1 + |Q| + |\delta| + \sum_{(p,m,q)\in\delta} \| m \|_{bin},$$

{eq:wnma} (2.4) $$\| \mathcal{A} \|_{max} = 1 + |Q| + |\delta| + \sum_{(p,m,q)\in\delta} \| m \|_{max}.$$

sec:NFAcc **2.3. Reductions and complexity classes.** We follow standard notation in complexity theory as it can be found for example in [58]. In particular, we assume the reader is familiar with the classes **P** and **NP** which denote the families of decision problems decidable on a Turing machine in deterministic (resp., nondeterministic) polynomial time.

Decision problems are encoded as subsets of Δ^* where Δ is a finite alphabet, for example, $\Delta = \{0, 1\}$. We define complexity classes via the notion of reductions, which are realized by (nondeterministic) Turing machines in the following sense. Let Γ and Σ be finite alphabets, and $f : \mathbb{N} \to \mathbb{N}$ be any function. Let T_f be a Turing machine with input alphabet Γ and a separate write-only output-tape, which is initially empty. We assume that T_f also satisfies the following property: on any input $u \in \Gamma^*$ of length n, every computation of T_f stops after at most $f(n)+1$ steps and produces some $v \in \Sigma^*$ on the output-tape. We write $v \in T_f(u)$ in this case. Note that by assumption we have $|v| \in \mathcal{O}(f(n))$.

Let $\mathcal{P} \subseteq \Gamma^*$ and $\mathcal{Q} \subseteq \Sigma^*$ be subsets. We say that $\mathcal{P}$ is DTIME(f) *(resp.,* NTIME(f)*) reducible* to $\mathcal{Q}$ if there exists a deterministic (resp., nondeterministic) Turing machine T_f with the above-mentioned properties such that $\forall u \in \Gamma^* : (u \in \mathcal{P} \iff \exists v \in \mathcal{Q} : v \in T_f(u))$. As a consequence, if $\mathcal{P}$ is DTIME(f) (resp., NTIME(f))

reducible to $\mathcal{Q}$ and if $\mathcal{Q}$ is DTIME(g) (resp., NTIME(g)) reducible to $\mathcal{R}$, then $\mathcal{P}$ is DTIME($g \circ f$) (resp., NTIME($g \circ f$)) reducible to $\mathcal{R}$.

A **P**-*reduction* (resp., **NP**-*reduction*) is a DTIME(f) (resp., NTIME(f)) reduction, where f is some polynomial. A **EXPTIME**-*reduction* (resp., **NEXPTIME**-*reduction*) is a DTIME(f) (resp., NTIME(f)) reduction, where f is a function of type 2^p where p is some polynomial.

If a problem $\mathcal{P}$ is DTIME(f) (resp., NTIME(f)) reducible to a singleton like $\{1\}$, then we say that $\mathcal{P}$ belongs to the complexity class DTIME(f) (resp., NTIME(f)). The classes $\mathbf{P} = \mathrm{DTIME}(n^{\mathcal{O}(1)})$ and $\mathbf{NP} = \mathrm{NTIME}(n^{\mathcal{O}(1)})$ are closed under **P** (resp., **NP**)-reductions.

sec:rba

2.4. Boolean algebras and relative Boolean algebras.

def:relBA

Definition 2.12. *Let U be any set and $\mathcal{B}$ be a family of subsets of U.*

- *We say that $\mathcal{B}$ is a* Boolean algebra*, if $\mathcal{B}$ is closed under finite union and complement.*
- *We say that $\mathcal{B}$ is a* relative Boolean algebra *if $\mathcal{B}$ is closed under finite union and relative complement: $L, K \in \mathcal{B} \implies L \setminus K \in \mathcal{B}$.*
- *We say that $\mathcal{B}$ is an* effective relative Boolean algebra *if each $K \in \mathcal{B}$ has an effective finite description, and there is an algorithm that, given descriptions of $K, L \in \mathcal{B}$, computes descriptions of $L \cup K$ and $L \setminus K$ and decides the emptiness of K.*

Every Boolean algebra is a relative Boolean algebra, and every relative Boolean algebra contains the empty set $\emptyset$. A relative Boolean algebra is closed under *nonempty* finite intersection. Indeed, $L \cap K = S \setminus ((S \setminus L) \cup (S \setminus K))$ where $S = L \cup K$. A relative Boolean algebra $\mathcal{B} \subseteq 2^U$ is a Boolean algebra if and only if $U \in \mathcal{B}$.

exs:classic

Examples 2.13. Let us list some classical examples of (relative) Boolean algebras.

1. If M is any f.g. semigroup, then the family of recognizable sets $\mathrm{Rec}(M)$ is an effective Boolean algebra. In particular, $\mathrm{Reg}(\Sigma^*)$ is an effective Boolean algebra if Σ is finite.
2. Let M_1 and M_2 be f.g. semigroups and let $M = M_1 * M_2$ denote their free product. If $\mathrm{Rat}(M_i)$ is an (effective) Boolean algebra for $i = 1, 2$, then $\mathrm{Rat}(M)$ is an (effective) Boolean algebra, see [66] and also [46] for a generalization.
3. Let M be a commutative semigroup. Rational sets in M are also called *semi-linear*: a semi-linear set is a finite union of *linear* sets, and a linear set (in additive notation) is a set of the form $c + \mathbb{N}d_1 + \cdots + \mathbb{N}d_t$, where $c, d_1, \ldots, d_t \in M$. The family of semi-linear sets forms a relative Boolean algebra by [27]. Using Presburger arithmetic, it can be shown that $\mathrm{Rat}(\mathbb{Z}^k)$ and $\mathrm{Rat}(\mathbb{N}^k)$ are actually effective Boolean algebras for all $k \in \mathbb{N}$. The decidability of Presburger arithmetic is a classical result due to Mojżesz Presburger [63].
4. Let $\mathbb{Q}$ be the additive group of the rational numbers. Then $\mathbb{Q}$ is not f.g. and every f.g. subgroup is isomorphic to $\mathbb{Z}$. As a consequence, $\mathrm{Rat}(\mathbb{Q})$ is an effective relative Boolean algebra, but not a Boolean algebra.
5. If G is a f.g. virtually free group, then the family of rational sets $\mathrm{Rat}(G)$ is an effective Boolean algebra. If G is an infinitely generated free group, then $\mathrm{Rat}(G)$ is a relative Boolean algebra, but not a Boolean algebra. The special case of f.g. free groups is due to Benois [10]. The extension to f.g. virtually free groups is in [33, 72, 74].

sec:tfl

3. The Fatou property and transfer results for rational subsets in groups.

Let G be a group and H be a subgroup. The aim of Section 3 is to prove Theorem 3.8.

It states that $L \subseteq H$ and $L \in \mathrm{Rat}(G)$ implies $L \in \mathrm{Rat}(H)$. This is called the Fatou property (see Remark 3.7 below for further discussion). This property does not hold for groups with respect to submonoids in general. Indeed, according to Remark 2.5 the f.g. group $\mathbb{Z} \times \mathbb{Z}$ contains a rational but not f.g. submonoid M. Since M is not f.g., we have $M \notin \mathrm{Rat}(M)$.

Theorem 3.8 holds without any assumption about G and its subgroup H: for example, the cardinalities the groups G, H, and the set of cosets G/H can be arbitrarily high. However, for effectiveness we need some restrictions. Therefore, we introduce the notion of *enumerable representation* in Definition 3.2. It is similar to the notion of "computably enumerable representation" as used, for example, in [76, Def. 1.1] for Boolean algebras, but differs from it in the sense that we do not require the equality relation to be computably enumerable. In particular, there are groups with an enumerable representation in which it is undecidable whether a given group element represents the neutral element.

We begin with Proposition 3.1. It gives a quasi-linear time complexity for deciding whether $L(\mathcal{A}) \subseteq \mathrm{SL}(2,\mathbb{Z})$ when $\mathcal{A}$ is a $\mathrm{GL}(2,\mathbb{Z})$-NFA. Its proof also serves as a warm-up example for the general proof strategy used later.

prop:GL2Z

PROPOSITION 3.1. *Let $\mathcal{A} = (Q, \delta, I, F)$ be a $\mathrm{GL}(2,\mathbb{Z})$-NFA of size $\| \mathcal{A} \|_{bin} = n$. Then we can construct in soft linear time with respect to n a $\mathrm{GL}(2,\mathbb{Z})$-NFA $\mathcal{A}'$ such that:*

- *$L(\mathcal{A}') = L(\mathcal{A})$.*
- *$\| \mathcal{A}' \|_{bin} \leq \| \mathcal{A} \|_{bin}$ and $\mathcal{A}'$ has at most $|Q|$ states.*
- *Moreover, $L(\mathcal{A}') \subseteq \mathrm{SL}(2,\mathbb{Z})$ if and only if all labels of transitions in $\mathcal{A}'$ have determinant 1. In particular, we can decide in time $\widetilde{\mathcal{O}}(n)$ whether $L(\mathcal{A}) \subseteq \mathrm{SL}(2,\mathbb{Z})$.*

Proof. In the first phase we trim $\mathcal{A}$, which can be done by standard algorithms in time $\widetilde{\mathcal{O}}(n)$ because $|Q| + |\delta| < n$. Henceforth, we assume without restriction that $\mathcal{A}$ is trim. In the second phase we mark all states in Q either by $+1$ or by -1. The corresponding states are called positive and negative respectively. All initial states are marked with $+1$, hence they are positive. As long as there is a transition $p \xrightarrow{m} q$, where p is marked and q is not marked, we mark q the same way as p if $\det(m) = 1$ and with the opposite marking of p if $\det(m) = -1$. After at most $|\delta|$ steps all states are marked. The marking procedure can also be implemented in time $\widetilde{\mathcal{O}}(n)$ using the fact that binary integers with n bits can be added and multiplied in $\widetilde{\mathcal{O}}(n)$ by the classical Schönhage-Strassen algorithm [68]. If we find a final state which is negative, then we have detected an accepted matrix with determinant -1. Hence $L(\mathcal{A})$ is not included in $\mathrm{SL}(2,\mathbb{Z})$, and we let $\mathcal{A}' = \mathcal{A}$ since $\mathcal{A}$ must have a transition whose label has determinant -1. We are done in this case.

Therefore, we may assume without restriction that all initial and final states are positive. In the third phase we relabel transitions using the matrix $s_{-1} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ of order two and with determinant -1. For that we consider all transitions $p \xrightarrow{m} q \in \delta$, one after another in some order. We either transform $p \xrightarrow{m} q$ into a transition $p \xrightarrow{m'} q$ such that $m' \in \mathrm{SL}(2,\mathbb{Z})$ or we detect that $L(\mathcal{A})$ is not included in $\mathrm{SL}(2,\mathbb{Z})$. Recall that $m = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with $a, b, c, d \in \mathbb{Z}$ and $\det(m) = \pm 1$. We make the following case distinction.

1. If both p and q are positive, then either we have $\det(m) = 1$ and we let $m' = m \in \mathrm{SL}(2,\mathbb{Z})$ or, if $\det(m) \neq 1$, we exit with an error message.
2. If p is positive and q is negative, then we have $\det(m) = -1$, and we let

$m' = ms_{-1} = \left(\begin{smallmatrix} a & -b \\ c & -d \end{smallmatrix}\right) \in \mathrm{SL}(2,\mathbb{Z})$ or, if $\det(m) \neq -1$, we exit with an error message.

3. If p is negative and q is positive, then either we have $\det(m) = -1$ and we let $m' = s_{-1}m = \left(\begin{smallmatrix} -a & -b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}(2,\mathbb{Z})$ or, if $\det(m) \neq -1$, we exit with an error message.
4. If p and q are negative, then either we have $\det(m) = 1$, and we let $m' = s_{-1}ms_{-1} = \left(\begin{smallmatrix} -a & b \\ c & -d \end{smallmatrix}\right) \in \mathrm{SL}(2,\mathbb{Z})$ or, if $\det(m) \neq 1$, we exit with an error message.

Since $\mathcal{A}$ is trim, an error message tells us that $\mathcal{A}$ accepts a matrix with determinant -1, which implies that $L(\mathcal{A}) \setminus \mathrm{SL}(2,\mathbb{Z}) \neq \emptyset$. To see this, recall the marking procedure. As noted above, we can assume without restriction that the initial and final states are positive. Since the NFA is trim, for every state $p \in Q$, there are matrices f_p and g_p such that f_p labels the path defined by the marking procedure from an initial state to p and g_p labels *any* path from p and to some final state. The marking procedure tells us that $\det(f_p)$ is positive if and only if the state p is positive. Assume that $\det(f_p) \neq \det(g_p)$. Then $\mathcal{A}$ accepts f_pg_p with $\det(f_p)\det(g_p) = -1$, and we know that $L(\mathcal{A}) \setminus L(\mathcal{A}) \neq \emptyset$. So, if $L(\mathcal{A}) \subseteq \mathrm{SL}(2,\mathbb{Z})$, then $\det(f_p) = \det(g_p)$ for every $p \in Q$. Now, consider any transition $p \xrightarrow{h} q$ in $\mathcal{A}$, then hg_q labels a path from p to some final state, and we conclude $f_phg_q \in L(\mathcal{A})$. Therefore, $L(\mathcal{A}) \subseteq \mathrm{SL}(2,\mathbb{Z})$ implies $\det(h) = \det(f_p)\det(g_q) = \det(f_p)\det(f_q)$. Now, $\det(f_p)\det(f_q) \neq \det(h)$ is exactly the situation when an error message occurs. Thus, an error message implies that $L(\mathcal{A})$ is not contained in $\mathrm{SL}(2,\mathbb{Z})$. In this case we stop and let $\mathcal{A}' = \mathcal{A}$.

Finally, assume there was no error message. In this case, the construction is finished, and it produces an $\mathrm{SL}(2,\mathbb{Z})$-NFA $\mathcal{A}' = (Q, \delta', I, F)$. It remains to verify $L(\mathcal{A}') = L(\mathcal{A})$. To see this, we first show that $L(\mathcal{A}) \subseteq L(\mathcal{A}') \subseteq \mathrm{SL}(2,\mathbb{Z})$. Consider a path π in $\mathcal{A}$ which begins in some initial state $\iota \in I$ and which ends in some final state $t \in F$ and which accepts $m \in \mathrm{GL}(2,\mathbb{Z})$. After the transformation we obtain a path π' having all labels in $\mathrm{SL}(2,\mathbb{Z})$.

We claim that the path π' accepts the same matrix m as before the transformation. Thus, the claim implies that $m \in \mathrm{SL}(2,\mathbb{Z})$ and $L(\mathcal{A}) \subseteq L(\mathcal{A}') \subseteq \mathrm{SL}(2,\mathbb{Z})$. We prove the claim by induction on the number of negative states on that path. Both states $\iota \in I$ and $t \in F$ are positive. If all states are positive, then the claim holds. Otherwise, the path π contains a subpath $p_0 \xrightarrow{m'_1} p_1 \xrightarrow{m'_2} \cdots p_{k-1} \xrightarrow{m'_k} p_k$ with $k \geq 2$ where p_0 and p_k are positive, but $p_1, \ldots, p_{k-1}$ are negative. This subpath corresponds to a path $p_0 \xrightarrow{m_1} p_1 \xrightarrow{m_2} \cdots p_{k-1} \xrightarrow{m_k} p_k$ in $\mathcal{A}$ such that $\det(m_1) = \det(m_k) = -1$ and $\det(m_i) = 1$ for $2 \leq i \leq k-1$. By definition of $\mathcal{A}'$ this yields $m'_1 = m_1s_{-1}$, $m'_k = s_{-1}m_k$, and $m'_i = s_{-1}m_is_{-1}$ for $2 \leq i \leq k-1$. Hence, $m_{1,k} = m_1 \cdots m_k = m'_1 \cdots m'_k \in \mathrm{SL}(2,\mathbb{Z})$. By Lemma 2.9, we can temporally add in $\mathcal{A}$ and $\mathcal{A}'$ the same transition $p_0 \xrightarrow{m_{1,k}} p_k$ between positive states without changing the accepted language. Note that adding these transitions does not affect the above procedure because both p_0 and p_k are positive and $\det(m_{1,k}) = 1$. However with the new transitions we obtain shorter accepting paths which visit less negative states. We are done by induction on the number of negative states on the accepting path π. Removing the newly added transition brings us back to the NFAs $\mathcal{A}$ and $\mathcal{A}'$. This shows the claim, which implies $L(\mathcal{A}) \subseteq L(\mathcal{A}')$.

To see that we also have the inclusion $L(\mathcal{A}') \subseteq L(\mathcal{A})$, consider an accepting path π' in $\mathcal{A}'$ that accepts a matrix $m' \in \mathrm{SL}(2,\mathbb{Z})$. Since there is a one-to-one correspondence between the transitions of $\mathcal{A}$ and $\mathcal{A}'$, we can construct and accepting path π in $\mathcal{A}$, which accepts some matrix m, such that π is transformed into π' by

the above procedure. It follows by the previous argument that $m = m'$ and hence $L(\mathcal{A}') \subseteq L(\mathcal{A})$. Figure 1 illustrates this transformation: the states p, v, w, t are positive and denoted as p_+, v_+, w_+, t_+ and q, u are negative and denoted as q_-, u_-. The labels before the transformation are in the upper line. The new labels are in the lower line.□

$$I \ni \iota_+ \overset{*}{\cdots\!\!\rightarrow} p_+ \xrightarrow{m_1} q_- \xrightarrow{m_2} u_- \xrightarrow{m_3} v_+ \xrightarrow{m_4} w_+ \overset{*}{\cdots\!\!\rightarrow} t_+ \in F$$

$$I \ni \iota_+ \overset{*}{\cdots\!\!\rightarrow} p_+ \xrightarrow{m_1 s_{-1}} q_- \xrightarrow{s_{-1} m_2 s_{-1}} u_- \xrightarrow{s_{-1} m_3} v_+ \xrightarrow{m_4} w_+ \overset{*}{\cdots\!\!\rightarrow} t_+ \in F$$

FIG. 1. *The upper line is a subpath of an accepting path π in $\mathcal{A}$. The lower line is the same subpath of π in $\mathcal{A}'$. Since $s_{-1}^2 = 1$ we have $m_1 \cdots m_4 = m_1' \cdots m_4' \in \mathrm{SL}(2, \mathbb{Z})$.* fig:SLZGLZ

def:en

DEFINITION 3.2. *We say that a semigroup S has an* enumerable representation *if there exist a finite alphabet Σ and a surjective mapping $\eta : W_S \to S$ such that the following holds:*

1. *The subset $W_S \subseteq \Sigma^+$ is decidable (as defined in formal language theory, [38]).*
2. *On input $u, v \in W_S$ we can compute a word $w \in W_S$ such that $\eta(u) \cdot \eta(v) = \eta(w)$.*

If S has an enumerable representation and $L \subseteq S$ is a subset, then we say that the membership problem for L *is decidable, if $\eta(L)^{-1} \subseteq \Sigma^+$ is a decidable language. We say that S has a* decidable word problem*, if on input $u, v \in W_S$ it is decidable whether $\eta(v) = \eta(w)$ in S.*

A pair (S, G), where S a semigroup and G is a subgroup of S, has an enumerable representation *if, in addition to the above, $\eta(G)^{-1} \subseteq W_S$ is decidable; and on input $w \in \eta(G)^{-1}$ we can compute some word $\widetilde{w} \in W_S$ such that $\eta(\widetilde{w}) = \eta(w)^{-1} \in G$. Moreover, if $S = M$ is a monoid, then we view $W_M \subseteq \Sigma^*$ by defining $\eta(1) = 1 \in M$.*

Finally, if $S = G$ is explicitly specified as a group, then we assume that for all $w \in W_G$ we can compute some word $\widetilde{w} \in W_G$ such that $\eta(\widetilde{w}) = \eta(w)^{-1} \in G$.

The Greek letter η used in the definition above stands for *evaluation*. In next proposition, there is also a letter ρ standing for *representation*.

rop:alpinv

PROPOSITION 3.3. *Let G be a group having an enumerable representation in the notation of Definition 3.2. If $\Gamma_+ \subseteq W_G$ is a finite subset, and $K \leq G$ is the subgroup generated by $\eta(\Gamma_+)$, then we can compute a finite set $\Gamma \subseteq W_G$ with an involution $^-$ and a mapping $\rho : \Gamma_+ \to \Gamma$ such that $\eta(\Gamma)$ generates K, $\eta(\rho(u)) = \eta(u)$ for all $u \in \Gamma_+$, and $\eta(\overline{u}) = \eta(u)^{-1}$ for all $u \in \Gamma$.*

Proof. Choosing a linear order on Σ, defines a shortlex-ordering $\leq$ on Σ^*. Therefore, $\leq$ is also a linear order on $W_G \subseteq \Sigma^*$. Given a word $w \in W_G$, we denote by $\widetilde{w}$ the word which is computed on input w and which satisfies $\eta(\widetilde{w}) = \eta(w)^{-1} \in G$. (The existence of such an algorithm is part of Definition 3.2.) In the beginning, we let $\rho(u) = u$ and $\overline{u} = \widetilde{u}$, for all $u \in \Gamma_+$, and let $\Gamma = \rho(\Gamma_+) \cup \{\overline{u} \mid u \in \rho(\Gamma_+)\}$, but Γ, ρ, and $^-$ will change dynamically. Note that initially $\rho = \mathrm{id}_{\Gamma_+}$, $\Gamma = \Gamma_+ \cup \overline{\Gamma_+}$, and $\eta(\Gamma)$ generates the subgroup K. Later we change ρ, and we extend $^-$ to an involution on Γ. During the construction, we will preserve the following invariants: $\eta(\Gamma)$ generates K, $\eta(\overline{u}) = \eta(u)^{-1}$ for all $u \in \rho(\Gamma_+)$, and $\eta(\rho(u)) = \eta(u)$ for all $u \in \Gamma_+$.

Next, we make $^-$ injective on $\rho(\Gamma_+)$. Namely, if it happens that there are $w_1, w_2 \in \rho(\Gamma_+)$ with $\overline{w_1} = \overline{w_2}$ and $w_1 < w_2$, then we remove w_2 from $\rho(\Gamma_+)$ and replace w_2

everywhere by w_1. For example, if we had $\overline{u} = w_2$ for some u, then now we have $\overline{u} = w_1$. If we had $\rho(u) = w_2$ for some u, then now we have $\rho(u) = w_1$. In particular, both $\rho(\Gamma_+)$ and Γ become smaller. Note that this modification preserves the above invariants because $\overline{w_1} = \overline{w_2}$ implies that $\eta(w_1) = \eta(w_2)$.

Since this procedure stops in a finite number of steps, the mappings $u \mapsto \overline{u}$ and ρ are computable. Finally, we extend $^-$ from $\rho(\Gamma_+)$ to an involution of Γ in a natural way: namely, if $v = \overline{u}$ for some $u \in \rho(\Gamma_+)$, then we define $\overline{v} = u$. Clearly, $\eta(\overline{u}) = \eta(u)^{-1}$ for all $u \in \Gamma$. ∎

rem:recen

Remark 3.4. Every finitely generated semigroup G has an enumerable representation by choosing a surjective homomorphism $\eta : \Sigma^+ \to G$ where Σ is finite and letting $W_G = \Sigma^+$. For f.g. monoids, the decidability of the word problem does neither depend on Σ nor on the homomorphism η. In this case, decidability of the word problem as defined in Definition 3.2 coincides verbatim with the standard definition for f.g. monoids as used for example in [12].

bo93springer

Note that one can construct a finitely presented semigroup with an undecidable word problem, see Markov [51]. It is considered to be the first undecidability result in algebra. The corresponding result for groups is more difficult. It was shown first in independent papers of Novikov and Boone [56, 13].

Markov

nov55,boone59

The group $\mathrm{GL}(n, \mathbb{Q})$ has an enumerable representation; and its word problem is decidable. It is also clear that $\mathrm{GL}(n, \mathbb{Q})$ is not finitely generated for $n \geq 1$. ◇

lem:finin

LEMMA 3.5. *Let H be a finite index subgroup of G. Then*

eq:finexHow

$$\{L \subseteq H \mid L \in \mathrm{Rat}(G)\} = \{L \cap H \mid L \in \mathrm{Rat}(G)\}. \tag{3.1}$$

Proof. The inclusion $\subseteq$ is trivial. The other inclusion is clear by Proposition 2.6 since $[G : H] < \infty$ implies $H \in \mathrm{Rec}(G)$. ∎

Lemma 3.5 cannot be extended to the case where H has infinite index. For example, the extension fails as soon G does not have the so-called Howson property. The *Howson property* states that the intersection of two f.g. subgroups is finitely generated.[5]

Howson54

The free groups satisfy the Howson property [39]. The following lemma shows that this is not the case for a direct product of nontrivial free groups. It is well-known and follows easily from [10] and standard results in trace theory [23]. For convenience, we provide a proof below.

berg99

dr95

lem:dr95H

LEMMA 3.6. *The direct product $G = F(a, b) \times F(c)$ does not satisfy the Howson property. Here $F(a, b)$ and $F(c)$ denote free groups of rank 2 and rank 1, respectively.*

More precisely, let H be the subgroup of in G which is generated by (a, c) and $(b, 1)$, and let L be the subgroup of G generated by $(a, 1)$ and (b, c). Then $K = H \cap L$ is not rational. (In particular, it is not finitely generated by Remark 2.5.)

Proof. By contradiction assume $K \in \mathrm{Rat}(G)$. Choose any set of monoid generators of $F(a, b)$ which includes the letters a and b. Let h be the canonical inclusion of the free monoid $\{a, b\}^*$ into $F(a, b)$. The family $\mathrm{Rat}(F(a, b))$ is closed under intersection by [10]. Hence, $R = \pi(K) \cap a^*b^* \in \mathrm{Rat}(F(a, b))$. By the second item of Proposition 2.6 there is a regular set $K \in \mathrm{Reg}(\{a, b\}^*)$ such that $h(K) = R$. A direct calculation shows $\{a^n b^n \mid n \in \mathbb{N}\} \subseteq K \subseteq \{w \in \{a, b\}^* \mid |w|_a = |w|_b\}$. But there is

berg99

[5] If G is not Howson, consider f.g. subgroups L and H such that $K = L \cap H$ is not f.g. Hence $K \subseteq H$ but $K \notin \mathrm{Rat}(G)$; thus Equation (3.1) fails.

no such regular set K because otherwise $\{a^n b^n \mid n \in \mathbb{N}\} = K \cap a^*b^* \in \mathrm{Reg}(\{a,b\}^*)$, which is not regular, see [38]. A contradiction. ∎

Remark 3.7. Lemma 3.6 also implies that $\mathrm{Rat}(\mathrm{SL}(4,\mathbb{Z}))$ is not closed under finite intersection because $\mathrm{SL}(4,\mathbb{Z})$ contains the product $\mathrm{SL}(2,\mathbb{Z}) \times \mathbb{Z}$ and $\mathrm{SL}(2,\mathbb{Z})$ contains a free group of rank 2. On the other hand, it is still open whether $\mathrm{SL}(3,\mathbb{Z})$ satisfies the Howson property, see [48]; and we also do not know whether $\mathrm{Rat}(\mathrm{SL}(3,\mathbb{Z}))$ is closed under finite intersection. ◇

Let M be a monoid and $N \leq M$ be a submonoid. Following the French school around Schützenberger, we say that (M,N) satisfies the *Fatou property*[6] if

$$\mathrm{Rat}(N) = \{L \in \mathrm{Rat}(M) \mid L \subseteq N\} \tag{3.2}$$

Even for f.g. commutative monoids the Fatou property does not hold in general. To see this let $M = \mathbb{N} \times \mathbb{N}$. Then $N = (0,0) \cup \{(m,n) \in M \mid m \geq 1\}$ is easily seen to be submonoid of M which is not finitely generated (see also Remark 2.5). Hence $N \notin \mathrm{Rat}(N)$. On the other hand, we have $N \in \mathrm{Rat}(M)$ because in the additive notation $\{(m,n) \in M \mid m \geq 1\}$ is the linear set $(1,0) + \mathbb{N}(1,0) + \mathbb{N}(0,1)$, and hence N is a semi-linear subset[7] of M.

Thus, we need some restrictions either on M or N, or both. In [11, 31] it is stated that the Fatou property holds for groups by similar arguments as given in [2]. However, the authors do not give any proofs. The first published proof (we are aware of) was given by Herbst using the notion of *star height*, see [37]. An immediate corollary of Theorem 3.8 is that the Fatou property holds for groups. (In order to have a reference, we state this explicitly in Corollary 3.9.) Our proof of Theorem 3.8 uses NFAs which is important for our complexity results. Under the assumption that G is f.g. and that the index $[G:H]$ is finite the Fatou property for groups has been shown in [33, 74] and, for f.g. virtually free groups, in [72]. To the best of our knowledge, our proof that works directly with NFA's without increasing their sizes was first published in the conference paper [22]. We apply it to $\mathrm{SL}(2,\mathbb{Z})$ and $\mathrm{GL}(2,\mathbb{Q})$. Here, $\mathrm{GL}(2,\mathbb{Q})$ is not finitely generated, and the index $[\mathrm{GL}(2,\mathbb{Q}) : \mathrm{SL}(2,\mathbb{Z})]$ is infinite.

THEOREM 3.8. *Let $\mathcal{A}$ be a G-NFA and K be the subgroup of a group G which is generated by $L(\mathcal{A}) \subseteq G$. Then there is a trim K-NFA $\mathcal{A}'$ which accepts $L(\mathcal{A})$ such that the number of states and transitions is bounded by that of $\mathcal{A}$. Moreover, if G has an enumerable representation and if the labels of $\mathcal{A}$ are given by words in the decidable set W_G as in Definition 3.2, then the construction of $\mathcal{A}'$ is effective.*

Proof. First, we trim the automaton $\mathcal{A}$. Therefore, from now on, we assume that every state p (and hence every transition) is on some accepting path. There is a finite set $\Gamma \subseteq G$ such that for every transition $p \xrightarrow{g} q$ we have both g and g^{-1} in Γ. For $g \in \Gamma$ we define $\overline{g}$ by $\overline{g} = g^{-1}$. Thus, Γ is finite set with involution. The inclusion $\eta : \Gamma \subseteq G$ induces a homomorphism $\psi : \Gamma^* \to G$ from the free monoid Γ^* with involution onto K. Recall that the involution on a word $a_1 \cdots a_k$ with $a_i \in \Gamma$ is defined by $\overline{a_k} \cdots \overline{a_1}$. Thus, ψ respects the involution.

In case when G has an enumerable representation, we know by assumption that all labels of $\mathcal{A}$ belong to a decidable set $W_G \subseteq \Sigma^*$ as in Definition 3.2. We let $\Gamma_+ \subseteq W_G$ be the finite set of labels $u \in W_G$ which appear on some transition $p \xrightarrow{u} q$.

[6] The notation was coined for groups in [11] as an analogue of a result of Fatou who published in 1904 that a rational series of $\mathbb{Q}[x]$ whose coefficients are all integers is a rational series of $\mathbb{Z}[x]$.

[7] The definition of semi-linear set is in the third item of Examples 2.13.

By Proposition 3.3, there is a computable mapping ρ from Γ_+ to some finite subset $\Gamma \subseteq W_G$ with involution $^-$ such that $\eta(\Gamma)$ generates the the same subgroup as $\eta(\Gamma_+)$, $\eta(\overline{u}) = \eta(u)^{-1}$ for all $u \in \Gamma$, and $\eta(\rho(u)) = \eta(u)$ for all $u \in \Gamma_+$. Thus, as in the case when $\Gamma \subseteq G$ above, η can be extended to a homomorphism $\psi : \Gamma^* \to G$ from the free monoid Γ^* to G which respects the involution. Using ρ, we relabel all transitions in $\mathcal{A}$ by letters in Γ.

Therefore we can use a unified notation for both cases $\Gamma \subseteq G$ and $\Gamma \subseteq W_G$. In particular, even for $\Gamma \subseteq G$ we write $\psi(L(\mathcal{A}))$ rather than $L(\mathcal{A})$. That is, we consider $\mathcal{A}$ as an automaton over the free monoid Γ^* rather than G since every sequence $g_1, \dots, g_k$ of k elements in G has a natural evaluation $g_1 \cdots g_k$ in G which coincides with $\psi(g_1 \cdots g_k)$. So, in our notation, K is the subgroup generated by $\psi(L(\mathcal{A}))$.

Since $\mathcal{A}$ is trim, for every state p of $\mathcal{A}$ there are shortest words $u_p, v_p \in \Gamma^*$ such that u_p is the label of a path from an initial state to p and v_p is the label of a path from p to a final state. Since $K = \langle \psi(L(\mathcal{A})) \rangle$ we have $\psi(u_p v_p) \in K$ for all $p \in Q$. We also have $\psi(\overline{u_p}) = \psi(u_p)^{-1}$ and $\psi(v_p) \in \psi(\overline{u_p})K$. Therefore the left-coset of $\psi(v_p)$ in G/K is unique: it depends on p and not on the choice of v_p. Thus, we can write $\psi(v_p) \in \psi(r_p)K$ with $r_p = \overline{u_p}$ for $p \notin I \cup F$. For $p \in I \cup F$, we can choose $r_p = \overline{r_p} = 1$, where 1 denotes the empty word in Γ^*. This choice is possible since for $p \in I$ (resp., $p \in F$) we have $u_p = 1$ (resp., $v_p = 1$), and hence $\psi(v_p) \in K$.

Next, we make Γ possibly larger such that Γ contains two letters r_p and $\overline{r_p}$ for all $p \notin I \cup F$. We define (respectively redefine if necessary) η for r_p and $\overline{r_p}$ by $\eta(r_p) = \psi(u_p)^{-1}$ and $\eta(\overline{r_p}) = \psi(u_p)$. As above, η induces a homomorphism $\psi : \Gamma^* \to G$ respecting the involution.

Having defined the coset representatives r_p, we transform the NFA $\mathcal{A}$ into an NFA $\mathcal{B}$ as follows. The state space of $\mathcal{B}$ is defined as the union $Q \cup \overline{Q}$ where $\overline{Q}$ is a disjoint copy of Q. We denote the copy of $p \in Q$ by $\overline{p} \in \overline{Q}$.

The transitions in $\mathcal{B}$ are defined in two steps. In the first step, we introduce for each $p \in Q$ an additional outgoing transition $p \xrightarrow{r_p} \overline{p}$ and an additional incoming transition $\overline{p} \xrightarrow{\overline{r_p}} p$. Since $\psi(r_p)\psi(\overline{r_p}) = 1 \in G$ for all $p \in Q$, this does not change the accepted language by Lemma 2.9. Recall that $r_p = \overline{r_p} = 1$ for all $p \in I \cup F$. Thus, an ε-transition (that is, a transition with label 1) leads from p to $\overline{p}$ and from $\overline{p}$ to p for all $p \in I \cup F$. Therefore we do not change the accepted language by enlarging the sets of initial and final states by $I \cup \overline{I}$ and $F \cup \overline{F}$, respectively.

In the second step, we consider every transition $p \xrightarrow{a} q \in \delta$ with $p, q \in Q$ in some order. Since $\varphi(u_p v_p) \in K$, $\varphi(u_p a v_q) \in K$, and $\psi(v_p)K = \psi(r_p)K$, we have $\psi(a v_q)K = \psi(v_p)K = \psi(r_p)K$. We also have $\psi(v_q)K = \psi(r_q)K$, and therefore $K = \psi(\overline{r_p})\psi(a)\psi(v_q)K = \psi(\overline{r_p} a r_q)K$, which is equivalent to $\psi(\overline{r_p}\, a\, r_q) \in K$.

Hence, defining $h = \overline{r_p}\, a\, r_q \in \Gamma^*$, we obtain $\psi(h) \in K$. Having this, we introduce for $\mathcal{B}$ a new transition $\overline{p} \xrightarrow{h} \overline{q}$. See Figure 2 for a visualization of the NFA $\mathcal{B}$.

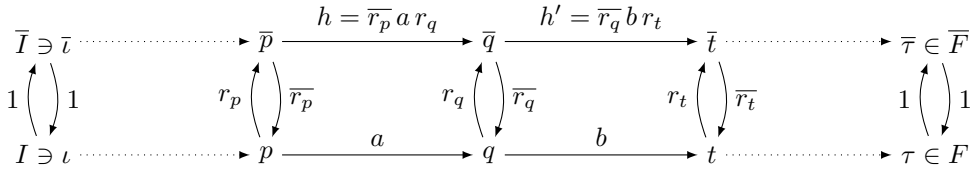

FIG. 2. *The construction of the NFA $\mathcal{B}$ yields $h = \overline{r_p}\, a r_q \in \psi^{-1}(H)$ and $h' = \overline{r_q}\, b r_t \in \psi^{-1}(H)$.* fig:cB

We claim that $\psi(L(\mathcal{A})) = \psi(L(\mathcal{B}))$. The inclusion $\psi(L(\mathcal{A})) \subseteq \psi(L(\mathcal{B}))$ is trivial. For the other direction we use Lemma 2.9: since $\psi(h) = \psi(\overline{r_p}\, a r_q)$ in G, we did not change $\psi(L(\mathcal{A}))$.

Finally, we define the NFA $\mathcal{A}'$ by removing from $\mathcal{B}$ all states in Q (together with the incident transitions). In particular, all the remaining transitions are of the form $\overline{p} \xrightarrow{h} \overline{q}$ with $\psi(h) \in K$, the set of initial states is $\overline{I}$, and the set of final states is $\overline{F}$. We can think of $\mathcal{A}'$ as a disjoint copy of $\mathcal{A}$ where a transition $p \xrightarrow{a} q$ with $a \in \Gamma$ has been replaced in its copy by the transition $\overline{p} \xrightarrow{h} \overline{q}$ with label $h = \overline{r_p}\, a\, r_q$ such that $\psi(h) \in K$. Note that the construction of $\mathcal{A}'$ is effective if G has an enumerable representation and if the labels of $\mathcal{A}$ are in the decidable set W_G.

Since we already know that $\psi(L(\mathcal{A})) = \psi(L(\mathcal{B}))$, it remains to show $\psi(L(\mathcal{A}')) = \psi(L(\mathcal{B}))$. For this we use a dual construction. Note that if we define $\overline{\overline{p}} = p$ for all $\overline{p} \in \overline{Q}$, then $Q \cup \overline{Q}$ becomes a set with involution. Now we perform the same construction as above starting with $\mathcal{A}'$ (which is the upper line in Figure 2) but replacing p with $\overline{p}$, r_p with $\overline{r}_p$, etc. In particular, we will have a transition $p \xrightarrow{r_p\, h\, \overline{r_q}} q$ between p and q in Figure 2 instead of $p \xrightarrow{a} q$. Let $\mathcal{B}'$ be the resulting automaton. Since $\psi(r_p h \overline{r_q}) = \psi(a)$, we conclude that $\psi(L(\mathcal{B}')) = \psi(L(\mathcal{B}))$. On the other hand, by Lemma 2.9, we have $\psi(L(\mathcal{A}')) = \psi(L(\mathcal{B}'))$. This completes the proof of the theorem. □

cor:silva

COROLLARY 3.9. *Let G be a group with a subgroup H and $\mathcal{A}$ be a G-NFA with n states and m transitions such that $L(\mathcal{A}) \subseteq H$. Then there is a (trim) H-NFA $\mathcal{A}'$ with at most n states and at most m transitions such that $L(\mathcal{A}') = L(\mathcal{A})$. In particular, the groups satisfy the* Fatou property *of Equation (3.2): that is, we have $\{L \subseteq H \mid L \in \mathrm{Rat}(G)\} = \mathrm{Rat}(H)$.*

rem:NyBrodda

NybergBrodda2023

Remark 3.10. Nyberg-Brodda has recently shown in [57] that there is finitely generated (and *context-free*) monoid M such that its group of units is a rational but not finitely generated. Thus, f.g. monoids fail to satisfy the Fatou property with respect to subgroups. In his example there is a set of three generators $\{a, b, c\}$. The defining relations are $\{(ab^i c)^2 = 1 \mid i \in \mathbb{N}\}$. The resulting semi-Thue system is easily seen to be confluent and Noetherian. It follows that M is not Dedekind-finite (since $acac = 1$ but $1 \neq caca$) and its group of units is the rational submonoid $F = (ab^*c)^*$. Thus, $U(M)$ is the free product $F = *_{n\in\mathbb{N}} \mathbb{Z}/2\mathbb{Z}$, which is not f.g. ◇

cor:silva

COROLLARY 3.11. *Let G have an enumerable representation and H be a subgroup such that the membership problem for H is decidable. Then, we can decide for a G-NFA $\mathcal{A}$, whose labels are given by words in the set W_G in the notation of Definition 3.2, whether $L(\mathcal{A}) \subseteq H$.*

Proof. Let K be the subgroup of G generated by $L(\mathcal{A})$. We apply Theorem 3.8 to effectively construct a K-NFA $\mathcal{A}'$ such that $L(\mathcal{A}') = L(\mathcal{A})$ and where the transitions in $\mathcal{A}'$ have labels in W_G such that their image in G generates the subgroup K. Therefore, $L(\mathcal{A}) \subseteq H$ is decidable because the membership problem for H is decidable, and hence we can check whether the labels of transitions of $\mathcal{A}'$ belong to H. □

cor:fine

COROLLARY 3.12. *Let G be a f.g. group and H a subgroup of finite index. Then* $\mathrm{Rat}(H)$ *is a Boolean algebra if and only if* $\mathrm{Rat}(G)$ *is a Boolean algebra. Moreover, the membership problem for rational subsets of H is decidable if and only if it is decidable for* $\mathrm{Rat}(G)$.

Proof. It is well-known and easy to see that G is f.g. if and only if H is f.g. Therefore both groups G and H are f.g. In particular, they have enumerable representations, which allows us to apply the effectiveness condition in Theorem 3.8.

Assume that $\mathrm{Rat}(G)$ is a Boolean algebra. Let us show that $\mathrm{Rat}(H)$ is a Boolean algebra, too. Note that $\mathrm{Rat}(H) \subseteq \mathrm{Rat}(G)$; and we have $H \in \mathrm{Rat}(G)$ since H is finitely generated. Thus, for every $R \in \mathrm{Rat}(H)$, we have $H \setminus R \in \mathrm{Rat}(G)$, and hence $H \setminus R \in \mathrm{Rat}(H)$ by Corollary 3.9. This shows that $\mathrm{Rat}(H)$ is a Boolean algebra. If the membership for rational sets of G is decidable, then the membership for rational sets of H is decidable because $\mathrm{Rat}(H) \subseteq \mathrm{Rat}(G)$.

For the other direction, assume $\mathrm{Rat}(H)$ is a Boolean algebra. In order to show that $\mathrm{Rat}(G)$ is a Boolean algebra let $R \in \mathrm{Rat}(G)$. We have to show that $G \setminus R \in \mathrm{Rat}(G)$. Since the index $[G : H]$ is finite, there is subgroup $N \leq H$ which is normal in G. (Actually, $N = \bigcap\{gHg^{-1} \mid g \in G\}$ and the intersection is finite since $[G : H] < \infty$.) Let $\varphi : G \to G/N$ be the canonical homomorphism. Then φ recognizes H.

Let $\{r_1, \ldots, r_k\} \subseteq G$ be representatives of left cosets of H, where $k = [G : H]$, such that for each $g \in G$ there is exactly one r_g with $g \in r_g H$. Thus, $g \notin R$ if and only if $r_g^{-1} g \in H \setminus r_g^{-1} R$. In other words, $G \setminus R = \bigcup_{i=1}^{k} r_i (H \setminus r_i^{-1} R) = \bigcup_{i=1}^{k} r_i (H \setminus (r_i^{-1} N \cap N))$.

By Proposition 2.6 we have $r_i^{-1} R \cap H \in \mathrm{Rat}(G)$ because H is recognizable. By Corollary 3.9 we have $r_i^{-1} R \cap H \in \mathrm{Rat}(H)$. Since $\mathrm{Rat}(H)$ is a Boolean algebra, $H \setminus (r_i^{-1} R \cap H) \in \mathrm{Rat}(H)$, and we conclude that $G \setminus R \in \mathrm{Rat}(H)$.

It remains to show that the membership for $\mathrm{Rat}(G)$ is decidable if the membership for $\mathrm{Rat}(H)$ is decidable. Since G is f.g. there is some finite generating subset $\Gamma \subseteq G \setminus \{1\}$ such that $\Gamma = \Gamma^{-1}$. Thus, every word in $w \in \Gamma^*$ has a natural interpretation in the group G. The *Schreier graph*, also called the *coset graph*, has been defined in [69] for H with respect to Γ. It is a directed graph where the set of vertices V is the finite set of all left cosets: $V = \{gH \mid g \in G\}$. The directed edges are labeled by generators and defined as $gH \xrightarrow{a} agH$ for all $a \in \Gamma$ and $gH \in V$. Thus, the out-degree of each vertex is $|\Gamma|$. We construct the Schreier graph of H by exhaustive search. The construction yields rooted tree T where the nodes $V(T)$ are words in Γ^*. We begin with $T = \{1\}$ where 1 the empty word representing the coset H. During the process some nodes without *children* will become a leaf in the final tree. For that we define a subset $L(T) \subseteq V(T)$ which initially is empty. The invariant is that all nodes in $L(T)$ are leaves.

Next, while $V(T) \setminus L(T) \neq \emptyset$ we repeat the following loop.

1. Choose any node $g \in V(T) \setminus L(T)$.
2. For each $a \in \Gamma$ (in some order) consider the word $ag \in \Gamma^*$, and decide whether $agH = hH$ for some $h \in V(T)$. (This is possible because the membership in $h^{-1}ag \in H$ is decidable.) If for all $h \in V(T)$ we have $h^{-1}ag \notin H$, then the word ag represents the coset agH (which was not represented in $V(T)$ so far) and we add ag to $V(T)$ as a child of g.
3. If g is still without any child by the previous step, then g becomes a leaf in the tree T. That is, we update $L(T)$ redefining it as $L(T) \cup \{g\}$.

Let us show that the algorithm terminates. The first observation is that $V(T)$ grows as long as $|V(T)|$ is less than the index of H in G. Thus, the algorithm reaches a point where $|V(T)| = [G : H]$. Having this, all nodes without children become leaves because Γ is finite. At this point the algorithm stops with $V(T) = \{r_1, \ldots, r_k\} \subseteq G$. The representatives $r_i \in V(T)$ are written as words in Γ^*.

After computing the coset representatives $\{r_1, \ldots, r_k\} \subseteq G$ using the above procedure, we can decide membership to $R \in \mathrm{Rat}(G)$ using the following equivalence: $g \notin R$ if and only if for some $i \in \{1, \ldots, k\}$, we have $r_i^{-1} g \in H \setminus (r_i^{-1} R \cap H)$. ∎

Remark 3.13. The algorithm in the proof Corollary 3.12 yields a coset enumer-

ation, and the algorithm is typically called the *Todd-Coxeter coset-enumeration*. Its original version in [75] was designed for finding a finite presentation for finite groups, only. For finitely presented groups the coset-enumeration yields an effective construction of the Schreier graph if $[G : H]$ is finite, see [49]. However, even for finitely presented groups there is no computable upper time bound for the Todd-Coxeter coset-enumeration in general.

4. Smith normal forms and commensurators. It is a classical fact from linear algebra that every matrix $m \in \mathbb{Q}^{n\times n}$ admits a *Smith normal form*. For $n = 2$, the Smith normal form of a non-zero $m \in \mathbb{Q}^{2\times 2}$ is a factorization

$$m = r\, e \begin{pmatrix} 1 & 0 \\ 0 & q \end{pmatrix} f \tag{4.1}$$

such that $r \in \mathbb{Q}$ is a positive rational number, $e, f \in \mathrm{SL}(2,\mathbb{Z})$, and $q \in \mathbb{Z}$. Note that we may assume that r is positive because $m \neq 0$ and $\left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right) \in \mathrm{SL}(2,\mathbb{Z})$. Since $r^2 q = \det(m)$, the sign of $\det(m)$ is determined by the sign of q. For $q \in \mathbb{Z}$ we fix the notation

$$s_q = \left(\begin{smallmatrix} 1 & 0 \\ 0 & q \end{smallmatrix}\right). \tag{4.2}$$

If we write $m = r\, e\, s_q\, f$ for $m \in \mathbb{Q}^{2\times 2}$, then we refer to it as the Smith normal form of m according to (4.1) and (4.2). We use Smith normal forms only when $n = 2$. The computation of Smith normal form is closely related to Gaussian elimination and relies on gcd-computations. More details are given in Section 4.1.

4.1. Computation of the Smith normal form. As mentioned above, a *Smith normal form* of a non-zero matrix m in $\mathbb{Q}^{2\times 2}$ is defined by a factorization $m = r \cdot e \left(\begin{smallmatrix} 1 & 0 \\ 0 & q \end{smallmatrix}\right) f$ where $0 < r \in \mathbb{Q}$, $e, f \in \mathrm{SL}(2,\mathbb{Z})$, and $q \in \mathbb{Z}$. Moreover, r and q are uniquely determined by the matrix m (but e and f are not unique). The uniqueness of r and q can be seen as follows. Let $m = r_1 \cdot e_1 \left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right) f_1 = r_2 \cdot e_2 \left(\begin{smallmatrix} 1 & 0 \\ 0 & q \end{smallmatrix}\right) f_2$ with $0 < r_i \in \mathbb{Q}$, $e_i, f_i \in \mathrm{SL}(2,\mathbb{Z})$ for $i = 1, 2$, and $p, q \in \mathbb{Z}$. Multiplying m on the left by ${r_2}^{-1} \cdot {e_2}^{-1}$ and on the right by ${f_1}^{-1}$ yields $\frac{r_1}{r_2} \cdot e \left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 & 0 \\ 0 & q \end{smallmatrix}\right) f$ with $e, f \in \mathrm{SL}(2,\mathbb{Z})$. Since $0 < \frac{r_1}{r_2}$ we can write $\frac{r_1}{r_2} = \frac{s}{t}$, where s, t are positive natural numbers such that $\gcd(s,t) = 1$. Therefore, it is enough to show that $\frac{s}{t} \cdot e \left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 & 0 \\ 0 & q \end{smallmatrix}\right) f$ implies $s/t = 1$ and $p = q$. Let $e = (e_{ij})$ and $f = (f_{ij})$, then

$$\begin{pmatrix} se_{11} & spe_{12} \\ se_{21} & spe_{22} \end{pmatrix} = \begin{pmatrix} tf_{11} & tf_{12} \\ tqf_{21} & tqf_{22} \end{pmatrix}.$$

Since $\gcd(s,t) = 1$, the positive integer t divides e_{11} and e_{21}. Hence, t divides $\det(e) = 1$. Thus, $t = 1$, and by symmetry we also have $s = 1$. Therefore, $e \left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 & 0 \\ 0 & q \end{smallmatrix}\right) f$, and hence $\det(\left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right)) = \det(\left(\begin{smallmatrix} 1 & 0 \\ 0 & q \end{smallmatrix}\right))$. Clearly, this implies $p = q$.

The following lemma is a special case of a polynomial-time result by Kannan and Bachem [40]. We include a proof because the result for 2×2 matrices is rather easy to show. Moreover, for 2×2 matrices we obtain a soft cubic time bound whereas [40] just states polynomial time.[8]

LEMMA 4.1. *On input $0 \neq m \in \mathbb{Q}^{2\times 2}$ with $n = \| m \|_{bin}$ we can compute $0 < r \in \mathbb{Q}$, matrices $e, f \in \mathrm{SL}(2,\mathbb{Z})$, and $q \in \mathbb{Z}$ in soft-cubic time $\widetilde{\mathcal{O}}(n^3)$ such that $m = r \cdot e \left(\begin{smallmatrix} 1 & 0 \\ 0 & q \end{smallmatrix}\right) f$.*

[8] We did not check whether "soft cubic time" is an upper bound for computing the Smith normal form in higher dimensions, too.

Our proof follows [40]. It relies on the fact that gcd's can be computed in cubic time. This fact is straightforward, but it is not optimal. For example, Schönhage [67] gives a $\mathcal{O}(n(\log n)^2(\log\log n))$ algorithm. Möller [54] gives another quasi-linear time algorithm which (according to Möller) runs slightly faster than earlier quasi-linear time algorithms.

Proof. On input m we calculate some positive integer p such that $p \cdot m = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ where $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathbb{Z}^{2\times 2}$. For example, we may choose the product over the denominators of all entries in m. Knowing the Smith normal form of A, we obtain the Smith normal form of m by multiplication with p^{-1}. Hence, w.l.o.g., we assume that $m = A$, and let $D = \det(A)$.

With the help of matrices $e, f \in \mathrm{SL}(2,\mathbb{Z})$ we may assume $a = \| A \|_{\max} \geq 1$. If $a = 1$, then we are done: we have $r = 1$ and $q = D$ because $\left(\begin{smallmatrix} 1 & 0 \\ -c & 1 \end{smallmatrix}\right) \cdot \left(\begin{smallmatrix} 1 & b \\ c & d \end{smallmatrix}\right) \cdot \left(\begin{smallmatrix} 1 & -b \\ 0 & 1 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 & 0 \\ 0 & d-bc \end{smallmatrix}\right)$. Hence, from now on we assume $a \geq 2$. In the first phase we reduce the problem to the case where A is a diagonal matrix. This is true if $b = c = 0$. By symmetry, we may assume in the first phase that $a \geq 2$ and $b \neq 0$.

First phase. Let $1 \leq g = \gcd(a,b) = pa + qb$ with $0 \leq q < a$. This is possible since $(p+b)a + (q-a)b = pa + qb$. Then $\left(\begin{smallmatrix} p & -b/g \\ q & a/g \end{smallmatrix}\right) \in \mathrm{SL}(2,\mathbb{Z})$, and hence

$$\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \cdot \left(\begin{smallmatrix} p & -b/g \\ q & a/g \end{smallmatrix}\right) = \left(\begin{smallmatrix} g & 0 \\ pc+qd & D/g \end{smallmatrix}\right).$$

If $\gcd(a,b) = a$, then we choose $p = 1$ and $q = 0$. Otherwise $a/g \geq 2$ and, since $b \neq 0$, we have $1 \leq |b| < a = \| A \|_{\max}$. Hence:

$$|p| = \left|\frac{g - qb}{a}\right| \leq \frac{g}{a} + \frac{(a-1)\,|b|}{a} \leq 1/2 + a - 1 < a.$$

Thus, after the first step and by left-right symmetry due to transposition of matrices, we may assume without restriction that we actually start with a matrix

$$A' \in \left\{ \left(\begin{smallmatrix} g & 0 \\ D' & D/g \end{smallmatrix}\right), \left(\begin{smallmatrix} g & D' \\ 0 & D/g \end{smallmatrix}\right) \right\},$$

where $g|a$ and $0 \leq |D'| < 2\|A\|^2_{\max}$. If $D' = 0$, then we stop because the matrix is diagonal which is the aim for this phase.

We now assume that $D' \neq 0$ and $A' = \left(\begin{smallmatrix} g & 0 \\ D' & D/g \end{smallmatrix}\right)$. Let $g' = \gcd(g, D') = pg + qD'$ with $0 \leq q < g$. We have $\left(\begin{smallmatrix} p & q \\ -D'/g' & g/g' \end{smallmatrix}\right) \in \mathrm{SL}(2,\mathbb{Z})$ and

$$\left(\begin{smallmatrix} p & q \\ -D'/g' & g/g' \end{smallmatrix}\right) \cdot \left(\begin{smallmatrix} g & 0 \\ D' & D/g \end{smallmatrix}\right) = \left(\begin{smallmatrix} g' & qD/g \\ 0 & D/g' \end{smallmatrix}\right).$$

If $g \mid D'$, then $q = 0$ and the above matrix is diagonal. So, we stop the first phase. Thus, without restriction $0 < q < g$, and, in particular, $g \neq 1$. Clearly: $g' \mid g \mid a$. Let $D'' = qD/g$. Then

$$0 \leq |D''| < |D| \leq 2\|A\|^2_{\max}.$$

Since $0 < q < g$, we have $g' < g$. Since each time we have either $g/g' \geq 2$ or $g \mid D'$, we finish after at most $\log \|A\|_{\max}$ steps. This completes the first phase.

Second phase. We continue with a matrix $A'' = \left(\begin{smallmatrix} g & 0 \\ 0 & D/g \end{smallmatrix}\right)$ for some $g \mid a$. If $g \mid D/g$ we are done. Thus, w.l.o.g. $D \neq 0$ and letting $d = D/g$ we write

$$\left(\begin{smallmatrix} g & 0 \\ 0 & d \end{smallmatrix}\right) = \left(\begin{smallmatrix} \gcd(g,d) & 0 \\ 0 & \gcd(g,d) \end{smallmatrix}\right) \cdot \left(\begin{smallmatrix} g/\gcd(g,d) & 0 \\ 0 & d/\gcd(g,d) \end{smallmatrix}\right).$$

Let $g' = g/\gcd(g,d)$ and $d' = d/\gcd(g,d)$. Note that $g' \geq 2$ because $g \neq \gcd(g,d)$. We add the right column of $\left(\begin{smallmatrix} g' & 0 \\ 0 & d' \end{smallmatrix}\right)$ to the left one by multiplying with the matrix $\left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$. We obtain the matrix $\left(\begin{smallmatrix} g' & 0 \\ d' & d' \end{smallmatrix}\right)$. We let $pg' + qd' = 1$ with $0 \leq q < g'$ and $p = (1 - qd')/g'$. Hence, $|p| \leq |d'| + 1/g' - |d'|\,/g' \leq |d'| \leq |D| \leq 2\|A\|^2_{\max}$. Then,

$$\left(\begin{smallmatrix} p & q \\ -d' & g' \end{smallmatrix}\right) \cdot \left(\begin{smallmatrix} g' & 0 \\ d' & d' \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 & qd' \\ 0 & g'd' \end{smallmatrix}\right).$$

Subtracting qd' times the left column from the right one by multiplying with the matrix $\left(\begin{smallmatrix} 1 & -qd' \\ 0 & 1 \end{smallmatrix}\right)$, we obtain the desired result. ∎

sec:commens

4.2. Commensurators. The notion of a commensurator is well established in group theory. Let G be a group and H be its subgroup. Then the *commensurator of H in G* is defined to be the set of all $g \in G$ such that $H \cap H^g$ has finite index in H and in H^g, see for example [25, Def. 5.17]. Here, and in the following, we abbreviate gHg^{-1} as H^g which is a standard notation in group theory. It is a known fact that the commensurator is a subgroup of G, see [25, Ex. 5.18].[9]

Now, let H be an arbitrary group. For the sake of brevity, we say that a group G containing H is *a commensurator of H* if for all $g \in G$ the subgroup $H \cap H^g$ has finite index in H.[10] Note that this also implies that $[H^g : H \cap H^g] = [H : H^{g^{-1}} \cap H]$ is finite for all $g \in G$. Hence G is the commensurator of H in G.

If H has finite index in G, then G is a commensurator of H because the intersection $H \cap H^g$ has finite index in G (and hence in H) for any $g \in G$. If $K \leq H$ are subgroups of G and G is a commensurator of K, then obviously H is a commensurator of K, too. We also use the following lemma in the proof of Proposition 4.3.

lem:index

LEMMA 4.2. *Let $K \leq H \leq G$ be a chain of subgroups such that the index $[H : K]$ is finite. Then G is a commensurator of K if and only if G is a commensurator of H.*

Proof. Suppose that G is a commensurator of K. Then for all $g \in G$ we have:

$$[H : H \cap H^g] \leq [H : K \cap K^g] = [H : K][K : K \cap K^g] < \infty.$$

Since $[H : K]$ is finite, G is a commensurator of H. For the other direction, it is enough to show that for all $g \in G$ we have $[K : K \cap K^g] \leq [H : H \cap H^g][H : K]$ since, by assumption, both $[H : H \cap H^g]$ and $[H : K]$ are finite. For that, we start with the following equation

eq:comm

$$\begin{aligned}[H : K][K : K \cap K^g] &= [H : K \cap K^g] \\ &= [H : H \cap H^g][H \cap H^g : K \cap H^g][K \cap H^g : K \cap K^g].\end{aligned} \tag{4.3}$$

Next, we use the fact that for all subgroups N and K of H, the set of left cosets $N/K \cap N$ embeds into H/K, and hence $[N : N \cap K] \leq [H : K]$. The fact implies that

$$[H \cap H^g : K \cap H^g] = [H \cap H^g : (H \cap H^g) \cap K] \leq [H : K] \quad \text{and}$$
$$[K \cap H^g : K \cap K^g] = [K \cap H^g : (K \cap H^g) \cap K^g] \leq [H^g : K^g] = [H : K].$$

Substituting these inequalities in (4.3) and dividing every term by $[H : K]$, we obtain that $[K : K \cap K^g] \leq [H : H \cap H^g][H : K]$, which proves the lemma. ∎

[9] Note that in geometric group theory there is a more general notion of an *abstract commensurator*, which is different from what we use here, see [25, Def. 5.13].

[10] In [44], a group G and its subgroup H that satisfy such property are called a *Hecke pair* (H, G).

The statement of the following Proposition 4.3 holds for all $n \in \mathbb{N}$. However, the case $n = 2$ has a short proof which is also given below.

prop:commens

PROPOSITION 4.3. *The group* $\mathrm{GL}(n,\mathbb{Q})$ *is a commensurator of* $\mathrm{SL}(n,\mathbb{Z})$ *and of any subgroup* $G \le \mathrm{GL}(n,\mathbb{Q})$ *which contain* $\mathrm{SL}(n,\mathbb{Z})$ *as a subgroup of finite index. In particular,* $\mathrm{GL}(n,\mathbb{Q})$ *is a commensurator of both* $\mathrm{SL}(n,\mathbb{Z})$ *and* $\mathrm{GL}(n,\mathbb{Z})$.

krieg1990hecke

Proof. In [44, Ch. V], it is shown that $\mathrm{GL}(n,\mathbb{Q})$ is a commensurator of $\mathrm{GL}(n,\mathbb{Z})$ for all $n \in \mathbb{N}$. Using Lemma 4.2, we see that $\mathrm{GL}(n,\mathbb{Q})$ is a commensurator of $\mathrm{SL}(n,\mathbb{Z})$, too. This implies the result. ∎

For $n = 2$ we give a short and direct proof of Proposition 4.3 based on the Smith normal form, which we have defined only for $n = 2$. For $n \ge 3$, such a proof becomes more technical (see [44, Ch. V]). Our applications only concern 2×2 matrices.

krieg1990hecke

Proof of Proposition 4.3 for $\boldsymbol{n = 2}$. It is enough to show that $\mathrm{GL}(2,\mathbb{Q})$ is a commensurator of $H = \mathrm{SL}(2,\mathbb{Z})$. To see this, recall that $s_q = \left(\begin{smallmatrix} 1 & 0 \\ 0 & q \end{smallmatrix}\right)$ (as in Equation (4.2)), where $q \in \mathbb{Z}$. Writing a matrix $g \in \mathrm{GL}(2,\mathbb{Q})$ in its Smith normal form yields $g = r\, e\, s_q\, f$ with $r \in \mathbb{Q}$ and $e, f \in \mathrm{SL}(2,\mathbb{Z})$. Then the index of $gHg^{-1} \cap H$ in H is the same as the index of $s_q H s_q^{-1} \cap H$ in H. We have $s_q^{-1} \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) s_q = \left(\begin{smallmatrix} a & qb \\ c/q & d \end{smallmatrix}\right)$. Hence, a matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in H$ belongs to the intersection $s_q H s_q^{-1} \cap H$ if and only if $c \in q\mathbb{Z}$. Thus, $\ker(\bmod q) \subseteq s_q H s_q^{-1} \cap H$, where $\bmod q : \mathrm{SL}(2,\mathbb{Z}) \to \mathrm{SL}(2,\mathbb{Z}/q\mathbb{Z})$ is the canonical homomorphism. Thus, the index of $s_q H s_q^{-1} \cap H$ in H is bounded by the size of the finite group $\mathrm{SL}(2,\mathbb{Z}/q\mathbb{Z})$. It follows that $\mathrm{GL}(2,\mathbb{Q})$ is a commensurator of $\mathrm{SL}(2,\mathbb{Z})$. ∎

The size of $\mathrm{SL}(2,\mathbb{Z}/q\mathbb{Z})$ is obviously bounded by some polynomial in q. This would be good enough for our purposes, but not good enough in practical applications. As a matter of fact, there is a better and more precise estimate for the index of $s_q H s_q^{-1} \cap H$ in H which is stated next.

prop:index

PROPOSITION 4.4. *As above, denote* $H = \mathrm{SL}(2,\mathbb{Z})$. *Let* $g \in \mathrm{GL}(2,\mathbb{Q})$ *and* $g = r\, e\, s_q\, f$ *be its Smith normal form with* $0 < r \in \mathbb{Q}$, $e, f \in \mathrm{GL}(2,\mathbb{Z})$, *and* $s_q = \left(\begin{smallmatrix} 1 & 0 \\ 0 & q \end{smallmatrix}\right)$. *Then*

$$[H : (gHg^{-1} \cap H)] = [H : (s_q H s_q^{-1} \cap H)] \in \mathcal{O}(|q| \log |q|).$$

Proof. We just have seen above that $[H : (gHg^{-1} \cap H)] = [H : (s_q H s_q^{-1} \cap H)]$, and that $s_q H s_q^{-1} \cap H$ consists of those matrices $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in H$ for which $c \in q\mathbb{Z}$. The subgroup $s_q H s_q^{-1} \cap H$ is also denoted as $\Gamma_0(q)$ in the literature. The index of $\Gamma_0(q)$ in H is equal to $|q| \prod_{p|q}(1 + 1/p)$, where the product is taken over all prime divisors of q, see [19, Ex. 1.2.3(e)].

DiamondS2005

We now estimate the above product $\prod_{p|q}(1 + 1/p)$. Note that

$$\ln \prod_{p|q} \left(1 + \frac{1}{p}\right) = \sum_{p|q} \ln \left(1 + \frac{1}{p}\right) \;\le\; \sum_{p|q} \frac{1}{p} \;\le\; \sum_{p \le |q|} \frac{1}{p} \;\le\; \ln \ln |q| + C$$

for some constant $C > 0$, where the sums and product are taken over all primes p such that $p \mid q$ or $p \le |q|$, respectively. The last inequality follows from Mertens's Second Theorem, see [35, p. 466]. Therefore, $\prod_{p|q}(1 + 1/p) \le e^C \ln |q| \in \mathcal{O}(\log |q|)$ and $[H : (s_q H s_q^{-1} \cap H)] \in \mathcal{O}(|q| \log |q|)$. ∎

HardyW2008

sec:algglq

5. Dichotomy in $\mathrm{GL}(2,\mathbb{Q})$**.** One of the main results of this paper is Theorem 5.4 stated below, which classifies the f.g. subgroups G sitting strictly between $\mathrm{GL}(2,\mathbb{Z})$ and $\mathrm{GL}(2,\mathbb{Q})$ into two mutually exclusive classes. An important consequence of this

dichotomy is that, for such subgroups, $\mathrm{Rat}(G)$ is never closed under intersection, and in particular it is not a relative Boolean algebra. This is a result of independent interest. In our proof of the dichotomy, the *Baumslag-Solitar group* $\mathrm{BS}(p,q)$ where $1 = p < q$ shows up.[11] Recall that $\mathrm{BS}(p,q) = \langle a, t \mid ta^pt^{-1} = a^q\rangle$ is actually defined for all $p, q \in \mathbb{Z}$, but up to isomorphism it is enough to impose $0 \leq p \leq |q|$. As we will see, $\mathrm{BS}(|q|, q)$ for $|q| \geq 2$ contains a direct product of a free group of rank two and $\mathbb{Z}$. This is a consequence of Bass-Serre theory [73], see for example [30].

The case $p = 0$ is not very interesting since $\mathrm{BS}(0,q)$ is isomorphic to the free product $\mathbb{Z} * (\mathbb{Z}/q\mathbb{Z})$. It is fairly easy to see that $\mathrm{BS}(p,q)$ has no free subgroup of finite index unless $pq = 0$, see [32]. As a consequence, in both cases of the dichotomy in Theorem 5.4, the group $\mathrm{GL}(2,\mathbb{Z})$ has infinite index in G when $\mathrm{GL}(2,\mathbb{Z}) < G \leq \mathrm{GL}(2,\mathbb{Q})$.

Actually, we prove more: if G contains a matrix of the form $\left(\begin{smallmatrix} r_1 & 0 \\ 0 & r_2 \end{smallmatrix}\right)$ with $|r_1| \neq |r_2|$ (which is the second case of the dichotomy), then G contains some $\mathrm{BS}(1,q)$ for $q \geq 2$ which has **infinite** index in G. It is wide open whether the membership for rational subsets of G can be decided in that second case.

For example, let $p \geq 2$ be a prime, and let G' be generated by $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$, $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, and $\left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right)$. In this case $\left(\begin{smallmatrix} p & 0 \\ 0 & p^{-1} \end{smallmatrix}\right)$ also belongs to G'. Let $\mathbb{Z}[1/p]$ denote the ring $\{p^n r \in \mathbb{Q} \mid n, r \in \mathbb{Z}\}$. It is known by [6] that $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$, $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, and $\left(\begin{smallmatrix} p & 0 \\ 0 & p^{-1} \end{smallmatrix}\right)$ generate the special linear group $\mathrm{SL}(2,\mathbb{Z}[1/p])$ of 2×2 matrices over $\mathbb{Z}[1/p]$. Hence, G' contains $\mathrm{SL}(2,\mathbb{Z}[1/p])$ as a subgroup. The structure of $\mathrm{SL}(2,\mathbb{Z}[1/p])$ is described in [73, II.1 Cor. 2]: it is an amalgam of two copies of $\mathrm{SL}(2,\mathbb{Z})$ over a common subgroup of finite index. It is however unknown how to decide subgroup membership for such amalgams. Moreover, $\left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right)$ acts by conjugation on $\mathrm{SL}(2,\mathbb{Z}[1/p])$, and since $\left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right)$ generates an infinite cyclic group, G' is a semi-direct product of the form $G' = \mathrm{SL}(2,\mathbb{Z}[1/p]) \rtimes \mathbb{Z}$. Hence, even if the subgroup membership for $\mathrm{SL}(2,\mathbb{Z}[1/p])$ were decidable, it could still be undecidable in G'. The situation is more friendly for the subgroup generated by the matrices $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right)$ because it is the group $\mathrm{UT}(2,\mathbb{Z}[1/p]) \rtimes \mathbb{Z} \cong Z[1/p] \rtimes \mathbb{Z} \cong \mathrm{BS}(1,p)$, where $\mathrm{UT}(2,\mathbb{Z}[1/p])$ is the group of 2×2 upper-unitriangular matrices over $\mathbb{Z}[1/p]$. The membership problem for rational subsets of $\mathrm{BS}(1,q)$ is decidable for all $q \geq 2$ by [14]. However, it is not clear how to generalize this result to extensions of $\mathrm{BS}(1,q)$ of infinite index.

It is also shown in [14, Ex. 3.7] that $\mathrm{Rat}(\mathrm{BS}(1,q))$ is not closed under finite intersection for $q \geq 2$. Using Theorem 3.8, we show next that this non-closure property holds whenever $0 \leq p \leq |q|$ and $|q| \geq 2$. In particular, it covers the "famous" Baumslag-Solitar group $\mathrm{BS}(2,3)$ and cases where q is negative. To the best of our knowledge the following dichotomy theorem for Baumslag-Solitar groups has not been stated explicitly or shown elsewhere.

thm:bsint

THEOREM 5.1. *Let $p, q \in \mathbb{Z}$ and $\mathrm{BS}(p,q)$ be the Baumslag-Solitar group. Then $\mathrm{Rat}(\mathrm{BS}(p,q))$ is a Boolean algebra if and only if it is closed under finite intersection if and only if $|pq| \leq 1$.*

Proof. We will use some well-known facts about Baumslag-Solitar groups. What we need for the proof can be found, for example, in [21, Sect. 8.4.2] and elsewhere in the literature. For example, as we mentioned above, we assume without restriction that $0 \leq p \leq |q|$. We let $\overline{a} = a^{-1}$ and $\overline{t} = t^{-1}$. The group $\mathrm{BS}(0,q)$ is the free

[11] The group $\mathrm{BS}(p,q)$ is an HNN-extension (named after Higman, Neumann, and Neumann) of $\mathbb{Z}$ over the subgroups $p\mathbb{Z}$ and $q\mathbb{Z}$ with a "stable letter" t.

product $\mathbb{Z} * (\mathbb{Z}/q\mathbb{Z})$, hence $\mathrm{Rat}(\mathrm{BS}(0,q))$ is a Boolean algebra by [66, 46]. Therefore, we only need to consider the case when $p \geq 1$. Let $p = |q| = 1$; then we know that $\mathrm{Rat}(\mathrm{BS}(1,1))$ is a Boolean algebra since $\mathrm{BS}(1,1) = \mathbb{Z} \times \mathbb{Z}$, and therefore the rational sets are the semi-linear subsets. Also, $\mathrm{Rat}(\mathrm{BS}(1,-1))$ is a Boolean algebra by Corollary 3.12 since it contains $\mathbb{Z} \times \mathbb{Z}$ as a subgroup of index two. It remains to show that $\mathrm{Rat}(\mathrm{BS}(p,q))$ is not closed under intersection for $|q| \geq 2$ and $1 \leq p \leq |q|$.

We treat the case $2 \leq p = |q|$ first. Consider the f.g. subgroup F of $\mathrm{BS}(|q|,q)$ which is generated by the two commutators $\alpha = [a,t]$ and $\beta = [a,t^2]$. Then we obtain a natural epimorphism ψ from the free group $F_{\alpha,\beta}$ onto F. Now, consider a non-trivial freely reduced word $w \neq 1$ in $F_{\alpha,\beta}$. Then a Britton-reduction (with respect to a and t) yields a nontrivial element in $\mathrm{BS}(|q|,q)$ since $|q| \geq 2$. For example, $\alpha\beta = at\bar{a}\bar{t}\ at^2\bar{a}\bar{t}^2$ is Britton-reduced, and the Britton reduction of $\alpha\beta^{-1}$ yields $at\bar{a}\bar{t}\ t^2a\bar{t}^2\bar{a} = at\bar{a}at\bar{t}^2\bar{a}$. Based on this observation, standard arguments with an induction on $|w|$ show that ψ is injective. Therefore F is a free group. It is easy to check that a^p commutes with α and β when $p = |q|$. This implies that the intersection of the infinite cyclic group $\langle a^p \rangle$ and F is trivial. Thus, $\mathrm{BS}(q,|q|)$ for $|q| \geq 2$ contains a direct product isomorphic to $F_2 \times \mathbb{Z}$, where F_2 is the free group of rank two generated by α, β and $\mathbb{Z}$ is generated by a^p. We have seen in Lemma 3.6 that $F_2 \times \mathbb{Z}$ does not satisfy the Howson property. Therefore $\mathrm{Rat}(\mathrm{BS}(q,|q|))$ is not closed under finite intersection.

In order to finish the proof it remains to consider $\mathrm{BS}(p,|q|)$ where $1 \leq p < |q|$. The proof has a different flavor than the one for $\mathrm{BS}(|q|,q)$ with $|q| \geq 2$. We let $A_+ = a^0 \cup \cdots \cup a^{p-1}$, and $A_- = A_+$ if q is positive and $A_- = a^0 \cup \cdots \cup a^{1-p}$ if q is negative. We consider the set $L = a^* \cap T^*a(\bar{t}\,\bar{t}\,)^*$ with $T = tA_-tA_+$. Then L is the intersection of two rational sets. We claim that L is not rational. By contradiction, assume that $L \in \mathrm{Rat}(\mathrm{BS}(p,q))$. Then, by Theorem 3.8, there is an $a^{\mathbb{Z}}$-NFA $\mathcal{A}$ which accepts L. The set L is not empty since $a \in L$. If $a^k \in L$, then $1 \leq k \in \mathbb{N}$ and we can write $a^k \in T^sat^{-2s}$ for some $s \in \mathbb{N}$. Thanks to the choice of A_+ and A_-, we can also state that for each $s \in \mathbb{N}$ there is a unique $k_s \in \mathbb{N}$ such that $a^{k_s} \in T^sat^{-2s}$ and $k_s \geq (q/p)^{2s}$. This can be shown by induction on s. More precisely, for each k_s there a unique k'_s such that $1 \leq k_s \leq k'_s \leq k_s + p - 1$ and $k'_s \in p\mathbb{N}$. Let us define $\ell_s = (q/p)k'_s$. Then we have $a^{\ell_s} \in tA_+T^sat^{-2s+1}$. Note that $\ell_s < 0 \iff q < 0$. Now, consider the unique ℓ'_s with $\ell_s \leq \ell'_s \leq \ell_s + p - 1$ if $q > 0$ (or with $\ell_s \geq \ell'_s \geq \ell_s + 1 - p$ if $q < 0$) such that $\ell'_s \in p\mathbb{Z}$. This leads to the next positive $k_{s+1} \in \mathbb{N}$ such that $k_{s+1} = (q/p)\ell'_s \geq (q/p)^2k_s$.

Putting things together, we have shown that $L = \{a^{k_s} \mid s \in \mathbb{N}\}$ with $k_s \geq (q/p)^{2s}$ for all $s \in \mathbb{N}$. The assumption $L \in \mathrm{Rat}(\mathrm{BS}(p,q))$ implies $L = \psi(K)$ for some homomorphism ψ and some regular language $K \subseteq \{a,t,\bar{t}\}^*$. By the pumping lemma for regular languages (also known as uvw-Theorem), we know that for all s there is some $r \neq s$ with $|k_s - k_r| \in \mathcal{O}(1)$. It is a contradiction with the above lower bound on k_s. ∎

Another ingredient to show the dichotomy is the next proposition and its corollary.

prop:964 PROPOSITION 5.2. *Let G, H, and A be groups, where the center $Z(H)$ is trivial and A is Abelian. If $\varphi : H \to G \times A$ is an injective homomorphism, then the induced homomorphism $\varphi_1 : H \to G$ is injective where $\varphi_1(h) = g$ for $\varphi(h) = (g,a)$.*

Proof. It is enough to show that $\varphi_1(h) = 1$ implies $h = 1$. To see this, take $h \in H$ with $\varphi(h) = (1,a)$. Then $(1,a)$ is in the center of $G \times A$ and therefore in the center of $\varphi(H) \cong H$. Therefore $h \in Z(H)$ which is trivial. Hence, $\varphi(h) = (1,a)$ implies $h = 1$ and we are done. ∎

The following corollary holds for all $\mathrm{BS}(p,q)$ where $1 \leq p < |q|$ because the center of those $\mathrm{BS}(p,q)$ is trivial. More generally, the center of an HNN-extension $\mathrm{HNN}(H,t,\varphi)$ with an isomorphism $\varphi : A \to B$ is trivial if $A \neq H$ and $\{a \in Z(A) \mid \varphi(a) = a\} = \{1\}$. However, we need Corollary 5.3 only for $\mathrm{BS}(1,q)$ with $q \geq 2$: in this case the proof is less technical.

cor:976 COROLLARY 5.3. *Let G and A be groups where A is Abelian. If $1 < |q|$ and the Baumslag-Solitar group $\mathrm{BS}(1,q)$ appears as a subgroup in the $G \times A$, then $\mathrm{BS}(1,q)$ appears in G.*

Proof. The group $\mathrm{BS}(1,q)$ is isomorphic to the semi-direct product $\mathbb{Z}[1/q] \rtimes \mathbb{Z}$. The elements of $\mathbb{Z}[1/q] \rtimes \mathbb{Z}$ are pairs (r,m) where $r = pq^e$ with $p, e, m \in \mathbb{Z}$ and the multiplication $(r,m) \cdot (s,n) = (r + q^m s, m + n)$. A direct verification shows that the center of $\mathbb{Z}[1/q] \rtimes \mathbb{Z}$ is trivial. Thus, Proposition 5.2 yields the result. ∎

thm:nofinex THEOREM 5.4. *Let G be a f.g. group such that $\mathrm{GL}(2,\mathbb{Z}) < G \leq \mathrm{GL}(2,\mathbb{Q})$. Then there are two mutually exclusive cases.*

1. *G is isomorphic to $\mathrm{GL}(2,\mathbb{Z}) \times \mathbb{Z}^k$ for some $k \geq 1$.*
2. *G contains a subgroup which is an extension of infinite index of $\mathrm{BS}(1,q)$ for some $q \geq 2$.*

Furthermore, in both cases of the dichotomy, $\mathrm{Rat}(G)$ is not closed under finite intersection.

Proof. We distinguish two cases. In the first case, we suppose that G is generated by $\mathrm{GL}(2,\mathbb{Z})$ and finitely many elements from the center $Z(G)$. Since $\mathrm{GL}(2,\mathbb{Z})$ is a subgroup of G, we see that $Z(G) \leq \{\left(\begin{smallmatrix} r & 0 \\ 0 & r \end{smallmatrix}\right) \mid r \in \mathbb{Q}^*\}$. Moreover, since $-1 \in \mathrm{GL}(2,\mathbb{Z}) < G$, the group G is generated by $\mathrm{GL}(2,\mathbb{Z})$ and a nontrivial f.g. subgroup $Z \leq \{\left(\begin{smallmatrix} r & 0 \\ 0 & r \end{smallmatrix}\right) \mid r \in \mathbb{Q}^* \wedge r > 0\}$. Hence $Z \cong \mathbb{Z}^k$ for some $k \geq 1$ because $\{r \in \mathbb{Q}^* \mid r > 0\}$ is torsion free and Z is finitely generated and Abelian. Since $\mathrm{GL}(2,\mathbb{Z})$ contains a free group of rank 2 and $k \geq 1$, Lemma 3.6 tells us that $\mathrm{Rat}(G)$ is not closed under finite intersection.

Assume we are not in the first case. Then consider any finite generating set of G and write the generators in their Smith normal form $re\left(\begin{smallmatrix} 1 & 0 \\ 0 & q \end{smallmatrix}\right)f$ with $0 < r \in \mathbb{Q}$, $e, f \in \mathrm{GL}(2,\mathbb{Z})$ and $q \in \mathbb{Z}$. Since $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right) \in \mathrm{GL}(2,\mathbb{Z}) < G$, the generators can be chosen from $\mathrm{GL}(2,\mathbb{Z})$ and matrices of the form $\left(\begin{smallmatrix} r & 0 \\ 0 & rq \end{smallmatrix}\right)$ with $0 < r \in \mathbb{Q}$ and $0 \neq q \in \mathbb{N}$. Note that there is at least one generator $s = \left(\begin{smallmatrix} r & 0 \\ 0 & rq \end{smallmatrix}\right)$ where $r > 0$ and $2 \leq q \in \mathbb{N}$, because otherwise we are in the first case.

As usual, we define $\mathrm{BS}(1,q) = \langle a, t \mid tat^{-1} = a^q \rangle$ in standard group generators a and t. Let $b = \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$ and $\varphi : \mathrm{BS}(1,q) \to G$ be a homomorphism such that $\varphi(a) = b$ and $\varphi(t) = s$. It is well-defined since $s\left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)s^{-1} = \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)^q$. Let $\mathrm{BS} = \varphi(\mathrm{BS}(1,q))$. We claim that φ is an isomorphism between $\mathrm{BS}(1,q)$ and BS. To see the claim we observe that every element $g \in \mathrm{BS}(1,q)$ can be written in the form $t^k b^x t^n$ where k, x, n are integers. Suppose $g = t^k b^x t^n$ and $\varphi(g) = 1$. Then $\left(\begin{smallmatrix} 1 & 0 \\ x & 1 \end{smallmatrix}\right) = \varphi(b^x) = \varphi(t^{-n-k}) = \left(\begin{smallmatrix} r & 0 \\ 0 & rq \end{smallmatrix}\right)^{-n-k}$ is a diagonal matrix and $x = 0$. Hence, $g = t^{k+n}$ and $\varphi(g) = s^{k+n} = 1$. This implies $k + n = 0$, and φ is injective. Hence, the claim.

Next, we show that BS has infinite index in G. Consider any $g \in \mathrm{BS} \cap \mathrm{SL}(2,\mathbb{Z})$. As above, consider $f = s^k \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)^x s^m$ with $x, k, m \in \mathbb{Z}$. Since by assumption $\det(f) = 1$, we obtain $m = -k$ and hence $f = \left(\begin{smallmatrix} 1 & 0 \\ q^k x & 1 \end{smallmatrix}\right) \in \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)^{\mathbb{Z}}$. Therefore $\mathrm{SL}(2,\mathbb{Z}) \cap \mathrm{BS}$ is the infinite cyclic group generated by $\left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$. It has infinite index in $\mathrm{SL}(2,\mathbb{Z})$. It follows that G contains an extension of $\mathrm{BS}(1,q)$ of infinite index.

Finally, let us show that $\mathrm{GL}(2,\mathbb{Z}) \times \mathbb{Z}^k$ cannot contain $\mathrm{BS}(1,q)$ for $k \geq 0$. Oth-

erwise, there is no dichotomy. For the sake of contradiction assume the contrary. By Proposition 5.2 this implies $\mathrm{BS}(1,q) \leq \mathrm{GL}(2,\mathbb{Z})$. We have seen in Section 2.4 that $\mathrm{Rat}(\mathrm{GL}(2,\mathbb{Z}))$ is a Boolean algebra because $\mathrm{GL}(2,\mathbb{Z})$ is a f.g. and virtually-free. This implies that for the f.g. subgroup $\mathrm{BS}(1,q)$, the set $\mathrm{Rat}(\mathrm{BS}(1,q))$ is a Boolean algebra. In particular, it is closed under finite intersection. This is a contradiction to Theorem 5.1. ∎

thm:undec

THEOREM 5.5. *Let G be isomorphic to $\mathrm{GL}(2,\mathbb{Z}) \times \mathbb{Z}^k$ with $k \geq 1$. Then, on input $L, R \in \mathrm{Rat}(G)$ it is undecidable whether $L = R$. However, on input $g \in G$ and $R \in \mathrm{Rat}(G)$ it is decidable whether $g \in R$.*

Proof. By Remark 2.1, we know that $\mathrm{GL}(2,\mathbb{Z})$ has a free subgroup F_2 of rank two and index 24. In particular, G contains the free partially commutative monoid $M = \{a,b\}^* \times \{c\}^*$ with $a \neq b$. It was proved by Aalbersberg and Hoogeboom in [1] that the equality problem is undecidable for $\mathrm{Rat}(M)$.

[AH89]

For the decidability, we use a result by Lohrey and Steinberg [47]. They showed that the membership problem for $\mathrm{Rat}(F_2 \times \mathbb{Z}^k)$ is decidable. Since $F_2 \times \mathbb{Z}^k$ has finite index in G, the membership problem for rational subsets in G is decidable by Corollary 3.12. ∎

[LohSte08]

sec:FRAT

6. Flat rational sets. In this section we introduce the notion of *flat rational set* for a semigroup M and a subset T. If $S = \langle T \rangle$ is a subsemigroup of M generated by T, then we can extend positive decidability results for $\mathrm{Rat}(S)$ to the larger family $\mathrm{FRat}(M,S)$. When $\mathrm{Rat}(M)$ is an effective Boolean algebra, then all the decision problems studied here are decidable. However, for a group G sitting between $\mathrm{GL}(2,\mathbb{Z})$ and $\mathrm{GL}(2,\mathbb{Q})$, the family $\mathrm{Rat}(G)$ is never a Boolean algebra unless $G = \mathrm{GL}(2,\mathbb{Z})$, see Theorem 5.4. The main result of this section is Theorem 6.6. It shows that the membership problem and (even stronger) the emptiness problem for Boolean combinations of flat rational sets are decidable for $\mathrm{FRat}(\mathrm{GL}(2,\mathbb{Q}), \mathrm{GL}(2,\mathbb{Z}))$.

The following definition is given for a semigroup M and a subset $T \subseteq M$. The main interest is when M is a monoid and T generates a submonoid $S = \langle T \rangle$. Below we also define when an M-NFA is flat over T. In this case T is a subset of labels of its transitions.

def:fratm

DEFINITION 6.1. *We say that $L \subseteq M$ is* flat rational *over a subset T if L is a finite union of languages of the form $L_0 g_1 L_1 \cdots g_t L_t$ where all $L_i \in \mathrm{Rat}(\langle T \rangle)$ and $g_i \in M$.*

The family of flat rational subsets over T is denoted by $\mathrm{FRat}(M,T)$. If $S = \langle T \rangle$, that is T generates the subsemigroup S of M, then Definition 6.1 implies $\mathrm{FRat}(M,T) = \mathrm{FRat}(M,S)$.

In order to specify a set L in $\mathrm{FRat}(M,S)$ for a subsemigroup S we can also use an M-NFA with a syntactic restriction as in Definition 6.2 with $T = S$. In this case, as soon as the membership to S is decidable, we can check whether an M-NFA is flat over S, and if it is, then we know that the accepted language belongs to $\mathrm{FRat}(M,S)$.

def:flatNFA

DEFINITION 6.2. *Let $T \subseteq M$. An M-NFA $\mathcal{A} = (Q, \delta, I, F)$ is called* flat over T *if no transition having a label outside T lies on a directed cycle.*

rem:fratm

Remark 6.3. As we mentioned in the introduction, the notion of $\mathrm{FRat}(M,S)$ *is a special case* of a *polynomial closure* $\mathrm{Pol}(M,\mathcal{L})$ introduced by Schützenberger in [71]: more precisely, in our special case we have $\mathcal{L} = \mathrm{Rat}(S)$, where S is a subsemigroup of M.[12] There is also a related notion of flatness in the context of finite control

[Sch76]

[12] The results in [71] characterize star-free (or aperiodic) languages as the polynomial closure over a

[Sch76]

systems, see [29] and its references.[13] ◇

The next theorem is a generalization of Theorem 3.8.

THEOREM 6.4. *Let M be a monoid such that all right-invertible elements are invertible*[14] *and H a subgroup of $U(M)$. Then the family $\mathrm{FRat}(M,H)$ is the least family $\mathcal{R}$ of subsets of M satisfying the following conditions:*

- *$\mathcal{R}$ contains all finite subsets of M,*
- *$\mathcal{R}$ is closed under finite union and concatenation,*
- *$\mathcal{R}$ is closed under taking the Kleene-star over subsets of H which belong to $\mathcal{R}$.*

In particular, this implies that $\{L \subseteq H \mid L \in \mathrm{FRat}(M,H)\} = \mathrm{Rat}(H)$.

Proof. Clearly, $\mathrm{Rat}(H) \subseteq \mathcal{R}$ and hence, all flat rational sets over H are contained in $\mathcal{R}$. To prove inclusion in the other direction, we need to show that the family of flat rational subsets of M over H (i) contains all finite subsets of M, (ii) is closed under finite union and concatenation, and (iii) is closed under taking the Kleene-star over subsets of H. The first two conditions are obvious. We show (iii) in two steps. Let L be a flat rational set over H such that $L \subseteq H$. First we show that $L \in \mathrm{Rat}(G)$, where $G = U(M)$ is the group of units of M. Since $L \in \mathrm{Rat}(M)$, there is some M-NFA $\mathcal{A}$ accepting L. After trimming, we may assume without restriction that every transition is used on some accepting path. Let g be any label of a transition. Then, thanks to trimming, there are $u, v \in M$ with $ugv \in L \subseteq H$. Hence, there is some $w \in H$ such that $ugvw = 1 \in G$. Therefore, u has a right-inverse. Since M is Dedekind-finite, we have $u \in G$ and $u^{-1}ugvwu = gvwu = 1$. It follows that g has a right-inverse; and therefore $g \in G$. This shows the first step: $L \in \mathrm{Rat}(G)$.

In the second step we apply Theorem 3.8. It shows $L \in \mathrm{Rat}(H)$. Hence, $L^* \in \mathrm{Rat}(H)$, which concludes the proof of (iii). So, $\mathrm{FRat}(M,H)$ is closed under all three closure properties. It also shows $\{L \subseteq H \mid L \in \mathrm{FRat}(M,H)\} = \mathrm{Rat}(H)$. □

Remark 6.5. In the literature a monoid M is called *Dedekind-finite* if all right-invertible elements are invertible. That is, $ab = 1$ implies $ba = 1$ for all $a, b \in M$. The notation appears for example in [28] and [3, Def. 2.3.2]. The class of Dedekind-finite monoids is closed under taking submonoids. It includes all finite monoids, all cancellative monoids and hence, all groups. If F is a field, then $F^{n\times n}$ is Dedekind-finite because a matrix in $F^{n\times n}$ is invertible if and only if its determinant is not zero. More results about Dedekind-finite monoids are in the classical textbook [17]. In our conference paper [22] the assertion of Theorem 6.4 was stated without the hypothesis that M is Dedekind-finite. However, in our applications we only considered those monoids. Further results in [22] were not affected by the missing hypothesis. The example in Remark 3.10 given by Nyberg-Brodda shows that Theorem 6.4 does not hold in general if M is not Dedekind-finite. ◇

THEOREM 6.6. *Let G be a group with an enumerable representation, and $H \leq G$ be a subgroup such that the following conditions hold:*

- *The family $\mathrm{Rat}(H)$ is an effective relative Boolean algebra.*
- *The group G is a commensurator of H, and on input $g \in G$, we can compute the index of $H_g = gHg^{-1} \cap H$ in H.*
- *The membership problem for H is decidable.*

language class by using prefix codes of bounded-synchronization delay. More results in this direction are in [70] and [24].

[13] In control theory the definition says that every control-state belongs to at most one loop.

[14] This means that M is Dedekind-finite, see Remark 6.5 for a short discussion of this notion.

Then $\mathrm{FRat}(G,H)$ *forms an effective relative Boolean algebra. In particular, given a finite Boolean combination*[15] B *of flat rational sets of* G *over* H*, we can decide the emptiness of* B*.*

Note that we do not require $\mathrm{Rat}(H)$ to be a Boolean algebra. In fact, it is a Boolean algebra if and only if $H \in \mathrm{Rat}(H)$ if and only if H is finitely generated. Before giving the proof of Theorem 6.6 let us first state one of its consequences.

cor:le

COROLLARY 6.7. *Let* $B \subseteq \mathrm{GL}(2,\mathbb{Q})$ *be a finite Boolean combination of flat rational sets of* $\mathrm{GL}(2,\mathbb{Q})$ *over* $\mathrm{GL}(2,\mathbb{Z})$*, then we can decide the emptiness of* B*.*

Proof. By Remark 2.1 the group $\mathrm{GL}(2,\mathbb{Z})$ is a finitely generated virtually free group. Hence, $\mathrm{Rat}(\mathrm{GL}(2,\mathbb{Z}))$ is an effective Boolean algebra by [74]. The group $\mathrm{GL}(2,\mathbb{Q})$ is infinitely generated, but obviously the group of matrices with rational entries has an enumerable representation in which the membership for $\mathrm{GL}(2,\mathbb{Z})$ is decidable. In Section 4, we showed that $\mathrm{GL}(2,\mathbb{Q})$ is a commensurator of its subgroup $\mathrm{GL}(2,\mathbb{Z})$. The index of $\mathrm{GL}(2,\mathbb{Z})_g = g\,\mathrm{GL}(2,\mathbb{Z})g^{-1} \cap \mathrm{GL}(2,\mathbb{Z})$ in $\mathrm{GL}(2,\mathbb{Z})$ is bounded by $|\,\mathrm{GL}(2,\mathbb{Z}/q\mathbb{Z})|$ if $g = re\left(\begin{smallmatrix}1&0\\0&q\end{smallmatrix}\right)f$ is the Smith normal form of g (see Proposition 4.4). Thus, all hypotheses of Theorem 6.6 hold. ∎

For the proof of Theorem 6.6 we will need the following lemma. Recall the notation $H_g = gHg^{-1} \cap H = \{h \in H \mid g^{-1}hg \in H\}$ for $H \le G$. Since we also defined H^g as gHg^{-1}, we have $H_g = H^g \cap H$.

lem:fri

LEMMA 6.8. *Let* G *be a group and* H *be a subgroup,* $L \in \mathrm{Rat}(H)$*, and* $g \in G$*. Then under the assumptions of Theorem 6.6 we can compute an* H*-NFA accepting* $g^{-1}(L \cap H_g)g$*.*

Proof. Since $H_g = gHg^{-1} \cap H$ is of finite index in H, we can compute an NFA $\mathcal{A}'$ accepting $L' = L \cap H_g \in \mathrm{Rat}(H_g)$ by Lemma 3.5. The labels of transitions are in H_g. We have $g^{-1}H_g g \subseteq H$. Hence it is enough to change every label h of transitions in $\mathcal{A}'$ to $g^{-1}hg$. This gives the NFA $\mathcal{A}$ for $g^{-1}(L \cap H_g)g$ over H. ∎

Proof of Theorem 6.6. Let $g \in G$ and $K \in \mathrm{Rat}(H)$. First, we claim that we can rewrite $Kg \in \mathrm{Rat}(G)$ as a finite union of languages $g'K'$ with $g' \in G$ and $K' \in \mathrm{Rat}(H)$. Let us show the claim.

The rewriting process for Kg begins with a computation of a set $U_g \subseteq H$ of left coset representatives of H_g such that $H = \bigcup\{uH_g \mid u \in U_g\}$. This is possible because, by assumption, the membership for H is decidable; and hence, the membership for gHg^{-1} and for $H_g = gHg^{-1} \cap H$ is decidable, too. Moreover, by the second assumption, we can compute the index $k = |H : H_g|$. Thus we can enumerate the elements of H until we find k elements that belong to k different left cosets of H_g. Checking if two elements belong to the same coset is decidable since the membership for H_g can be decided. Thus,

$$\begin{aligned} Kg &= \bigcup\{K \cap uH_g \mid u \in U_g\}g = \bigcup\{ug\,g^{-1}(u^{-1}K \cap H_g)g \mid u \in U_g\} \\ &= \bigcup\{g'g^{-1}(gg'^{-1}K \cap H_g)g \mid g' \in U_g g\}. \end{aligned}$$

Using Lemma 6.8 we obtain $g^{-1}(gg'^{-1}K \cap H_g)g = K' \in \mathrm{Rat}(H)$. This shows the claim.

Note that since membership for H is decidable, we can effectively enumerate a set S of all distinct representatives of the right cosets of H, and moreover for

[15] Complementation in the Boolean combination is taken with respect to G.

each $g \in G$ find a representative $g' \in S$ such that $g \in g'H$.

Let $L \in \mathrm{FRat}(G, H)$. Hence L is a finite union of languages $L_0 g_1 L_1 \cdots g_t L_t$ where all $L_i \in \mathrm{Rat}(H)$. Using the claim, we can write L as a finite union of languages gK with $g \in G$ and $K \in \mathrm{Rat}(H)$. By the above observation, we have $g = g'h$ for some $h \in H$ which can be effectively found. Hence we can write $gK = g'(hK)$, where $hK \in \mathrm{Rat}(H)$. Therefore, every flat rational set L can be written as a union $L = \bigcup_{i=1}^{n} g_i K_i$, where $g_i \in S$ and $K_i \in \mathrm{Rat}(H)$. Since $gK_1 \cup gK_2 = g(K_1 \cup K_2)$, we may assume that all g_i in the expression $L = \bigcup_{i=1}^{n} g_i K_i$ are different.

Now let L and R be two flat rational sets. By the above argument we may assume that $L = \bigcup_{i=1}^{n} a_i L_i$ and $R = \bigcup_{j=1}^{m} b_j R_j$, where $a_i, b_j \in S$ and $L_i, R_j \in \mathrm{Rat}(H)$. Then we have $L \setminus R = \bigcup_{i=1}^{n} \left(a_i L_i \setminus \bigcup_{j=1}^{m} b_j R_j\right)$. Note that if $a_i \notin \{b_1, \ldots, b_m\}$, then $a_i L_i \setminus \bigcup_{j=1}^{m} b_j R_j = a_i L_i$, but if $a_i = b_j$ for some j then $a_i L_i \setminus \bigcup_{j=1}^{m} b_j R_j = a_i(L_i \setminus R_j)$. Since $\mathrm{Rat}(H)$ is an effective relative Boolean algebra, we can compute the rational expression for $L_i \setminus R_j$ in H. Hence we can compute the flat rational expression for $L \setminus R$.

As a consequence, given any language B as a Boolean combination of flat rational sets, we find a flat rational expression for B. Every flat rational expression is a rational expression (over G). Deciding emptiness of a rational expression in a monoid with an enumerable representation is trivial. ∎

sec:GH

For the remainder of this section we let M be a monoid, G be a subgroup of its group of units, and H be a finite index subgroup of G.

Since $\mathrm{Rat}(H) \subseteq \mathrm{Rat}(G)$, the membership problem of $\mathrm{FRat}(M, H)$ is a special case of the membership problem of $\mathrm{FRat}(M, G)$. The aim is to prove the converse: the membership problem of $\mathrm{FRat}(M, G)$ is reducible to the membership problem of $\mathrm{FRat}(M, H)$.

thm:dirfin

THEOREM 6.9. *Let G be a subgroup of the group of units in M and $H \leq G$ be its finite index subgroup. Then we have $\mathrm{FRat}(M, G) = \mathrm{FRat}(M, H)$ and, for every M-NFA $\mathcal{A}$ which is flat over G, there exists an M-NFA $\mathcal{B}$ which is flat over H such that $|\mathcal{B}|$ is polynomial in $|\mathcal{A}|$ and such that $L(\mathcal{A}) = L(\mathcal{B})$.*

Moreover, suppose that $[G : H]$ is known and that the monoid M has an enumerable representation as in Definition 3.2. If both the membership problems for H and for G are decidable, then the construction of the NFA $\mathcal{B}$ is effective.

The main ingredient of the following proof of Theorem 6.9 is the application of Theorem 3.8.

Proof. Clearly, it is enough to show that $\mathrm{FRat}(M, G) \subseteq \mathrm{FRat}(M, H)$. W.l.o.g., we assume that the input is specified by a trim M-NFA $\mathcal{A} = (Q, \delta, q_{\mathrm{in}}, q_{\mathrm{fin}})$, which is flat over G, such that q_{in} is the unique initial state without any incoming transition and q_{fin} the unique final state without any outgoing transition. Moreover, $q_{\mathrm{in}} \neq q_{\mathrm{fin}}$. By adding, if necessary, ε-self-loops[16] we may assume that all other states have incoming and outgoing transitions.

For $i = 1, \ldots, t$, let $\mathcal{A}_i = (Q_i, \delta_i, I_i, F_i)$ be the set of (disjoint) subautomata of $\mathcal{A}$ which are induced by the strongly connected components of $\mathcal{A}$ with a nonempty set of transitions. Thus, $q_{\mathrm{in}}, q_{\mathrm{fin}}$ are the only states which do not appear in any $\mathcal{A}_i$. The initial states I_i (resp., the final states F_i) are defined as those states of $\mathcal{A}_i$ that have incoming (resp., outgoing) transitions in $\mathcal{A}$ which do not belong to $\mathcal{A}_i$. By

[16] Recall that an ε-transition in an M-NFA is a transition $p \xrightarrow{1} q$, where 1 is the neutral element of M.

Definition 6.2, each $\mathcal{A}_i$ is a G-NFA. Let $1 \in R \subseteq G$ be a finite set of right coset representatives for H in G. That is, G is the disjoint union $G = \bigcup_{f\in R} Hf$ with $1 \in R$.

For each $1 \le i \le t$ and $f \in R$, there is a trim G-NFA $\mathcal{A}_{i,f} = (Q_{i,f}, \delta_{i,f}, I_{i,f}, F_{i,f})$ of polynomial size in $|\mathcal{A}| \cdot [G : H]$ such that $Q_{i,f} = Q_i \times R$ and $L(\mathcal{A}_{i,f}) = L(\mathcal{A}_i) \cap Hf$. Note that we have $|\delta_{i,f}| \le |\delta_i|$ because for each $i \in \{1, \dots, t\}$ and $(p, a, q) \in \delta$ there is at most one transition $(p, r_p) \xrightarrow{a} (q, r_q) \in \delta_{i,f}$, where r_p and r_q are the right-cosets given by any path from any state in $I_{i,f}$ to p and q, respectively. This can be shown by using the same idea as in the proof of Theorem 3.8. Hence, $\sum_{1\le i\le t} |\delta_{i,f}| \le |\delta|$. Moreover, we can construct $\mathcal{A}_{i,f}$ in such a way that $|I_{i,f}| \le |I_i|$ and $|F_{i,f}| \le |F_i|$.

If M has an enumerable representation and the membership problems for H and G are decidable, then the construction of each $\mathcal{A}_{i,f}$ is effective: By exhaustive search we can find right-coset representatives for pairwise different cosets until $[G : H]$ of them are found.

Introduce a new final state $p_{i,f}$, and for each $p \in F_{i,f}$ a new transition $p \xrightarrow{f^{-1}} p_{i,f}$. This leads to a new G-NFA $\mathcal{A}'_{i,f} = (Q'_{i,f}, \delta'_{i,f}, I_{i,f}, \{p_{i,f}\})$ such that $L(\mathcal{A}'_{i,f}) = L(\mathcal{A}_{i,f})f^{-1} \subseteq H$. Since $L(\mathcal{A}'_{i,f}) \in \mathrm{Rat}(G)$, we may apply Theorem 3.8. After renaming, we obtain an H-NFA $\mathcal{B}_{i,f} = (Q'_{i,f}, \delta''_{i,f}, I_{i,f}, \{p_{i,f}\})$ such that $L(\mathcal{B}_{i,f}) = L(\mathcal{A}'_{i,f})$.

To finish the construction of $\mathcal{B}$, consider a disjoint union of NFAs

{eq:cB} (6.1) $$\mathcal{B} = \{q_{\mathrm{in}}, q_{\mathrm{fin}}\} \cup \bigcup_{1\le i\le t, f\in R} \mathcal{B}_{i,f}$$

Thus, q_{in} and q_{fin} are reintroduced for the same purpose: q_{in} becomes the unique initial state and q_{fin} becomes the unique final state.

For all $f \in F$, we let $Q_{0,f} = \{q_{\mathrm{in}}\}$ and $Q_{t+1,f} = \{q_{\mathrm{fin}}\}$. One after another, consider all pairs (i, j) where $0 \le i, j \le t + 1$ and $i \ne j$. Then introduce for every transition $p_i \xrightarrow{m_{i,j}} q_j \in \delta$ with $p_i \in Q_i$ and $q_j \in Q_j$ and every $f \in R$, a new transition $p_{i,f} \xrightarrow{f\, m_{i,j}} q_{j,f}$ in $\mathcal{B}$ for every $q_{j,f} \in I_{j,f}$, where $p_{i,f}$ is the unique final state in $\mathcal{B}_{i,f}$. This completes the construction of $\mathcal{B}$. ∎

cor:Pr COROLLARY 6.10. *We have* $\mathrm{FRat}(\mathbb{Q}^{2\times2}, \mathrm{GL}(2,\mathbb{Z})) = \mathrm{FRat}(\mathbb{Q}^{2\times2}, H)$ *for every finite index subgroup* H *of* $\mathrm{GL}(2,\mathbb{Z})$. *Moreover, there is a polynomial time reduction of the membership problem for* $\mathrm{FRat}(\mathbb{Q}^{2\times2}, \mathrm{GL}(2,\mathbb{Z}))$ *to the membership problem for* $\mathrm{FRat}(\mathbb{Q}^{2\times2}, H)$. *For the reduction we assume that matrices in* $\mathbb{Q}^{2\times2}$ *are encoded as* 4*-tuples of rational numbers written as quotients of binary integers.*

More precisely, there is a polynomial $p(n)$ *such that the following task can be computed in* $\mathrm{DTIME}(p(n))$*: the input is a* $\mathbb{Q}^{2\times2}$*-NFA* $\mathcal{A}$*, which is flat over* $\mathrm{GL}(2,\mathbb{Z})$*. The input size is* $\|\mathcal{A}\|_{bin}$*, and the output is a* $\mathbb{Q}^{2\times2}$*-NFA* $\mathcal{B}$ *with* $\|\mathcal{B}\|_{bin} \le p(\|\mathcal{A}\|_{bin})$ *which is flat over* H *and satisfies* $L(\mathcal{B}) = L(\mathcal{A})$.

Proof. Again, it is enough to show that $\mathrm{FRat}(\mathbb{Q}^{2\times2}, \mathrm{GL}(2,\mathbb{Z})) \subseteq \mathrm{FRat}(\mathbb{Q}^{2\times2}, H)$. It is also obvious that all effectiveness assumptions stated in Theorem 6.9 are satisfied for $\mathbb{Q}^{2\times2}$, $G = \mathrm{GL}(2,\mathbb{Z})$, and $H \le \mathrm{GL}(2,\mathbb{Z})$ because $H \in \mathrm{Rec}(\mathrm{GL}(2,\mathbb{Z}))$. Since H is not part of the input, we assume that the index $[\mathrm{GL}(2,\mathbb{Z}) : H]$ and a set R of right-coset representatives is given to us in advance[17], and we can write $\mathrm{GL}(2,\mathbb{Z})$ as a disjoint union over right-cosets $\mathrm{GL}(2,\mathbb{Z}) = \bigcup_{r\in R} Hr$. Following the proof of Theorem 6.9 step by step, we see that the algorithm runs in polynomial time because

[17] In case when H is given by a finite set of generators in $\mathrm{GL}(2,\mathbb{Z})$, we can compute in a preprocessing phase the index $[\mathrm{GL}(2,\mathbb{Z}) : H]$ and a set of right-coset representatives.

addition, multiplication, and division of binary integers is possible in polynomial time. Thus, the proof of the corollary is the same as that of Theorem 6.9 by plugging in concrete complexities. ∎

In the special case of $H = \mathrm{SL}(2,\mathbb{Z})$, we have $\mathrm{GL}(2,\mathbb{Z}) = H \cup \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right) H$. So, we can replace the application of Theorem 3.8 inside the proof of Corollary 6.10 by the simpler construction of Proposition 3.1 (that was illustrating a special case for the use of Theorem 3.8).

sec:membfrat

7. The membership problem for $\mathrm{FRat}(\mathrm{GL}(2,\mathbb{Q}), S)$ with $\mathrm{GL}(2,\mathbb{Z}) \subseteq S$. The aim of Section 7 is to prove Theorems 7.1 and 7.2. With respect to decidability Theorem 7.2 is stronger than Theorem 7.1 but the known upper bounds on the complexities are different.

thm:main

THEOREM 7.1. *On input $g \in \mathrm{GL}(2,\mathbb{Q})$ and a $\mathrm{GL}(2,\mathbb{Q})$-NFA $\mathcal{A}$ that is flat over $\mathrm{GL}(2,\mathbb{Z})$, where the input size is $n = \| g \|_{bin} + \| \mathcal{A} \|_{bin}$, it is decidable whether $g \in L(\mathcal{A})$ in singly exponential time* $\mathbf{EXPTIME} = \mathrm{DTIME}(2^{n^{\mathcal{O}(1)}})$.

thm:nsmfl

THEOREM 7.2. *On input $g \in \mathrm{GL}(2,\mathbb{Q})$ and a $\mathrm{GL}(2,\mathbb{Q})$-NFA $\mathcal{A}$ that is flat over the monoid $\mathrm{GL}(2,\mathbb{Z}) \cup \{h \in \mathrm{GL}(2,\mathbb{Q}) \mid |\det(h)| > 1\}$, where the input size is $n = \| g \|_{bin} + \| \mathcal{A} \|_{bin}$, it is decidable whether $g \in L(\mathcal{A})$ in doubly exponential time* $\mathrm{DTIME}(2^{2^{n^{\mathcal{O}(1)}}})$.

The proof of Theorem 7.1 is given in Section 7.2 and the proof of Theorem 7.2 is in Section 7.3, which is a reduction to the assertion in Theorem 7.1. The main difficulty is to show decidability of the membership problem for $\mathrm{FRat}(\mathrm{GL}(2,\mathbb{Q}), \mathrm{GL}(2,\mathbb{Z}))$. The complexity follows by a careful, but straightforward, analysis of the decidability proof.

sec:memslz

7.1. The membership problem for $\mathrm{Rat}(\mathrm{GL}(2,\mathbb{Z}))$. In this subsection we consider a special instance of Theorem 7.1, where the input is an NFA $\mathcal{A}$ such that all labels of transitions are in $\mathrm{GL}(2,\mathbb{Z})$, and the problem is to decide whether $1 \in L(\mathcal{A})$. A special case of this problem was studied in [9] by Bell et al. Their main result states that the membership problem for subsemigroups of $\mathrm{GL}(2,\mathbb{Z})$ is **NP**-complete. The proof in [9] is technically demanding and quite elaborate.

BeHP2023

In Theorem 7.3, we show a pseudo-polynomial time complexity[18] for deciding whether $1 \in L(\mathcal{A})$. Our proof is rather simple and avoids compression techniques from [9]. It also keeps the paper self-contained at this point. We are mainly interested in DTIME-complexities, and **NP**-completeness means that there is little hope to find a sub-exponential deterministic decision algorithm. Note that another instance of this problem, the subgroup membership problem in $\mathrm{GL}(2,\mathbb{Z})$, was shown to be decidable in polynomial time by Lohrey in [45].

Lohrey2310cs

thm:ratsl

THEOREM 7.3. *The following problem can be decided in* $\mathrm{DTIME}(\| \mathcal{A} \|_{max}^{\mathcal{O}(1)})$.
INPUT: A $\mathrm{GL}(2,\mathbb{Z})$-NFA $\mathcal{A}$ whose unary input size is $\| \mathcal{A} \|_{max}$.
QUESTION: $1 \in L(\mathcal{A})$?

Proof. The commutator subgroup of $\mathrm{SL}(2,\mathbb{Z})$ is a free subgroup of rank 2, and it has index 24 in $\mathrm{GL}(2,\mathbb{Z})$ by [55] as we discussed in Remark 2.1. By Corollary 6.10 we can reduce in polynomial time the problem of deciding $1 \in L(\mathcal{A})$ to the special instance where all matrices are in the free subgroup $F \leq \mathrm{SL}(2,\mathbb{Z}) \leq \mathrm{GL}(2,\mathbb{Z})$. The ambient group $\mathrm{SL}(2,\mathbb{Z})$ is generated by the matrices $S = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$ of order 4 and $R = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 1 \end{smallmatrix}\right)$ of order 6. This is a well known classical result, see, for example, [21, Ch. 8.12]. The

Newman62

Jedam16

[18] The complexity of a problem involving integers is called *pseudo-polynomial* if it is polynomial time when integers are given in unary representation.

free subgroup F has a finite (and symmetric) generating set $\Sigma = \Sigma^{-1} \subseteq \mathrm{SL}(2,\mathbb{Z})$ (of size at most 48) such that each generator in Σ can be written as a product over the matrices S and R of constant length.

The inclusion $\Sigma \subseteq F$ induces a canonical homomorphism ψ of Σ^* onto F. In another polynomial time reduction with respect to the unary input size $\|\mathcal{A}\|_{\max}$, we GurevichS07 replace matrices in F by words over Σ. Using the ideas of Gurevich and Schupp in [34] for the projective linear group $\mathrm{PSL}(2,\mathbb{Z})$, it is possible to replace a matrix in F of unary size $\|F\|_{\max}$ by a word over Σ of length $\mathcal{O}(\|F\|_{\max})$. This is also explained, for example, in edam16 [21, Ch. 8.12]. Next, using more transitions, we can assume that each transition is labeled with a letter in Σ. The number of additional transitions is in $\mathcal{O}(\|\mathcal{A}\|_{\max})$. It is this step which would exponentially blow-up the NFA if we used binary representation of integers. For example, we have $\|\left(\begin{smallmatrix}1 & n\\ 0 & 1\end{smallmatrix}\right)\|_{\mathrm{bin}} \in \mathcal{O}(\log n)$, but $\left(\begin{smallmatrix}1 & n\\ 0 & 1\end{smallmatrix}\right)$ written as a word $w_n \in \Sigma^*$ has $|w_n| \in \Omega(n)$.

Formally, we obtain Σ-NFA $\mathcal{B}$ with $\psi(L(\mathcal{B})) = L(\mathcal{A})$. More details are in DiekertElder2020ijac [20, Prop. 15.4]. Having constructed $\mathcal{B}$, it remains to decide whether $1 \in \psi(L(\mathcal{B}))$. For ben69 that we use a construction of Benois in [10]. Her aim was to show that $\mathrm{Rat}(F)$ is closed under complementation. For that she transforms first an Σ-NFA $\mathcal{B}$ into another Σ-NFA $\mathcal{B}'$ such that first, $L(\mathcal{B})$ only accepts freely reduced words (these are words without any factor aa^{-1} for $a \in \Sigma$) and second, $\psi(L(\mathcal{B}')) = \psi(L(\mathcal{B}))$. Let us explain why her transformation of $\mathcal{B}$ into $\mathcal{B}'$ can be performed in polynomial time in $|\mathcal{B}|$. It uses a so-called "flooding algorithm" where it is temporarily allowed to use labels in $\Sigma \cup \{1\}$. For $\mathcal{B} = (Q, \delta, I, F)$ and $p, q \in Q$, we let $\mathcal{B}[p,q] = (Q, \delta, \{p\}, \{q\})$. As long there is a letter $a \in \Sigma$ with $aa^{-1} \in L(\mathcal{B}[p,q])$, we introduce an ε-transition $p \xrightarrow{1} q$ into $\mathcal{B}$, unless $1 \in L(\mathcal{B}[p,q])$. Next, we remove all ε-transitions by standard methods. So, each time $\mathcal{B}$ is changed the number of pairs (p,q) with $1 \in L(\mathcal{B}[p,q])$ increases. Therefore the flooding stops after at most $|Q|^2$ rounds. Once it is finished, we see that if $uaa^{-1}v$ is accepted, then uv is accepted, too. Since Σ is fixed, the time complexity of the entire transformation is polynomial in $|Q| + |\delta|$. The construction begins and ends with NFAs without ε-transitions. Since $L(\mathcal{B}') = L(\mathcal{B})$ and $1 \in L(\mathcal{B}')$ if and only if at least one initial state is final, we are done.[19] ∎

rem:lohrey11 *Remark* 7.4. If we started with an input where matrices are written in binary, then the proof of Theorem 7.3 shows decidability in **EXPTIME**. ⋄

sec:memfrat **7.2. Proof of Theorem 7.1.** The proof of Theorem 7.1 begins with an input matrix $g \in \mathrm{GL}(2,\mathbb{Q})$ and a $\mathrm{GL}(2,\mathbb{Q})$-NFA $\mathcal{A}$ which is flat over $\mathrm{GL}(2,\mathbb{Z})$. Since the input g is nonsingular, we can assume that $g = \left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$. By Corollary 6.10, we transform the NFA in polynomial time to a $\mathrm{GL}(2,\mathbb{Q})$-NFA which is flat over $\mathrm{SL}(2,\mathbb{Z})$. Thus, without restriction, the input $\mathrm{GL}(2,\mathbb{Q})$-NFA $\mathcal{A}$ is flat over $\mathrm{SL}(2,\mathbb{Z})$. The problem is to decide whether the identity matrix is accepted by $\mathcal{A}$.

If $1 \in L(\mathcal{A})$, then there is an accepting path such that the transitions outside $\mathrm{SL}(2,\mathbb{Z})$ are used t times, where t is less than the number of strongly connected components of $\mathcal{A}$. (Otherwise, the NFA $\mathcal{A}$ were not flat over $\mathrm{SL}(2,\mathbb{Z})$.) Since the contribution of every such transition to $\|\mathcal{A}\|_{\mathrm{bin}}$ is at least 2, we have $2t < 1 + |Q| + 2t \leq \|\mathcal{A}\|_{\mathrm{bin}}$ and hence $t < \|\mathcal{A}\|_{\mathrm{bin}}/2$. Thus, we can nondeterministically guess in polynomial time an initial state q_0 and a sequence of t transitions $q_{j-1} \xrightarrow{g_j} q_j$ for $1 \leq j \leq t$ such that all other transitions (which are used on that path) are labeled with matrices from $\mathrm{SL}(2,\mathbb{Z})$. We may assume that $t \geq 1$ because for $t = 1$ at least one initial state is final, and then we have a proof for $1 \in L(\mathcal{A})$.

[19] The algorithm of Benois works in a more general setting, for example, see edam16 [21, Sec. 8.9].

Using the above guess, we compute in polynomial time t subautomata $\mathcal{A}_j$ of $\mathcal{A}$ for $1 \leq j \leq t$ such that

$$1 \in L(\mathcal{A}) \iff 1 \in g_1 L(\mathcal{A}_1) g_2 L(\mathcal{A}_2) \cdots g_t L(\mathcal{A}_t).$$

Note that $\sum_j \| g_j \|_{\text{bin}} + \| \mathcal{A}_j \|_{\text{bin}}$ is bounded by a polynomial in $\| g \|_{\text{bin}} + \| \mathcal{A} \|_{\text{bin}}$. Next, we write each matrix g_j in its Smith normal form as $g_j = r_j e_j \left(\begin{smallmatrix} 1 & 0 \\ 0 & q_j \end{smallmatrix}\right) f_j$, where $0 < r_j \in \mathbb{Q}$, $e_j, f_j \in \mathrm{SL}(2,\mathbb{Z})$, and $0 \neq q_j \in \mathbb{Z}$. Let $r = \prod_{j=1}^{t} r_j$ and $q = \prod_{j=1}^{t} q_j$, then

$$1 \in g_1 L(\mathcal{A}_1) g_2 L(\mathcal{A}_2) \cdots g_t L(\mathcal{A}_t)$$

implies $r^2 q = 1$. Thus, $0 < 1/r \in \mathbb{N}$ and $0 < q \in \mathbb{N}$. For $m = 1/r$ we obtain:

{eq:rdivm} (7.1)
$$g \in L(\mathcal{A}) \iff \left(\begin{smallmatrix} m & 0 \\ 0 & m \end{smallmatrix}\right) \in e_1 \left(\begin{smallmatrix} 1 & 0 \\ 0 & q_1 \end{smallmatrix}\right) f_1 L(\mathcal{A}_1) e_2 \left(\begin{smallmatrix} 1 & 0 \\ 0 & q_2 \end{smallmatrix}\right) f_2 L(\mathcal{A}_2) \cdots e_t \left(\begin{smallmatrix} 1 & 0 \\ 0 & q_t \end{smallmatrix}\right) f_t L(\mathcal{A}_t).$$

def:Uq DEFINITION 7.5. *Let $0 \neq q \in \mathbb{N}$. Then we define two subgroups:*

$$H_{L,q} = \{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}(2,\mathbb{Z}) \mid b \equiv 0 \bmod q \} \quad \textit{and}$$
$$H_{U,q} = \{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}(2,\mathbb{Z}) \mid c \equiv 0 \bmod q \}.$$

The images of $H_{L,q}$ and $H_{U,q} \bmod q$ are the subgroups of *lower* and *upper* triangular matrices in $\mathrm{SL}(2,\mathbb{Z}/q\mathbb{Z})$, which explains the choice of letters L and U.

lem:HLq LEMMA 7.6. *The subgroups $H_{L,q}$ and $H_{U,q}$ of $\mathrm{SL}(2,\mathbb{Z})$ are conjugate in $\mathrm{GL}(2,\mathbb{Q})$:*

{eq:cqbq} (7.2)
$$\left(\begin{smallmatrix} 1 & 0 \\ 0 & q \end{smallmatrix}\right)^{-1} H_{U,q} \left(\begin{smallmatrix} 1 & 0 \\ 0 & q \end{smallmatrix}\right) = H_{L,q}.$$

Moreover their indices in $\mathrm{SL}(2,\mathbb{Z})$ are in $\mathcal{O}(q \log q)$. In particular, they are of finite index and therefore recognizable subsets in $\mathrm{SL}(2,\mathbb{Z})$ and $\mathrm{GL}(2,\mathbb{Z})$.

Proof. Equation (7.2) is straightforward since $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1/q \end{smallmatrix}\right) \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \left(\begin{smallmatrix} 1 & 0 \\ 0 & q \end{smallmatrix}\right) = \left(\begin{smallmatrix} a & bq \\ c/q & d \end{smallmatrix}\right)$. Proposition 4.4 shows the estimation of the index. ∎

lem:num LEMMA 7.7. *Let $0 \neq q \in \mathbb{Z}$ and $\gcd(b,d) = 1$. Then there are integers $1 \leq x, y < |q|$ such that $\gcd(x,y) = 1$ and $xb + yd \equiv 0 \bmod q$.*

Proof. For $|q| = 1$, the numbers $x = y = 1$ are coprime; and they satisfy $xb + yd \equiv 0 \bmod q$ because all integers are congruent modulo 1. Hence, we may assume $2 \leq |q|$.

Let P_1 be the set of primes p such that $\gcd(p,d) = 1$ and P_2 be the set of primes p such that $p \mid d$. Write $q = q_1 \cdot q_2$ such that q_i uses primes from P_i, only. For every prime p we have

{eq:hugd1} (7.3)
$$p \in P_1 \implies \gcd(p,d) = 1$$

{eq:hugd2} (7.4)
$$p \in P_2 \implies \gcd(p,b) = 1, \text{ because } \gcd(b,d) = 1.$$

Hence, d is invertible in $\mathbb{Z}/q_1\mathbb{Z}$ and b is invertible in $\mathbb{Z}/q_2\mathbb{Z}$. Therefore we can solve the following congruences.

{eq:con1} (7.5)
$$x_1 \equiv 1 \bmod q_1 \text{ and } y_1 \equiv -bd^{-1} \bmod q_1$$

{eq:con2} (7.6)
$$y_2 \equiv 1 \bmod q_2 \text{ and } x_2 \equiv -db^{-1} \bmod q_2$$

Since $\gcd(q_1, q_2) = 1$ we obtain by the Chinese remainder theorem x, y with $1 \leq x, y < |q|$ such that

$$x \equiv 1 \bmod q_1 \text{ and } x \equiv -db^{-1} \bmod q_2 \tag{7.7}$$

$$y \equiv 1 \bmod q_2 \text{ and } y \equiv -bd^{-1} \bmod q_1 \tag{7.8}$$

The congruences in (7.7) and (7.8) tell us that these x, y with $1 \leq x, y < |q|$ satisfy

$$xb + yd \equiv 0 \bmod q. \tag{7.9}$$

Indeed, the congruence in (7.9) holds $\bmod q_1$ and $\bmod q_2$, hence it holds $\bmod q$.

We claim that $\gcd(x, y, q) = 1$. To see this, let $p \mid q$. Then $p \mid q_i$ for exactly one $i \in \{1, 2\}$. Say $i = 1$, then $x \equiv 1 \bmod q_1$ implies $x \equiv 1 \bmod p$ because $p \mid q_1$. Hence, $\gcd(x, q_1) = 1$. For $i = 2$ we obtain $\gcd(y, q_2) = 1$ and therefore $\gcd(x, y, q) = 1$ since $q = q_1 q_2$, which shows the claim.

It is still possible that there is a prime p such that $p \mid x$ and $p \mid y$. However, since $\gcd(x, y, q) = 1$ such a prime p is invertible in $\mathbb{Z}/q\mathbb{Z}$. Thus,

$$\frac{x}{p}b + \frac{y}{p}d \equiv 0 \bmod q. \tag{7.10}$$

The property $\gcd(\frac{x}{p}, \frac{y}{p}, q) = 1$ is inherited. So we can make x and y smaller. Repeating this process a finite number of times, we obtain desired x and y such that $\gcd(x, y) = 1$. ∎

We use the following well-known fact based on the extended Euclidian algorithm.

LEMMA 7.8. *Given two n-bits integers a and b, we can compute in deterministic polynomial time in n integers x and y such that $ax + by = \gcd(a, b)$ with $|x|, |y| \leq \max\{|a|, |b|\}$.*

Actually, using the fact that multiplication and division of n-bits integers is possible in soft-linear time, we can give a soft-quadratic time bound for Lemma 7.8.

LEMMA 7.9. *Let $q \in \mathbb{Z}$ and $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}(2, \mathbb{Z})$ be given in binary encoding. Then for $T \in \{L, U\}$ there is a matrix $M = \left(\begin{smallmatrix} x & y \\ z & w \end{smallmatrix}\right) \in \mathrm{SL}(2, \mathbb{Z})$, where $\| M \|_{max} \leq q$, such that $g \in M \cdot H_{T,q}$. In particular, since $\| M \|_{bin} \leq \log(q)$, we can guess the matrix M nondeterministically and verify in polynomial time that $M^{-1} \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in H_{T,q}$.*

Proof. Since $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}(2, \mathbb{Z})$, the entries b and d are coprime. By Lemma 7.7 there are coprime x and y such that firstly, $xb + yd \equiv 0 \bmod q$ and secondly, $1 \leq x, y \leq |q|$. We can guess these x, y. Next, we apply Lemma 7.8 to obtain w, z in polynomial time such that firstly, $xw - yz = 1$ and secondly, $|w|, |z| \leq \max\{x, y\} \leq |q|$. We obtain $\left(\begin{smallmatrix} x & y \\ z & w \end{smallmatrix}\right) \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in H_{L,q}$. This shows the result for $T = L$. The result for $T = U$ is symmetric. ∎

Recall that we have reduced via an **NP**-reduction the problem of deciding $g \in L(\mathcal{A})$ to the problem of deciding in the notation of (7.1) the following membership problem:

$$\left(\begin{smallmatrix} m & 0 \\ 0 & m \end{smallmatrix}\right) \in e_1 \left(\begin{smallmatrix} 1 & 0 \\ 0 & q_1 \end{smallmatrix}\right) f_1 L(\mathcal{A}_1) e_2 \left(\begin{smallmatrix} 1 & 0 \\ 0 & q_2 \end{smallmatrix}\right) f_2 L(\mathcal{A}_2) \cdots e_t \left(\begin{smallmatrix} 1 & 0 \\ 0 & q_t \end{smallmatrix}\right) f_t L(\mathcal{A}_t).$$

Firstly, we will do the following preprocessing steps. We conjugate the above equation with e_1 to move it to the end of the expression. By making, if necessary, t larger and adding dummy NFAs $\mathcal{A}_i$ of constant size with $L(\mathcal{A}_i) = \{1\}$ we can assume that

$t \leq \| \mathcal{A} \|_{\text{bin}}$ with $2 \leq t \in 2^{\mathbb{N}}$. In the new notation, we let $m_t = m$ and construct (in polynomial time) NFAs $\mathcal{A}_{j,t}$, for $1 \leq j \leq t$, such that $L(\mathcal{A}_{j,t}) = f_j L(\mathcal{A}_j) e_{j+1}$ for $j < t$, and $L(\mathcal{A}_{t,t}) = f_t L(\mathcal{A}_t) e_1$. For convenience, we assume without restriction that each $\mathcal{A}_{j,t}$ is trim with a single initial state $p_{j,t}$ without incoming transition and a single outgoing transition $p_{j,t} \xrightarrow{f_j} p'_{j,t}$.

The last step finishes the preprocessing phase, and the problem becomes to decide whether

{eq:fprepr}
$$\left(\begin{smallmatrix} m_t & 0 \\ 0 & m_t \end{smallmatrix}\right) \in \left(\begin{smallmatrix} 1 & 0 \\ 0 & q_{1,t} \end{smallmatrix}\right) L(\mathcal{A}_{1,t}) \cdots \left(\begin{smallmatrix} 1 & 0 \\ 0 & q_{t,t} \end{smallmatrix}\right) L(\mathcal{A}_{t,t}). \tag{7.11}$$

We now perform at most $\log_2 t \leq \log_2 \| \mathcal{A} \|_{\text{bin}}$ rounds. In the k-th round we will have $s = t/2^{k-1}$. Each round starts with the problem:

{eq:ron}
$$\left(\begin{smallmatrix} m_s & 0 \\ 0 & m_s \end{smallmatrix}\right) \in \left(\begin{smallmatrix} 1 & 0 \\ 0 & q_{1,s} \end{smallmatrix}\right) R_{1,s} \cdots \left(\begin{smallmatrix} 1 & 0 \\ 0 & q_{s,s} \end{smallmatrix}\right) R_{s,s} \tag{7.12}$$

where $0 < m_s \in \mathbb{N}$ and $R_{i,s} = L(\mathcal{A}_{i,s})$ such that for all (i, s) we have $0 \neq q_{i,s} \in \mathbb{Z}$ and $\mathcal{A}_{i,s}$ is an $\mathrm{SL}(2,\mathbb{Z})$-NFA. In first round, we have $s = t$ and we start with the problem in (7.11). Each round will halve the number s until either s becomes 1 or we know that $g \notin L(\mathcal{A})$, for example because $m_s^2 \neq \prod_{i=1}^{s} q_{i,s}$. In such a case, we stop. In the k-th round, we perform the following steps from **1** to **10**.

1. For the sake of simplifying the notation in (7.12), we rename m_s, $R_{i,s}$, and $L(\mathcal{A}_{i,s})$ as m, R_i, and $L(\mathcal{A}_i)$, respectively. Thus, the problem in the k-th round becomes to decide whether the following holds

{eq:newQ}
$$\left(\begin{smallmatrix} m & 0 \\ 0 & m \end{smallmatrix}\right) \in \left(\begin{smallmatrix} 1 & 0 \\ 0 & q_1 \end{smallmatrix}\right) R_1 \cdots \left(\begin{smallmatrix} 1 & 0 \\ 0 & q_s \end{smallmatrix}\right) R_s. \tag{7.13}$$

   Without restriction we have $m^2 = \prod_{i=1}^{s} q_{i,s}$ because $R_i \subseteq \mathrm{SL}(2,\mathbb{Z})$ for all $1 \leq i \leq s$.
2. Assuming that the assertion in (7.13) holds, we guess for all even $2 \leq i \leq s$ matrices $M_i \in \mathrm{SL}(2,\mathbb{Z})$ such that $\| M_i \|_{\max} \leq q_i$ and $R_{i-1} \cap M_i H_{U,q_i} \neq \emptyset$. This is possible because if (7.13) holds, then such M_i's *exist* and Lemma 7.9 gives an upper bound on $\| M_i \|_{\max}$.
3. For all odd i between 1 and s, we rename in $\mathcal{A}_i$ the label f_i of the unique outgoing transition from the initial state p_i to some p'_i with $M_{i+1}^{-1} f_i$. We obtain an $\mathrm{SL}(2,\mathbb{Z})$-NFA $\widetilde{\mathcal{A}}_i$ such that $L(\widetilde{\mathcal{A}}_i) = M_{i+1}^{-1} L(\mathcal{A}_i)$ for all odd i. We do not touch the $\mathcal{A}_i$'s for even i.
4. By Lemma 7.6, we know that the index of H_{U,q_i} in $\mathrm{SL}(2,\mathbb{Z})$ is in $\mathcal{O}(|q_i| \log |q_i|)$. In particular, H_{U,q_i} is a recognizable subset of $\mathrm{SL}(2,\mathbb{Z})$. This implies that ${M_i}^{-1} R_{i-1} \cap H_{U,q_i}$ is rational in $\mathrm{SL}(2,\mathbb{Z})$. More precisely, for all even i, using Corollary 6.10 we construct in polynomial time an NFA $\mathcal{B}_i$ such that firstly, the NFA $\mathcal{B}_i$ accepts ${M_i}^{-1} R_{i-1} \cap H_{U,q_i}$ and secondly, all labels of the transitions are in H_{U,q_i}. It is also easy to see that the construction keeps the invariant that $\mathcal{B}_i$ has a unique initial state with a single outgoing transition but no incoming transition.
5. For every even i, we write

$$\begin{aligned}
(R_{i-1} \cap M_i H_{U,q_i}) \left(\begin{smallmatrix} 1 & 0 \\ 0 & q_i \end{smallmatrix}\right) &= M_i ({M_i}^{-1} R_{i-1} \cap H_{U,q_i}) \left(\begin{smallmatrix} 1 & 0 \\ 0 & q_i \end{smallmatrix}\right) \\
&= M_i \left(\begin{smallmatrix} 1 & 0 \\ 0 & q_i \end{smallmatrix}\right) \left(\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1/q_i \end{smallmatrix}\right) ({M_i}^{-1} R_{i-1} \cap H_{U,q_i}) \left(\begin{smallmatrix} 1 & 0 \\ 0 & q_i \end{smallmatrix}\right) \right) \\
&= M_i \left(\begin{smallmatrix} 1 & 0 \\ 0 & q_i \end{smallmatrix}\right) \left(\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1/q_i \end{smallmatrix}\right) L(\mathcal{B}_i) \left(\begin{smallmatrix} 1 & 0 \\ 0 & q_i \end{smallmatrix}\right) \right).
\end{aligned}$$

6. Define $K_i = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1/q_i \end{smallmatrix}\right) L(\mathcal{B}_i) \left(\begin{smallmatrix} 1 & 0 \\ 0 & q_i \end{smallmatrix}\right)$. The NFA for accepting K_i is the NFA $\mathcal{B}_i$ where every label h of a transition is replaced by $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1/q_i \end{smallmatrix}\right) h \left(\begin{smallmatrix} 1 & 0 \\ 0 & q_i \end{smallmatrix}\right)$. Since $h \in H_{U,q_i}$, the new labels belong to the subgroup H_{L,q_i} of $\mathrm{SL}(2,\mathbb{Z})$.
7. Define $R'_i = K_i \cdot R_i$ and let $g'_i = \left(\begin{smallmatrix} 1 & 0 \\ 0 & q_{i-1} \end{smallmatrix}\right) M_i \left(\begin{smallmatrix} 1 & 0 \\ 0 & q_i \end{smallmatrix}\right)$ for all even i. For each g'_i, compute its Smith normal form $g'_i = r'_i e'_i \left(\begin{smallmatrix} 1 & 0 \\ 0 & q'_i \end{smallmatrix}\right) f'_i$. Thanks to Lemma 4.1, it is possible to do in time polynomial in $n = \| g \|_{\mathrm{bin}} + \| \mathcal{A} \|_{\mathrm{bin}}$ from inputs q_{i-1}, q_i and M_i since $\| q_{i-1} \|_{\mathrm{bin}}$, $\| q_i \|_{\mathrm{bin}}$, and $\| M_i \|_{\mathrm{bin}}$ are all bounded by a polynomial in n.
8. Similar to the preprocessing phase which led to (7.11), we push the positive rationals r'_i, for each even i, to the left by multiplying both sides in (7.13) with $1/r'_i$. This yields a new positive natural number $m_{s/2}$ on the left side. Otherwise we have $g \notin R$. Since each $1/r'_i$ is pushed to the left, the new g'_i is equal to $e'_i \left(\begin{smallmatrix} 1 & 0 \\ 0 & q'_i \end{smallmatrix}\right) f'_i$.
9. We conjugate (7.13) with e'_2 to move it to the end of the expression. For even i, define $R''_i = f'_i R'_i e'_{i+2}$, where $e'_{s+2} = e'_2$. Overall, we have to verify:

$$\left(\begin{smallmatrix} m_{s/2} & 0 \\ 0 & m_{s/2} \end{smallmatrix}\right) \in \left(\begin{smallmatrix} 1 & 0 \\ 0 & q'_2 \end{smallmatrix}\right) R''_2 \cdots \left(\begin{smallmatrix} 1 & 0 \\ 0 & q'_s \end{smallmatrix}\right) R''_s.$$

   Note that the concatenation uses only even indices. We must have $m^2_{s/2} = \prod_i q'_{2i}$ since, otherwise, we have $g \notin L(\mathcal{A})$. Finally, we let $q_{i,s/2} = q'_{2i}$ and $\mathcal{A}_{i,s/2}$ be the NFA for R''_{2i}. This finishes one round of the reduction.
10. If $s/2 = 1$, then we must have $q_{1,1} = m_1^2$. Otherwise, we have $g \notin R$. If $s/2 > 1$, then we go back to step **1** with the new problem, where the new value of s becomes $s/2$.

If the procedure above terminates with $s/2 = 1$, then we end up with the problem of deciding $\left(\begin{smallmatrix} m_1 & 0 \\ 0 & m_1 \end{smallmatrix}\right) \in \left(\begin{smallmatrix} 1 & 0 \\ 0 & m_1^2 \end{smallmatrix}\right) R_{1,1}$. Due to uniqueness of the Smith normal form, we must have $m_1 = 1$ and hence the problem reduces to deciding whether $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right) \in R_{1,1}$, which can be done using Theorem 7.3.

It remains to analyze the time complexity of the procedure for the input size $n = \| g \|_{\mathrm{bin}} + \| \mathcal{A} \|_{\mathrm{bin}}$. Note that the **NP**-reduction in the preprocessing step can be replaced by a $\mathrm{DTIME}(2^{n^{\mathcal{O}(1)}})$-reduction. The main procedure stops after at most $\log_2 t \leq \log_2 n$ rounds. After each round the largest value $|q_i|$ is bounded by m^2 using the inequality in (7.13), and hence $|q_i| \in 2^{\mathcal{O}(n)}$. At every stage, in step **4**, we rely on product automata construction with an automaton of size $\mathcal{O}(|q_i| \log |q_i|)$. This requires $|q_i|^{\mathcal{O}(n)} n^{\mathcal{O}(1)}$ time. We also need to compute Smith normal forms in step **7**, which can be done in time polynomial in n. In step **2**, we used a nondeterministic guesses. In a deterministic simulation, we need to run through $\mathcal{O}(q_{max}^{4n})$ possibilities, where $q_{max} = \max\{|q_i|\}$. Hence, the reduction runs in $\mathrm{DTIME}(2^{n^{\mathcal{O}(1)}}) = \mathbf{EXPTIME}$ time.

Finally, if we reach $s = 1$, then we apply Theorem 7.3 which eventually decides whether $g \in L(\mathcal{A})$ by checking in time $\mathrm{DTIME}(2^{n^{\mathcal{O}(1)}}) = \mathbf{EXPTIME}$ whether $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right) \in L(\mathcal{A}_{1,1})$. This concludes the proof.

sec:memDET

7.3. Proof of Theorem 7.2. Recall that the statement of Theorem 7.2 says that on input $g \in \mathrm{GL}(2,\mathbb{Q})$ and a $\mathrm{GL}(2,\mathbb{Q})$-NFA $\mathcal{A}$ that is flat over the monoid $S = \mathrm{GL}(2,\mathbb{Z}) \cup \{h \in \mathrm{GL}(2,\mathbb{Q}) \mid |\det(h)| > 1\}$, it is decidable whether $g \in L(\mathcal{A})$ in time $\mathrm{DTIME}(2^{2^{n^{\mathcal{O}(1)}}})$, where input size n is defined as $n = \| g \|_{\mathrm{bin}} + \| \mathcal{A} \|_{\mathrm{bin}}$. Clearly, $g \in L(\mathcal{A}) \iff 1 \in g^{-1} L(\mathcal{A})$. Actually, since $\mathcal{A}$ is flat over S, we can construct in

polynomial time S-NFAs $\mathcal{A}_i$ and matrices $f_i \in \mathrm{GL}(2,\mathbb{Q}) \setminus S$ for $1 \le i \le \ell \in \mathcal{O}(n)$ such that

$$\left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right) \in L(\mathcal{A}_0) f_1 L(\mathcal{A}_1) \cdots f_\ell L(\mathcal{A}_\ell) \iff g \in L(\mathcal{A}). \tag{7.14}$$

Let $1 \le m \in \mathbb{N}$ be the greatest common divisor of the denominators of entries in f_i for all $1 \le i \le \ell$, which can be computed in polynomial time. Multiplying both side in (7.14) with m, we obtain:

$$\left(\begin{smallmatrix}m & 0\\ 0 & m\end{smallmatrix}\right) \in L(\mathcal{A}_0) g_1 L(\mathcal{A}_1) \cdots g_\ell L(\mathcal{A}_\ell) \iff g \in L(\mathcal{A}), \tag{7.15}$$

where all the g_i's have integer entries. In particular, $g_i \in S$ for all $1 \le i \le \ell$. Thus, in polynomial time we find an S-NFA $\mathcal{A}'$ having the property

$$\left(\begin{smallmatrix}m & 0\\ 0 & m\end{smallmatrix}\right) \in L(\mathcal{A}') \iff g \in L(\mathcal{A}). \tag{7.16}$$

Since we can replace the input size n by any polynomial in n, we assume for simplicity and without restriction that $\| m \|_{\mathrm{bin}} \le n$ and $\| h \|_{\mathrm{bin}} \le n$ whenever h appears as a label of a transition in $\mathcal{A}'$. This implies $m \le 2^n$ and $|\det(h)| \ge 1 + 2^{-2n}$ whenever $|\det(h)| > 1$. Assume $\left(\begin{smallmatrix}m & 0\\ 0 & m\end{smallmatrix}\right) \in L(\mathcal{A}')$ and let t be the maximal number of times a transition is used on an accepting path which is labeled by h, where $|\det(h)| > 1$. Since $(1 + 2^{-2n})^k > 1 + k2^{-2n}$ for all k, we obtain $t \le 2^{2n}(m^2 - 1) \le 2^{4n}$ because $m \le 2^n$. Next, we nondeterministically guess $t \le 2^{4n}$ transitions labeled by h_i with $|\det(h_i)| > 1$ and $\mathrm{GL}(2,\mathbb{Z})$-subautomata $\mathcal{A}'_i$ of $\mathcal{A}'$ for $1 \le i \le t$ such that

$$g \in L(\mathcal{A}) \iff \left(\begin{smallmatrix}m & 0\\ 0 & m\end{smallmatrix}\right) \in L(\mathcal{A}') \iff \left(\begin{smallmatrix}m & 0\\ 0 & m\end{smallmatrix}\right) \in L(\mathcal{A}'_0) h_1 L(\mathcal{A}'_1) \cdots h_t L(\mathcal{A}'_t). \tag{7.17}$$

Let $L = L(\mathcal{A}'_0) h_1 L(\mathcal{A}'_1) \cdots h_t L(\mathcal{A}'_t)$. Then we have $L \in \mathrm{FRat}(\mathrm{GL}(2,\mathbb{Q}), \mathrm{GL}(2,\mathbb{Z}))$ and the language L can be represented by some NFA $\mathcal{B}$, flat over $\mathrm{GL}(2,\mathbb{Z})$, which can be constructed in deterministic time $2^{\|\mathcal{A}\|_{\mathrm{bin}}^{\mathcal{O}(1)}}$. By Theorem 7.1, we can decide $\left(\begin{smallmatrix}m & 0\\ 0 & m\end{smallmatrix}\right) \in L(\mathcal{B})$ in deterministic time $2^{\|\mathcal{B}\|_{\mathrm{bin}}^{\mathcal{O}(1)}}$. Altogether, we obtain a deterministic doubly exponential time algorithm to decide $g \in L(\mathcal{A})$ as stated in Theorem 7.2.

8. Singular target matrices.

The aim of this section is to prove the following two theorems.

THEOREM 8.1 (The mortality problem). *Given as input a $\mathbb{Q}^{2\times 2}$-NFA $\mathcal{A}$ which is flat over the monoid generated by $\mathrm{GL}(2,\mathbb{Z}) \cup \mathbb{Q} \cup \{s_0\}$, it is decidable whether $0 \in L(\mathcal{A})$ in singly exponential time $\mathrm{DTIME}(2^{n^{\mathcal{O}(1)}})$ with respect to the input size $n = \| \mathcal{A} \|_{bin}$.*

THEOREM 8.2. *Given as inputs a matrix $g \in \mathbb{Q}^{2\times 2}$ and a $\mathbb{Q}^{2\times 2}$-NFA $\mathcal{A}$ which is flat over the monoid generated by $\mathrm{GL}(2,\mathbb{Z}) \cup \{r \in \mathbb{Q} \mid r > 1\} \cup \{0, s_0\}$, it is decidable whether $g \in L(\mathcal{A})$ in doubly exponential time $\mathrm{DTIME}(2^{2^{n^{\mathcal{O}(1)}}})$ with respect to the input size $n = \| g \|_{bin} + \| \mathcal{A} \|_{bin}$.*

The proofs of these theorems are given in Section 8.3 and Section 8.4, respectively.

8.1. Preliminary calculations.

In this section, we will use the following definitions.

DEFINITION 8.3. *For $a \in \mathbb{Z}$ we define*

$$M_{i,j}(a) = \{\left(\begin{smallmatrix}g_{11} & g_{12}\\ g_{21} & g_{22}\end{smallmatrix}\right) \in \mathrm{GL}(2,\mathbb{Q}) \cap \mathbb{Z}^{2\times 2} \mid g_{ij} = a\}.$$

For $0 \ne a \in \mathbb{Z}$ we define

$$M(a,0) = \{\left(\begin{smallmatrix}g_{11} & g_{12}\\ g_{21} & g_{22}\end{smallmatrix}\right) \in \mathrm{GL}(2,\mathbb{Q}) \cap \mathbb{Z}^{2\times 2} \mid g_{11} = a \text{ and } g_{21} = 0\}.$$

In other words, $M_{i,j}(a)$ is the set of nonsingular 2×2 integer matrices where the entry i,j is equal to a; and $M(a,0)$ is the subset of upper triangular matrices in $M_{1,1}(a)$.

prob:m33 *Problem* 8.4. INPUT: An integer $a \in \mathbb{Z}$ and a $(\mathrm{GL}(2,\mathbb{Q}) \cap \mathbb{Z}^{2\times 2})$-NFA $\mathcal{B}$ which is flat over $\mathrm{GL}(2,\mathbb{Z})$ and where the input size is $\| a \|_{\mathrm{bin}} + \| \mathcal{B} \|_{\mathrm{bin}}$.
QUESTION: $M_{i,j}(a) \cap L(\mathcal{B}) \neq \emptyset$?

prob:m36 *Problem* 8.5. INPUT: An integer $0 \neq a \in \mathbb{Z}$ and a $(\mathrm{GL}(2,\mathbb{Q}) \cap \mathbb{Z}^{2\times 2})$-NFA $\mathcal{B}$ which is flat over $\mathrm{GL}(2,\mathbb{Z})$ and where the input size is $\| a \|_{\mathrm{bin}} + \| \mathcal{B} \|_{\mathrm{bin}}$.
QUESTION: $M(a,0) \cap L(\mathcal{B}) \neq \emptyset$?

prob:gLg *Problem* 8.6. INPUT: $g \in \mathrm{GL}(2,\mathbb{Q})$ and a $\mathrm{GL}(2,\mathbb{Q})$-NFA $\mathcal{B}$ that is flat over $\mathrm{GL}(2,\mathbb{Z})$, where the input size is $\| g \|_{\mathrm{bin}} + \| \mathcal{A} \|_{\mathrm{bin}}$.
QUESTION: $g \in L(\mathcal{B})$?

Recall that Problem 8.6 is decidable in singly exponential time $\mathrm{DTIME}(2^{N^{\mathcal{O}(1)}})$ for $N = \| g \|_{\mathrm{bin}} + \| \mathcal{A} \|_{\mathrm{bin}}$ by Theorem 7.1. The reason to use the letter N here instead of n is that we will apply Theorem 7.1 later for singular matrices where N is exponential is another parameter n.

lem:freduc LEMMA 8.7. *There are* **NP***-reductions of Problems 8.4 and 8.5 to Problem 8.6. In particular, we can solve both problems in* **EXPTIME** *by Theorem 7.1.*

Proof. We begin with Problem 8.4. Let $N = \| a \|_{\mathrm{bin}} + \| \mathcal{B} \|_{\mathrm{bin}}$ be the input size, and $L(\mathcal{B}) = g_1 L(\mathcal{B}_1) g_2 L(\mathcal{B}_2) \cdots g_t L(\mathcal{B}_t)$, where $L(\mathcal{B}_j) \subseteq \mathrm{GL}(2,\mathbb{Z})$.

Note that $\mathrm{GL}(2,\mathbb{Z})$ contains the matrix $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ such that multiplying any matrix $m \in \mathbb{Q}^{2\times 2}$ with $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ on the left (resp., on the right) swaps the rows (resp., columns) of m. Hence, without restriction, we can assume that $i = j = 1$, and the problem is to decide whether $M_{1,1}(a) \cap L(\mathcal{B}) \neq \emptyset$.

Since $\mathcal{B}$ is flat over $\mathrm{GL}(2,\mathbb{Z})$, it follows that $\| D \|_{\mathrm{bin}}$ is bounded by some polynomial in N for every $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in L(\mathcal{B})$, where N is the input size. If $a = 0$, then we have $D = -bc$, and so we can guess b and c. Note that

$$\left(\begin{smallmatrix} 0 & b \\ c & d \end{smallmatrix}\right)\left(\begin{smallmatrix} 1 & x \\ 0 & 1 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 0 & b \\ c & d+cx \end{smallmatrix}\right).$$

Hence we can guess $0 \leq d' \leq |c|$ such that

$$\left(\begin{smallmatrix} 0 & b \\ c & d \end{smallmatrix}\right) \in L(\mathcal{B}) \iff \left(\begin{smallmatrix} 0 & b \\ c & d' \end{smallmatrix}\right) \in L(\mathcal{B})\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)^{\mathbb{Z}},$$

where the question "$\left(\begin{smallmatrix} 0 & b \\ c & d' \end{smallmatrix}\right) \in L(\mathcal{B})\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)^{\mathbb{Z}}$?" is an instance of Problem 8.6. Here, and in the following, $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)^{\mathbb{Z}}$ is a shortcut for $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)^{\mathbb{Z}} = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)^{*} \cup \left(\begin{smallmatrix} 1 & -1 \\ 0 & 1 \end{smallmatrix}\right)^{*}$.

Thus, we assume $a \neq 0$ for the rest of the proof. For all $x, y \in \mathbb{Z}$, a straightforward calculation shows:

$$\left(\begin{smallmatrix} 1 & 0 \\ y & 1 \end{smallmatrix}\right)\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)\left(\begin{smallmatrix} 1 & x \\ 0 & 1 \end{smallmatrix}\right) = \left(\begin{smallmatrix} a & b+ax \\ c+ay & d+cx+by+axy \end{smallmatrix}\right).$$

As a consequence, there are integers b', c', d' with $0 \leq |b'|, |c'| \leq |a|$ such that

eq:lead (8.1) $$\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in L(\mathcal{B}) \iff \left(\begin{smallmatrix} a & b' \\ c' & d' \end{smallmatrix}\right) \in \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)^{\mathbb{Z}} L(\mathcal{B}) \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)^{\mathbb{Z}}.$$

Since $a \neq 0$ and $ad' = (D + b'c')$, the binary sizes of the integers b', c', and d' are polynomially bounded in n. Thus, we can guess the integers b', c' among exponentially many candidates and compute d'. The right-hand side in (8.1) is again an instance of Problem 8.6, and we are done with Problem 8.4.

It remains to show an **NP**-reduction for Problem 8.5. Recall that the problem is to decide whether there exist $b, d \in \mathbb{Z}$ such that $\left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right) \in L(\mathcal{B})$.

Again, since $\mathcal{B}$ is flat over $\mathrm{GL}(2,\mathbb{Z})$, it follows that $\| D \|_{\mathrm{bin}}$ is bounded by some polynomial in N. Since $D = ad$ and $|D| = |g_1 \cdots g_t|$, where the g_i's are the nonsingular integer matrices defined above, we know that $d = \pm g_1 \cdots g_t / a \in \mathbb{Z}$. So there are only two options for d, and we can compute both possibilities in polynomial time if D and all g_i's are written in binary. Note that

$$\left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right) \left(\begin{smallmatrix} 1 & x \\ 0 & 1 \end{smallmatrix}\right) = \left(\begin{smallmatrix} a & b+ax \\ 0 & d \end{smallmatrix}\right).$$

Hence we can guess $0 \le b' \le |a|$ such that

$$\left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right) \in L(\mathcal{B}) \iff \left(\begin{smallmatrix} a & b' \\ 0 & d \end{smallmatrix}\right) \in L(\mathcal{B}) \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)^{\mathbb{Z}},$$

where the question "$\left(\begin{smallmatrix} a & b' \\ 0 & d \end{smallmatrix}\right) \in L(\mathcal{B}) \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)^{\mathbb{Z}}$?" is again an instance of Problem 8.6. ∎

sec:inflood

8.2. The flooding procedure. Recall that a zero-transition is a transition whose label is the zero matrix, and s_0 denotes the matrix $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$. In the following, a rank-1 transition means a transition with label $a \cdot s_0 = \left(\begin{smallmatrix} a & 0 \\ 0 & 0 \end{smallmatrix}\right)$ where $0 \neq a \in \mathbb{Z}$. (The notation is justified in our context since every $\mathbb{Z}^{2\times 2}$-matrix of rank one is in $\mathrm{SL}(2,\mathbb{Z}) \left(\begin{smallmatrix} a & 0 \\ 0 & 0 \end{smallmatrix}\right) \mathrm{SL}(2,\mathbb{Z})$ with $0 \neq a \in \mathbb{Z}$.)

The aim of this section we prove Lemma 8.8, which will be used to show Theorems 8.1 and 8.2. The key ingredient of Lemma 8.8 is the procedure FLOODING$(m, \mathcal{A})$; and we begin with an informal description. It has two parameters: a natural number $m \in \mathbb{N}$ and a $\mathbb{Z}^{2\times 2}$-NFA $\mathcal{A}$ of input size $\| \mathcal{A} \|_{\mathrm{bin}} = n$ which is flat over $\mathrm{GL}(2,\mathbb{Z}) \cup \mathbb{Z}s_0$. We rewrite in DTIME$\big(n^{\mathcal{O}(1)}\big)$ every label in its Smith normal form. This makes it possible to assume that the procedure is called only if $\mathcal{A}$ is a $\mathbb{Z}^{2\times 2}$-NFA where each label of a nonzero transition is either in $\mathrm{GL}(2,\mathbb{Q}) \cap \mathbb{Z}^{2\times 2}$ or a rank-1 matrix.

The idea of the "flooding" is to introduce more rank-1 transitions that can be used as shortcuts for accepting paths without changing the accepted language $L(\mathcal{A})$.[20] Actually, there are only three cases in the proof of Lemma 8.8. Firstly, for $m = 0$, the procedure stops as soon as a transition with the zero-matrix as a label appears. This is the witness that the zero-matrix is accepted, and hence we stop. For $m \neq 0$, we first remove all zero-transitions and we never introduce any zero-transition. In the remaining two cases, the procedure either exits with the correct output that $ms_0 \notin L(\mathcal{A})$ or, in the third cases, it transforms $\mathcal{A}$ into an NFA $\mathcal{B}$ with $L(\mathcal{B}) = L(\mathcal{A})$ such that if $ms_0 \in L(\mathcal{B})$, then it is accepted by a path in $\mathcal{B}$ which uses a rank-1 transition exactly once. Clearly, the third case is impossible for $m = 0$. The formal description is in Figure 3.

lem:flood

LEMMA 8.8. *Let $m \in \mathbb{N}$ and $\mathcal{A}$ be a $\mathbb{Z}^{2\times 2}$-NFA which is flat over $\mathrm{GL}(2,\mathbb{Z}) \cup \mathbb{Z}s_0$. Then, for $m = 0$ we have $\left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right) \in L(\mathcal{A})$ if and only if the procedure* FLOODING$(0, \mathcal{A})$ *in Figure 3 stops with that output.*

For $m \neq 0$, the procedure works as follows. It either stops and correctly outputs that $ms_0 \notin L(\mathcal{A})$ or it terminates with a trim $\mathbb{Z}^{2\times 2}$-NFA which is flat over $\mathrm{GL}(2,\mathbb{Z}) \cup \mathbb{Z}s_0$ and has the following two properties:

1. *We have $L(\mathcal{B}) = L(\mathcal{A})$.*
2. *If $ms_0 \in L(\mathcal{A})$, then ms_0 is accepted by some path where a rank-1 transition is used exactly once.*

[20] A similar idea was also used in Section 7.1 and, as mentioned there, goes back to [10]. ben69

1: **procedure** FLOODING($m, \mathcal{A}$)
2: Run the trimming procedure and denote by $\mathcal{B}$ its output.
3: If $\mathcal{B}$ contains a zero-transition and $m = 0$, then output $\left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right) \in L(\mathcal{A})$ and exit.
4: Otherwise, remove all zero-transitions in $\mathcal{B}$ and trim it again.
5: **repeat** ▷ *In the beginning of each round, there are no zero-transitions and $\mathcal{B}$ is trim.*
6: **if** there is no transition with label $a \cdot s_0$ such that $a \mid m$ **then**
7: Output $\left(\begin{smallmatrix} m & 0 \\ 0 & 0 \end{smallmatrix}\right) \notin L(\mathcal{A})$ and exit the procedure.
8: **end if**
9: Create a list $\mathcal{L}$ containing all triples (t, t', a), where $0 \neq a \in \mathbb{Z}$ and t, t' are rank-1 transitions $t = (q \xrightarrow{a' \cdot s_0} p)$ with $t' = (p' \xrightarrow{a'' s_0} q')$ such that the product $a'aa''$ divides m. ▷ *Otherwise, $a'aa''$ cannot be used as a label.*
10: **for all** $(t, t', a) \in \mathcal{L}$ **do**
11: Let $\mathcal{B}_{[p,p']}$ be a sub-automaton of $\mathcal{B}$ containing all transitions with labels from $\mathrm{GL}(2, \mathbb{Q}) \cap \mathbb{Z}^{2\times 2}$ in which p is the unique initial and p' is the unique final state. ▷ *The procedure behaves differently for $m = 0$ and $m \neq 0$.*
12: **if** $m = 0$ and $M_{1,1}(0) \cap L(\mathcal{B}_{[p,p']}) \neq \emptyset$ **then**
13: Output $\left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right) \in L(\mathcal{B})$ and exit the procedure. ▷ *We are done because $\mathcal{B}$ is trim.*
14: **else if** $m \neq 0$ and $M_{1,1}(a) \cap L(\mathcal{B}_{[p,p']}) \neq \emptyset$ **then**
15: Introduce a transition $q \xrightarrow{a'aa'' \cdot s_0} q'$ (unless it is already present in $\mathcal{B}$).
16: **end if** ▷ *See Figure 4 for an illustration.*
17: **end for**
18: **until** the NFA $\mathcal{B}$ stabilizes during the body of the repeat-loop in lines 3:–15:
19: **end procedure**

FIG. 3. *The code of the flooding procedure.* fig:flo

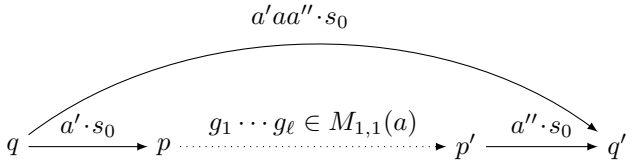

FIG. 4. *The flooding procedure introduces a rank-1 transition with label $a'aa'' \cdot s_0$* fig:flood

Moreover, the flooding procedure can be implemented in $\mathrm{DTIME}(2^{N^{\mathcal{O}(1)}})$, *where* $N = \| m \|_{bin} + \| \mathcal{A} \|_{bin}$.

Proof. Using Lemma 2.9, it is easy to see that $L(\mathcal{B}) = L(\mathcal{A})$ is an invariant throughout the procedure. If we see a zero-transition after the initial trimming then we give the correct answer for $m = 0$ in $\mathrm{DTIME}(N^{\mathcal{O}(1)})$ and we are done in this case. Thus, we may assume that we enter the repeat-loop at least once with a trim $\mathbb{Z}^{2\times 2}$-NFA $\mathcal{B}$ without zero-transitions. If all labels of transitions in $\mathcal{B}$ are invertible, we cannot accept any singular matrix. Thus, if $ms_0 \in L(\mathcal{A})$, then on every iteration of the loop, $\mathcal{B}$ must contain a transition with label as_0 such that $a \mid m$ because m and all labels are integer matrices. Thus, the procedure gives the correct answer $\left(\begin{smallmatrix} m & 0 \\ 0 & 0 \end{smallmatrix}\right) \notin L(\mathcal{A})$ whenever it exits in the body of the outer repeat-loop with a negative answer. Inside the inner for-loop, the procedure can stop and exit with another correct answer $\left(\begin{smallmatrix} m & 0 \\ 0 & 0 \end{smallmatrix}\right) \in L(\mathcal{A})$ for $m = 0$.

The considerations above handle all possible exits, and each time the answer is correct. It remains to deal with the case when there are no such exits at all. The

termination is clear because the flooding must stop eventually. We claim that every iteration of the repeat-loop shortens every accepting path for $ms_0 \in L(\mathcal{B})$.

Suppose $m \neq 0$ and $ms_0 \in L(\mathcal{B})$. Let ℓ be the minimal number of rank-1 transitions used on an accepting path for m. For $\ell = 1$ there is nothing to do. Hence we may assume $\ell \geq 2$. Let $q \xrightarrow{a's_0} p$ be the first and $p' \xrightarrow{a''s_0} q'$ be the second rank-1 transition on that path. Then $L(B_{[p,p']})$ contains some nonsingular matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, and therefore we have $a'aa''s_0 \in L(B_{[q,q']})$. We must have $a \neq 0$ and $a \mid m$ because $m \neq 0$. After that the flooding procedure can proceed by adding a new transition $q \xrightarrow{a'aa''s_0} q'$, which does not change $L(B)$ because it is just a short cut of an existing path with label $\left(\begin{smallmatrix} a'aa'' & 0 \\ 0 & 0 \end{smallmatrix}\right)$. It is indeed new because ℓ was chosen to be minimal, and with $q \xrightarrow{a'aa''s_0} q'$ we can find another path which uses less rank-1 transitions to accept ms_0 than before. Therefore, for $m \neq 0$, the flooding leads to an accepting path which uses a rank-1 transition exactly once. Note that this argument also shows that for $m = 0$ the procedure must eventually find a witness for $\left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right) \in L(\mathcal{B})$ and exit because the zero-matrix cannot be accepted by a path that uses exactly one rank-1 transition. This shows the correctness of the algorithm in Figure 3.

In order to finish the proof of Lemma 8.8, it remains to analyze its complexity. This is done as follows. Firstly, the procedure trims $\mathcal{A}$ andproduces an output $\mathcal{B}$. Trimming does not change the accepted language and does not increase the number of states or transitions. After that the procedure does not change the state set of $\mathcal{B}$ anymore. The number of states is less than N. The set of pairs (q, q') in $\mathcal{B}$ with an outgoing (resp., incoming) rank-1 transition is not changed, and hence their number is less than N^2. The number of divisors a of m is at most $\log(m) \leq \log(2^N)$. Hence, it is polynomial in N. Therefore, the number of repeat loops is bounded by a polynomial in N. Constructing the list $\mathcal{L}$ can be performed in $\mathrm{DTIME}(N^{\mathcal{O}(1)})$. Thus, it remains to show that the inner for-all-loop can by implemented to run in $\mathrm{DTIME}(2^{N^{\mathcal{O}(1)}})$. Within each loop we have to solve an instance of Problem 8.4. We can answer Problem 8.4 in $\mathrm{DTIME}(2^{N^{\mathcal{O}(1)}})$ by Lemma 8.7. Therefore, Lemma 8.8 is proved. ∎

sec:mort

8.3. Deciding the mortality problem: proof of Theorem 8.1. Recall that Theorem 8.1 says that, given as input a $\mathbb{Q}^{2\times 2}$-NFA $\mathcal{A}$ which is flat over the monoid generated by $\mathrm{GL}(2,\mathbb{Z}) \cup \mathbb{Q} \cup \{s_0\}$ of size $n = \| \mathcal{A} \|_{\mathrm{bin}}$, it is decidable in singly exponential time $\mathrm{DTIME}(2^{n^{\mathcal{O}(1)}})$ whether $\left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right) \in L(\mathcal{A})$. Without restriction, we assume that $\mathcal{A}$ is trim and does not have zero-transitions.

In a preprocessing phase, we compute in polynomial time for every transition $p \xrightarrow{h} p'$ the Smith normal form of its label as $h = e \left(\begin{smallmatrix} r & 0 \\ 0 & rq \end{smallmatrix}\right) f$, where $e, f \in \mathrm{GL}(2,\mathbb{Z})$, $q \in \mathbb{Z}$, and $0 < r \in \mathbb{Q}$. After that we replace the label h by $e \left(\begin{smallmatrix} 1 & 0 \\ 0 & q \end{smallmatrix}\right) f$ which does not change the property whether $\left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ is accepted. Splitting each transition into at most three transitions we obtain an NFA $\mathcal{A}'$ of size N which is polynomial in n such that every transition has its label in $\mathrm{GL}(2,\mathbb{Z}) \cup \{s_q \mid q \in \mathbb{Z}\}$. The NFA $\mathcal{A}'$ is a $\mathbb{Z}^{2\times 2}$-NFA which is flat over the set $\mathrm{GL}(2,\mathbb{Z}) \cup \{s_0\}$. Since $N = \| \mathcal{A}' \|_{\mathrm{bin}}$ is polynomial in n, we rename $\mathcal{A}'$ as $\mathcal{A}$ and assume that $n = N$.

After this preprocessing, we run the procedure $\textsc{Flooding}(m, \mathcal{A})$ which, assuming $n = N$, stops in time $2^{n^{\mathcal{O}(1)}}$. Recall that if the procedure did not exit with the answer $\left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right) \in L(\mathcal{A})$, then we must have $\left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right) \notin L(\mathcal{A})$ because otherwise it would accept $\left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ using a path where a rank-1 matrix appears at most once and all other labels are invertible matrices, which is impossible. Therefore, Theorem 8.1 is shown.

sec:sing

8.4. Proof of Theorem 8.2. Theorem 8.2 states that given as inputs a matrix $g \in \mathbb{Q}^{2\times 2}$ and a $\mathbb{Q}^{2\times 2}$-NFA $\mathcal{A}$ which is flat over the monoid generated by $\mathrm{GL}(2,\mathbb{Z}) \cup \{r \in \mathbb{Q} \mid r > 1\} \cup \{0, s_0\}$, it is decidable whether $g \in L(\mathcal{A})$ in $\mathrm{DTIME}(2^{2^{n^{\mathcal{O}(1)}}})$ with respect to $n = \| g \|_{\mathrm{bin}} + \| \mathcal{A} \|_{\mathrm{bin}}$. Thanks to Theorem 7.2 and Theorem 8.1, it is enough to prove Theorem 8.2 when the input g is singular but not zero. As usual, we may assume that $\mathcal{A}$ is a trim $\mathbb{Q}^{2\times 2}$-NFA without any zero-transition and which is flat over the monoid generated by $\mathrm{GL}(2,\mathbb{Z}) \cup \{r \in \mathbb{Q} \mid r > 1\} \cup \{s_0\}$. Note that the assertion of Theorem 8.2 does not change if we replace n by some $n' \in n^{\mathcal{O}(1)}$. This allows us to rename n' as n whenever convenient.

As in the proof of Theorem 8.1, we start with a preprocessing phase. We begin by computing in polynomial time the Smith normal form of the target matrix $g = e_g \left(\begin{smallmatrix} r_g & 0 \\ 0 & 0 \end{smallmatrix}\right) f_g$ with $0 < r_g \in \mathbb{Q}$. Multiplying g by the denominator of r_g and changing $\mathcal{A}$ by adding to it new initial and final transitions with labels e_g^{-1} and f_g^{-1}, respectively, we assume without restriction that $g = \left(\begin{smallmatrix} m_g & 0 \\ 0 & 0 \end{smallmatrix}\right) = m_g s_0$ with $1 < m_g \in \mathbb{N}$ and that the modified NFA is still called $\mathcal{A}$ with $n = \| \mathcal{A} \|_{\mathrm{bin}}$. Next, we compute for each transition $p \xrightarrow{h} p'$ the Smith normal form of h as $h = er \left(\begin{smallmatrix} 1 & 0 \\ 0 & q \end{smallmatrix}\right) f = ers_q f$ with $e, f \in \mathrm{GL}(2,\mathbb{Z})$, $q \in \mathbb{Z}$, and $1 < r \in \mathbb{Q}$. We also split the transition $p \xrightarrow{srs_q f} p'$ into at most 3 transitions such that all labels are either in $\mathrm{GL}(2,\mathbb{Z})$ or of the form rs_q with $0 < r \in \mathbb{Q}$ and $q \in \mathbb{Z}$. Since the Smith normal form was computed in polynomial time, we can write r as a fraction $r = n_r/m_r$ where n_r and m_r are positive natural numbers in $2^{n^{\mathcal{O}(1)}}$.

Again, we assume that the NFA is still called $\mathcal{A}$ with $n = \| \mathcal{A} \|_{\mathrm{bin}}$. Since $\mathcal{A}$ is flat over the monoid generated by $\mathrm{GL}(2,\mathbb{Z}) \cup \{r \in \mathbb{Q} \mid r > 1\} \cup \{s_0\}$, there are at most n transitions with a label $rs_q \notin \mathrm{GL}(2,\mathbb{Z})$ where $0 < r \leq 1$. Multiplying g and the labels of these transitions with appropriate positive integers in $2^{n^{\mathcal{O}(1)}}$, we may assume that $2 \leq r \in \mathbb{N}$ for all these transitions and the target matrix is changed to $g' = Km_g s_0$ with $K \in 2^{n^{\mathcal{O}(1)}}$. To simplify the notation, we rename g' as g and assume that the new automaton is called $\mathcal{A}$.

This finishes the first phase in the preprocessing. At this point we have the following situation: the target matrix g is of the form ms_0 with $1 \leq m \in \mathbb{N}$. The labels of $\mathcal{A}$ are in either in $\mathrm{GL}(2,\mathbb{Z}) \cup \{s_0\}$ or of the form rs_q with $q \in \mathbb{Z}$ and $1 + 2^{-n} \leq r \in \mathbb{Q}$, where $n = \| \mathcal{A} \|_{\mathrm{bin}}$.

For the second phase of the preprocessing, we define a subset $\mathcal{T}$ of transitions:

{eq:Ttau} (8.2) $$\mathcal{T} = \{p \xrightarrow{h} p' \mid \exists k, \ell \in \mathbb{N}\, \exists q \in \mathbb{Z} :\ h = (k/\ell)s_q \text{ and } k/\ell > 1\}.$$

Since we have $n = \| \mathcal{A} \|_{\mathrm{bin}}$, the label h of every transition in $\mathcal{A}$ satisfies $\| h \|_{\mathrm{bin}} < n$. Hence, if the label is $rs_q = \left(\begin{smallmatrix} k/\ell & 0 \\ 0 & kq/\ell \end{smallmatrix}\right)$ with $\gcd(k,\ell) = 1$, then $\| k \|_{\mathrm{bin}} + \| \ell \|_{\mathrm{bin}} < n$. Thus, $k < 2^n$ and $\ell < 2^n$. Moreover, we also have $q < 2^n$ according to the definitions in Section 2.2.

Suppose that $ms_0 \in L(\mathcal{A})$. Then there is an accepting path using t transitions $\tau_j \in \mathcal{T}$ such that all other transitions on that path are labeled by nonzero integer matrices. Recall that every τ_j has a label $r_j s_{q_j}$ with $r_j \geq 1 + 2^{-n}$. Since all other matrices on the chosen accepting path have integer entries and r_j's commute with all matrices, we obtain that $(1+2^{-n})^t \leq m < 2^n$. Since $1 + t2^{-n} \leq (1+2^{-n})^t$, we obtain $t2^{-n} \leq 2^n$, which means that $t \leq 2^{2n}$.

Next, we perform the following $\mathrm{NTIME}(2^{\mathcal{O}(n)})$-reduction which defines a $(\mathbb{Q}^{2\times 2} \setminus \{0\})$-NFA $\mathcal{A}'$ by guessing a sequence of t transitions $\tau_j \in \mathcal{T}$ with label $r_j s_{q_j}$ and $t+1$

subautomata $\mathcal{A}_j$ of $\mathcal{A}$ where all labels of transitions in $\mathcal{A}_j$ belong to $\mathbb{Z}^{2\times 2}$ such that:

$$g \in L(\mathcal{A}) \iff ms_0 \in L(\mathcal{A}') = L(\mathcal{A}_0)r_1 s_{q_1} L(\mathcal{A}_1)\cdots r_t s_{q_t} L(\mathcal{A}_t). \tag{8.3}$$

Since each $\mathcal{A}_i$ is a subautomaton of $\mathcal{A}$ which does not have transition from $\mathcal{T}$, it must be flat over $\mathrm{GL}(2,\mathbb{Z}) \cup \{s_0\}$. We also have $\| \mathcal{A}_i \|_{\mathrm{bin}} \leq n$. Recall that we have calculated each r_j as a fraction $r_j = k_j/\ell_j$ where k_j, ℓ_j are nonzero natural numbers with $k_j, \ell_j < 2^n$. Thus, in $\mathrm{DTIME}\big(2^{\mathcal{O}(n)}\big)$ we can construct a $\mathbb{Z}^{2\times 2}$-NFA $\mathcal{A}''$ such that (8.3) becomes

$$g \in L(\mathcal{A}) \iff m\Big(\prod_{j=1}^{t} \ell_j\Big)s_0 \in L(\mathcal{A}'') = L(\mathcal{A}_0)k_1 s_{q_1} L(\mathcal{A}_1)\cdots k_t s_{q_t} L(\mathcal{A}_t). \tag{8.4}$$

Note that we have $\prod_{j=1}^{t} \ell_j \leq 2^{n2^{2n}}$. Thus, $\prod_{j=1}^{t} \ell_j$ can be very large number which needs $2^{n^{\mathcal{O}(1)}}$ bits in binary notation. We conclude that we have $\| \mathcal{A}'' \|_{\mathrm{bin}} \leq N$ where $N \in \mathbb{N}$ is some computable number in $2^{\mathcal{O}(n)}$. The automaton $\mathcal{A}''$ is large, but it is a $\mathbb{Z}^{2\times 2}$-NFA which is flat over $\mathrm{GL}(2,\mathbb{Z}) \cup \{s_0\}$. Hence we can call the procedure $\textsc{Flooding}(\ell, \mathcal{A}'')$ according to Figure 3 where $\ell = m\prod_{j=1}^{t} \ell_j \in 2^{N^{\mathcal{O}(1)}}$.

Since we assumed that $ms_0 \in L(\mathcal{A})$, we can guess the automaton $\mathcal{A}'$ in (8.3) correctly and assume that $\ell s_0 \in L(\mathcal{A}'')$. The output of $\textsc{Flooding}(\ell, \mathcal{A}'')$ is a $\mathbb{Z}^{2\times 2}$-NFA $\mathcal{B}$ which is flat over $\mathrm{GL}(2,\mathbb{Z}) \cup \mathbb{Z}s_0$ such that $\ell s_0 \in L(\mathcal{B})$ if and only if ℓs_0 is accepted by some path which uses a rank-1 transition τ exactly once.

We guess $\tau = as_0$ and remove all other rank-1 transitions from $\mathcal{B}$, which yields a sub-automaton $\mathcal{B}'$ of $\mathcal{B}$. Note that if $\ell s_0 \in L(\mathcal{B}')$, then a must divide ℓ. Hence we can assume without restriction that $a = 1$. Next, we guess two sub-automata $\mathcal{C}_1$ and $\mathcal{C}_2$ of $\mathcal{B}'$ such that $\mathcal{C}_1$ and $\mathcal{C}_2$ are both $\mathbb{Z}^{2\times 2} \cap \mathrm{GL}(2,\mathbb{Q})$-NFA which are flat over $\mathrm{GL}(2,\mathbb{Z})$ and we have

$$g \in L(\mathcal{A}) \iff \ell s_0 \in L(\mathcal{C}_1)s_0 L(\mathcal{C}_2) = L(\mathcal{C}_1)s_0 \cdot s_0 L(\mathcal{C}_2). \tag{8.5}$$

Clearly, the assertion in (8.5) holds if and only if for $j \in \{1,2\}$ there are invertible $\mathbb{Z}^{2\times 2}$ matrices $\left(\begin{smallmatrix} a_j & b_j \\ c_j & d_j \end{smallmatrix}\right) \in L(\mathcal{C}_j)$ with $\left(\begin{smallmatrix} \ell & 0 \\ 0 & 0 \end{smallmatrix}\right) = \left(\begin{smallmatrix} a_1 & 0 \\ c_1 & 0 \end{smallmatrix}\right)\left(\begin{smallmatrix} a_2 & b_2 \\ 0 & 0 \end{smallmatrix}\right)$. The last equality holds if and only if $a_1a_2 = \ell$, $c_1a_2 = 0$, and $b_2a_1 = 0$. Since $\ell \neq 0$, we conclude $c_1 = 0$ and $b_2 = 0$. There are only $\log(\ell) \in N^{\mathcal{O}(1)}$ possibilities to write $a_1a_2 = \ell$ in nonzero integers a_1 and a_2. Hence we guess them and the assertion in (8.5) is equivalent to the conjunction of the following two assertions:

$$\exists b_1, d_1 \in \mathbb{Z} : \left(\begin{smallmatrix} a_1 & b_1 \\ 0 & d_1 \end{smallmatrix}\right) \in L(\mathcal{C}_1) \tag{8.6}$$

$$\exists c_2, d_2 \in \mathbb{Z} : \left(\begin{smallmatrix} a_2 & 0 \\ c_2 & d_2 \end{smallmatrix}\right) \in L(\mathcal{C}_2) \tag{8.7}$$

Using transpositions of matrices, the assertion in (8.7) is equivalent to $\exists c_2, d_2 \in \mathbb{Z} : \left(\begin{smallmatrix} a_2 & c_2 \\ 0 & d_2 \end{smallmatrix}\right) \in L(\mathcal{C}_2^T)$, where $\mathcal{C}_2^T$ is obtained by reversing the direction of all transitions, interchanging initial and final states, and by replacing every label $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ by its transposition $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)^T = \left(\begin{smallmatrix} a & c \\ b & d \end{smallmatrix}\right)$. Thus, after this observation, we only need to decide the assertion in (8.6). This is an instance of Problem 8.5 which can be decided in $\mathrm{DTIME}(2^{N^{\mathcal{O}(1)}}) \subseteq \mathrm{DTIME}(2^{2^{n^{\mathcal{O}(1)}}})$ by Lemma 8.7. This concludes the proof of Theorem 8.2.

sec:conclusion

9. Conclusion and open problems. The decidability of membership problems in group theory has a long history going back to the work of Dehn (and others) at the beginning of the 20th century. Of particular interest are the membership problems for $\mathrm{GL}(n,\mathbb{Z})$ and $\mathrm{GL}(n,\mathbb{Q})$ but as soon as $n \geq 3$ various natural decision problems become undecidable, whereas the corresponding problems remain open for $\mathrm{GL}(2,\mathbb{Q})$.

The contributions of the paper are as follows. On a conceptual level, we draw the attention to the family of flat rational sets $\mathrm{FRat}(M,S)$ of a semigroup M with respect to a subsemigroup S. By definition, $\mathrm{FRat}(M,S)$ contains $\mathrm{Rat}(S)$, and it is a subfamily of $\mathrm{Rat}(M)$. For us, the most interesting case is when $S = H$ is a group.[21] In this case $\mathrm{FRat}(M,H)$ has an inductive definition without reference to a particular presentation of M or H, see Theorem 6.4. This is a rather strong result. It has a remote analogue for finite semigroups when Schützenberger [71] characterized aperiodic semigroups by allowing the star over certain prefix codes of bounded synchronization delay.

sch76

Another main contribution is the dichotomy stated in Theorem 5.4. It shows that if a subgroup G of $\mathrm{GL}(2,\mathbb{Q})$ contains $\mathrm{GL}(2,\mathbb{Z})$ and, in addition, a diagonal but not central matrix like $\left(\begin{smallmatrix} a & 0 \\ 0 & d \end{smallmatrix}\right)$ with $|a| \neq |d|$, then G contains a Baumslag-Solitar group $\mathrm{BS}(1,q)$ with $q \geq 2$ which has infinite index in G. As a consequence, there is no hyperbolic subgroup in $\mathrm{GL}(2,\mathbb{Q})$ which has $\mathrm{GL}(2,\mathbb{Z})$ as a proper subgroup. In particular, with respect to inclusion, $\mathrm{GL}(2,\mathbb{Z})$ is a maximal virtually free group and also a maximal hyperbolic group in $\mathrm{GL}(2,\mathbb{Q})$.

We have the following natural hierarchy of decision problems in terms of their increasing complexity:

- The membership problem for f.g. subgroups.
- The membership problem for f.g. subsemigroups.
- The membership problem for rational subsets.
- Inclusion of rational subsets.

For $\mathrm{GL}(2,\mathbb{Z})$, the inclusion and hence the equality of rational subsets is decidable because the family $\mathrm{Rat}(\mathrm{GL}(2,\mathbb{Z}))$ is an effective Boolean algebra. The dichotomy implies that for any subgroup G in $\mathrm{GL}(2,\mathbb{Q})$, which is larger than $\mathrm{GL}(2,\mathbb{Z})$, either membership for rational subsets is decidable but equality of rational subsets is undecidable or, in the other case, we do not know (when the paper is written) whether membership for f.g. subgroups of G is decidable. These facts were the main motivation to define the notion of a flat rational sets. It pushes the positive decidability results for $\mathrm{GL}(2,\mathbb{Z})$ further to the relative Boolean algebra $\mathrm{FRat}(\mathrm{GL}(2,\mathbb{Q}),\mathrm{GL}(2,\mathbb{Z}))$ (and beyond if we include nonsingular matrices). Using several structural results for flat rational sets, we proved our main positive decidability results in Theorem 7.2 for nonsingular matrices and in Theorem 8.1 and Theorem 8.2 for singular matrices.

Open problems. Potential directions for future research include the following items.

- Find other applications of flat rational sets to natural membership problems. For example, when considering $\mathrm{GL}(2,k)$ where k is either an algebraic field over $\mathbb{Q}$ or a function field in one variable over a finite field.
- Let G be the subgroup of $\mathrm{GL}(2,\mathbb{Q})$ which is generated by $\mathrm{GL}(2,\mathbb{Z})$ and $\left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right)$ where p is prime. Is the subgroup membership problem for G decidable?
- Several statements of our paper contain complexity bounds but we do not know whether they are sharp. For example, Problem 8.6 is **NP**-hard, but a

[21] Recall that $\mathrm{FRat}(M,S)$ is polynomial closure of $\mathrm{Rat}(S)$ in the terminology of [71]. However, we are not aware if his concept was used for decision problems in group theory elsewhere.

sch76

proof for **NP**-completeness is still missing to date, although a recent work [9] might suggest a positive answer.

- Is the mortality problem decidable for rational subsets of $\mathbb{Q}^{2\times 2}$? This problem is equivalent to the following question: given a $\mathrm{GL}(2,\mathbb{Q})\cap\mathbb{Z}^{2\times 2}$-NFA $\mathcal{A}$, do there exist $b,c,d\in\mathbb{Z}$ such that $\left(\begin{smallmatrix}0 & b\\ c & d\end{smallmatrix}\right)\in L(\mathcal{A})$.

REFERENCES

[1] IJ. J. Aalbersberg and H. J. Hoogeboom, *Characterizations of the decidability of some problems for regular trace languages*, Math. Syst. Theory, 22 (1989), pp. 1–19.

[2] A. V. Anisimow and F. D. Seifert, *Zur algebraischen Charakteristik der durch kontext-freie Sprachen definierten Gruppen*, Elektron. Informationsv. Kybernetik, 11 (1975), pp. 695–702.

[3] A. A. Antoniou, *On Product and Sum Decompositions of Sets: The Factorization Theory of Power Monoids*, Ohio State University, Department of Mathematics, 2019.

[4] L. Babai, R. Beals, J.-Y. Cai, G. Ivanyos, and E. M. Luks, *Multiplicative equations over commuting matrices*, in Proceedings of the Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '96, Philadelphia, PA, USA, 1996, Society for Industrial and Applied Mathematics, pp. 498–507.

[5] G. Baumslag and D. Solitar, *Some two-generator one-relator non-Hopfian groups*, Bull. Amer. Math. Soc., 68 (1962), pp. 199–201.

[6] H. Behr and J. Mennicke, *A presentation of the groups PSL(2, p)*, Canadian Journal of Mathematics, 20 (1968), pp. 1432–1438.

[7] P. Bell, V. Halava, T. Harju, J. Karhumäki, and I. Potapov, *Matrix equations and Hilbert's tenth problem*, International Journal of Algebra and Computation, 18 (2008), pp. 1231–1241.

[8] P. Bell, I. Potapov, and P. Semukhin, *On the mortality problem: From multiplicative matrix equations to linear recurrence sequences and beyond*, in Proc. 44th MFCS, LIPIcs, 2019, pp. 83:1–83:15.

[9] P. C. Bell, M. Hirvensalo, and I. Potapov, *The membership problem for subsemigroups of $GL_2(\mathbb{Z})$ is NP-complete*, Information and Computation, (2023), pp. 105–132.

[10] M. Benois, *Parties rationelles du groupe libre*, C. R. Acad. Sci. Paris, Sér. A, 269 (1969), pp. 1188–1190.

[11] J. Berstel and J. Sakarovitch, *Recent results in the theory of rational sets*, in Proc. MFCS 1986, Bratislava, Czechoslovakia, J. Gruska, B. Rovan, and J. Wiedermann, eds., vol. 233 of Lecture Notes in Computer Science, Springer, 1986, pp. 15–28.

[12] R. Book and F. Otto, *String-Rewriting Systems*, Springer-Verlag, 1993.

[13] W. W. Boone, *The Word Problem*, Ann. of Math., 70 (1959), pp. 207–265.

[14] M. Cadilhac, D. Chistikov, and G. Zetzsche, *Rational subsets of Baumslag-Solitar groups*, in 47th International Colloquium on Automata, Languages, and Programming, ICALP 2020, July 8-11, 2020, Saarbrücken, Germany (Virtual Conference), A. Czumaj, A. Dawar, and E. Merelli, eds., vol. 168 of LIPIcs, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020, pp. 116:1–116:16.

[15] J. Cassaigne, V. Halava, T. Harju, and F. Nicolas, *Tighter undecidability bounds for matrix mortality, zero-in-the-corner problems, and more*, arXiv eprints, abs/1404.0644 (2014).

[16] É. Charlier and J. Honkala, *The freeness problem over matrix semigroups and bounded languages*, Inf. Comp., 237 (2014), pp. 243–256.

[17] A. H. Clifford and G. B. Preston, *The algebraic theory of semigroups*, vol. 1,2, American Mathematical Society, 1961,1967.

[18] T. Colcombet, J. Ouaknine, P. Semukhin, and J. Worrell, *On reachability problems for low-dimensional matrix semigroups*, in 46th International Colloquium on Automata, Languages, and Programming, ICALP 2019, July 9-12, 2019, Patras, Greece, C. Baier, I. Chatzigiannakis, P. Flocchini, and S. Leonardi, eds., vol. 132 of LIPIcs, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019, pp. 44:1–44:15.

[19] F. Diamond and J. Shurman, *A first course in modular forms*, vol. 228 of Graduate Texts in Mathematics, Springer-Verlag, New York, 2005.

[20] V. Diekert and M. Elder, *Solutions to twisted word equations and equations in virtually free groups*, International Journal of Algebra and Computation, 30 (2020), pp. 731–819. Based on the conference in LIPIcs.ICALP.2017.96:1–96:14.

[21] V. Diekert, M. Kufleitner, G. Rosenberger, and U. Hertrampf, *Discrete Algebraic Methods. Arithmetic, Cryptography, Automata and Groups*, De Gruyter, 2016.

[22] V. Diekert, I. Potapov, and P. Semukhin, *Decidability of membership problems for flat rational subsets of GL(2,Z) and singular matrices*, in Proc. International Symposium on Symbolic and Algebraic Computation, ISSAC '20, Kalamata, Greece, July 20-23, 2020, I. Z. Emiris and L. Zhi, eds., ACM, 2020, pp. 122–129.

[23] V. Diekert and G. Rozenberg, eds., *The Book of Traces*, World Scientific, Singapore, 1995.

[24] V. Diekert and T. Walter, *Characterizing classes of regular languages using prefix codes of bounded synchronization delay*, International Journal of Algebra and Computation, 27 (2017), pp. 561–590.

[25] C. Druţu and M. Kapovich, *Geometric Group Theory*, vol. 63 of Colloquium Publications, American Mathematical Society, Providence (RI), 2018.

[26] S. Eilenberg, *Automata, Languages, and Machines*, vol. A, Academic Press, New York and London, 1974.

[27] S. Eilenberg and M.-P. Schützenberger, *Rational sets in commutative monoids*, Journal of Algebra, 13 (1969), pp. 173–191.

[28] C. Faith, *Dedekind finite rings and a theorem of Kaplansky*, Communications in Algebra, 31 (2003), pp. 4175–4178.

[29] A. Finkel and M. Praveen, *Verification of flat FIFO systems*, in Proc. 30th CONCUR 2019, August 27-30, 2019, Amsterdam, The Netherlands, W. J. Fokkink and R. van Glabbeek, eds., vol. 140 of LIPIcs, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019, pp. 12:1–12:17.

[30] E. M. Freden and T. Knudson, *Recent growth results*, in Groups St. Andrews 2005. Vol. 1, vol. 339 of London Math. Soc. Lecture Note Ser., Cambridge Univ. Press, Cambridge, 2007, pp. 341–355.

[31] C. Frougny, J. Sakarovitch, and P. Schupp, *Finiteness conditions on subgroups and formal language theory*, Proc. London Math. Soc. (3), 58 (1989), pp. 74–88.

[32] S. M. Gersten, *Dehn functions and l_1-norms of finite presentations*, in Algorithms and classification in combinatorial group theory (Berkeley, CA, 1989), vol. 23 of Math. Sci. Res. Inst. Publ., Springer, New York, 1992, pp. 195–224.

[33] Z. Grunschlag, *Algorithms in Geometric Group Theory*, PhD thesis, University of California, 1999.

[34] Y. Gurevich and P. Schupp, *Membership problem for the modular group*, SIAM J. Comput., 37 (2007), pp. 425–459.

[35] G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, Oxford University Press, Oxford, sixth ed., 2008. Revised by D. R. Heath-Brown and J. H. Silverman, With a foreword by Andrew Wiles.

[36] T. Harju, *Post correspondence problem and small dimensional matrices*, in 13th International Conference, DLT 2009, Stuttgart, Germany, June 30–July 3, 2009, Proceedings, V. Diekert and D. Nowotka, eds., vol. 5583 of Lecture Notes in Computer Science, Springer-Verlag, 2009, pp. 39–46.

[37] T. Herbst, *On a subclass of context-free groups*, RAIRO-Theor. Inf. Appl., 25 (1991), pp. 255–272.

[38] J. E. Hopcroft and J. D. Ullman, *Introduction to Automata Theory, Languages and Computation*, Addison-Wesley, 1979.

[39] A. G. Howson, *On the intersection of finitely generated free groups*, J. London Math. Soc., 29 (1954), pp. 428–434.

[40] R. Kannan and A. Bachem, *Polynomial algorithms for computing the Smith and Hermite normal forms of an integer matrix*, SIAM, 8 (1979), pp. 499–507.

[41] S. C. Kleene, *Representation of events in nerve nets and finite automata*, in Automata Studies, C. E. Shannon and J. McCarthy, eds., no. 34 in Annals of Mathematics Studies, Princeton University Press, 1956, pp. 3–40.

[42] S. Ko, R. Niskanen, and I. Potapov, *On the identity problem for the special linear group and the heisenberg group*, in 45th International Colloquium on Automata, Languages, and Programming, ICALP 2018, July 9-13, 2018, Prague, Czech Republic, I. Chatzigiannakis, C. Kaklamanis, D. Marx, and D. Sannella, eds., vol. 107 of LIPIcs, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018, pp. 132:1–132:15.

[43] D. König, M. Lohrey, and G. Zetzsche, *Knapsack and subset sum problems in nilpotent, polycyclic, and co-context-free groups*, in Algebra and Computer Science, vol. 677 of Contemporary Mathematics, AMS, 2016.

[44] A. Krieg, *Hecke Algebras*, no. 435 in American Mathematical Society: Memoirs of the American Mathematical Society, Vol. 87, Providence, Rhode Island, United States, 1990.

[45] M. Lohrey, *Subgroup membership in GL(2,Z)*, Theory of Computing Systems, (2023).

[46] M. Lohrey and G. Sénizergues, *Theories of HNN-extensions and amalgamated products*, in ICALP, M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener, eds., vol. 4052 of Lecture Notes in Computer Science, Springer, 2006, pp. 504–515.

[47] M. Lohrey and B. Steinberg, *The submonoid and rational subset membership problems for graph groups*, Journal of Algebra, 320 (2008), pp. 728–755.

[48] D. D. Long and A. W. Reid, *Small subgroups of $SL(3,\mathbb{Z})$*, Experimental Mathematics, 20 (2011), p. 412–425.

[49] R. Lyndon and P. Schupp, *Combinatorial Group Theory*, Classics in Mathematics, Springer, 2001. First edition 1977.

[50] A. A. Markov, *On certain insoluble problems concerning matrices*, Dokl. Akad. Nauk SSSR, 57 (1947), pp. 539–542.

[51] A. A. Markov, *On the impossibility of certain algorithms in the theory of associative systems*, Dokl. Akad. Nauk SSSR, 55 (1947), pp. 587–590.

[52] J. D. McKnight, *Kleene quotient theorem*, Pacific J. of Mathematics, (1964), pp. 1343–1352.

[53] K. A. Mihailova, *The occurrence problem for direct products of groups*, Dokl. Akad. Nauk SSSR, 119 (1958), pp. 1103–1105. English translation in: Math. USSR Sbornik, 70: 241–251, 1966.

[54] N. Möller, *On Schönhage's algorithm and subquadratic integer GCD computation*, Math. Comp., 77 (2008), pp. 589–607.

[55] M. Newman, *The structure of some subgroups of the modular group*, Illinois J. Math., 6 (1962), pp. 480–487.

[56] P. S. Novikov, *On the algorithmic unsolvability of the word problem in group theory*, Trudy Mat. Inst. Steklov, (1955), pp. 1–143. In Russian.

[57] C.-F. Nyberg-Brodda, *Non-finitely generated maximal subgroups of context-free monoids*, Journal of Algebra, 616 (2023), pp. 227–238.

[58] Ch. H. Papadimitriou, *Computational Complexity*, Addison Wesley, 1994.

[59] M. S. Paterson, *Unsolvability in 3×3 matrices*, Stud. Appl. Mathematics, 49 (1970), pp. 105–107, https://onlinelibrary.wiley.com/doi/10.1002/sapm1970491105.

[60] I. Potapov, *Reachability problems in matrix semigroups*, Dagstuhl Reports, 9 (2019), pp. 95–98.

[61] I. Potapov and P. Semukhin, *Decidability of the membership problem for 2×2 integer matrices*, in Proc. 28th SODA, 2017, pp. 170–186.

[62] I. Potapov and P. Semukhin, *Membership problem in GL(2, $\mathbb{Z}$) extended by singular matrices*, in Proc. 42nd MFCS, 2017, pp. 44:1–44:13.

[63] M. Presburger, *Über die Vollständigkeit eines gewissen Systems der Arithmetik ganzer Zahlen, in welchem die Addition als einzige Operation hervortritt*, Comptes Rendus du I congrès de Mathématiciens des Pays Slaves, (1929), pp. 92–101. English translation by Ryan Stansifer avalable at https://ecommons.cornell.edu/items/7ea46cc7-5938-4cfc-8c94-bc0baba39152.

[64] J. L. Rhodes and B. Steinberg, *The $\mathfrak{q}$-theory of finite semigroups.*, Springer Monographs in Mathematics, Springer, 2009.

[65] N. S. Romanovskiĭ, *Some algorithmic problems for solvable groups*, Algebra i Logika, 13 (1974), pp. 26–34, 121.

[66] J. Sakarovitch, *The "last" decision problem for rational trace languages*, in Proc. 1st Latin American Symposium on Theoretical Informatics (LATIN'92), I. Simon, ed., vol. 583 of Lecture Notes in Computer Science, Heidelberg, 1992, Springer-Verlag, pp. 460–473.

[67] A. Schönhage, *Schnelle Berechnung von Kettenbruchentwicklungen*, Acta Informatica, (1971), pp. 139–144.

[68] A. Schönhage and V. Strassen, *Schnelle Multiplikation großer Zahlen*, Computing, 7 (1971), pp. 281–292.

[69] O. Schreier, *Die Untergruppen der freien Gruppen*, Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg, 5 (1927), pp. 161–183.

[70] M.-P. Schützenberger, *Sur les monoides finis dont les groupes sont commutatifs*, Rev. Française Automat. Informat. Recherche Opérationnelle Sér. Rouge, 8 (1974), pp. 55–61.

[71] M.-P. Schützenberger, *Sur le produit de concaténation non ambigu*, Semigroup Forum, 13 (1976), pp. 47–75.

[72] G. Sénizergues, *On the rational subsets of the free group*, Acta Informatica, 33 (1996), pp. 281–296.

[73] J.-P. Serre, *Trees*, Springer, 1980. French original 1977.

[74] P. V. Silva, *Recognizable subsets of a group: finite extensions and the abelian case*, Bulletin of the EATCS, 77 (2002), pp. 195–215.

[75] J. A. Todd and H. S. M. Coxeter, *A practical method for enumerating cosets of a finite abstract group*, Proceedings of the Edinburgh Mathematical Society, 5 (1936), pp. 26–34.

[76] Y.-Y. Tran, *Computably Enumerable Boolean Algebras*, PhD thesis, Cornell University, Dept. Mathematics, 2018, https://ecommons.cornell.edu/items/825e1b2b-4f48-4df7-8cfa-b6716ea96459.