The Asymptotic Size of Finite Irreducible Semigroups of Rational Matrices

Stefan Kiefer
Department of Computer Science, University of Oxford, UK

Andrew Ryzhikov
University of Warsaw, Poland

Abstract

We study finite semigroups of $n \times n$ matrices with rational entries. Such semigroups provide a rich generalization of transition monoids of unambiguous (and, in particular, deterministic) finite automata. In this paper we determine the maximum size of finite semigroups of rational $n \times n$ matrices, with the goal of shedding more light on the structure of such matrix semigroups.

While in general such semigroups can be arbitrarily large in terms of n, a classical result of Schützenberger from 1962 implies an upper bound of $2^{\mathcal{O}(n^2 \log n)}$ for irreducible semigroups, i.e., the only subspaces of $\mathbb{Q}^n$ that are invariant for all matrices in the semigroup are $\mathbb{Q}^n$ and the subspace consisting only of the zero vector. Irreducible matrix semigroups can be viewed as the building blocks of general matrix semigroups, and as such play an important role in mathematics and computer science. From the point of view of automata theory, they generalize strongly connected automata.

Using a very different technique from that of Schützenberger, we improve the upper bound on the cardinality to 3^{n^2}. This is the main result of the paper. The bound is in some sense tight, as we show that there exists, for every n, a finite irreducible semigroup with $3^{\lfloor n^2/4 \rfloor}$ rational matrices. Our main result also leads to an improvement of a bound, due to Almeida and Steinberg, on the mortality threshold. The mortality threshold is a number ℓ such that if the zero matrix is in the semigroup, then the zero matrix can be written as a product of at most ℓ matrices from any subset that generates the semigroup.

2012 ACM Subject Classification Theory of computation → Formal languages and automata theory; Computing methodologies → Symbolic and algebraic manipulation

Keywords and phrases finite matrix semigroups, irreducible matrix semigroups, matrix mortality, aperiodic semigroups, unambiguous automata, transition monoids

Digital Object Identifier 10.4230/LIPIcs.STACS.2026.60

Related Version *Full Version*: `https://arxiv.org/abs/2601.01236` [14]

Funding *Andrew Ryzhikov*: Supported by Polish National Science Centre SONATA BIS-12 grant number 2022/46/E/ST6/00230.

1 Introduction

Given a finite set $\mathcal{A}$ of $n \times n$ matrices, the semigroup generated by $\mathcal{A}$ is the set of all possible products of matrices from $\mathcal{A}$. Matrix semigroups appear naturally in many areas of computer science, such as automata theory, dynamical systems, program analysis and formal verification. Let us illustrate that with the following two examples.

Two examples. Consider the linear loop in Figure 1 (left), where the conditions for both exiting the loop and choosing one of the two conditional branches are abstracted out and denoted by $*$. We assume that, in each iteration, one of the two linear operators is nondeterministically applied to the vector $\binom{x}{y}$ of variables. The overall set of linear operators that can be applied to this vector during the execution of the loop is thus the matrix semigroup generated by $\mathcal{A}_\ell$. This can be seen as the set of all behaviours of the loop.

43rd International Symposium on Theoretical Aspects of Computer Science (STACS 2026).
Editors: Meena Mahajan, Florin Manea, Annabelle McIver, and Nguyễn Kim Thắng
Article No. 60; pp. 60:1–60:20
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

```
while (*)
  if (*)
    x = -y
    y = x - y
  else
    x = x - y
    y = -y
```

$$\mathcal{A}_\ell = \left\{ \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}, \begin{pmatrix} 1 & -1 \\ 0 & -1 \end{pmatrix} \right\}$$

$$\mathcal{A}_r = \left\{ \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix} \right\}$$

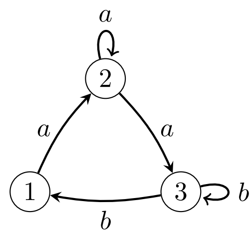


Figure 1 Left: a linear loop with nondeterministic branching. Right: an NFA. Centre: generating sets of the corresponding matrix semigroups.

Similarly, the set of all possible behaviours of a nondeterministic finite automaton (NFA) is represented by its transition monoid. Note that the transition monoid does not depend on initial and final states. In this context, an NFA is called unambiguous if for every pair p, q of states, each word labels at most one path from p to q in it. Clearly, every deterministic NFA is unambiguous. The transition monoid of an unambiguous NFA can be seen as the monoid generated by the set of transition matrices of the letters ($\mathcal{A}_r$ in the example in Figure 1 (right)) *with the usual addition and multiplication of the integers*. This fact allows us to consider unambiguous NFAs as automata with multiplicities (or, more generally, weighted automata), which significantly extend the variety of applicable techniques [13].

General motivation. In algorithmic applications, there is always a trade-off between the expressiveness of a model and the tractability of deciding its properties. This is especially important for matrix semigroups: for example, the question whether the semigroup generated by a given set of matrices contains the zero matrix is undecidable already for 3×3 integer matrices [20]. For such problems, the known decidable special cases are usually obtained by restricting the dimension [8, 3, 9] or considering only matrices with nonnegative entries [22].

In this paper, we consider a different restriction: finiteness of the generated semigroup. It constitutes a "middle ground" between the two applications above. From the loop analysis point of view, it describes loops that have a finite set of behaviours regardless of the initial values of the variables. On the other hand, finite rational matrix semigroups closely correspond to weighted automata over $\mathbb{Q}$ that have a finite image set (see, e.g., [7, Section 4.1]), a rich generalization of unambiguous and deterministic NFAs. Finite matrix semigroups are also building blocks of noncommutative power series of polynomial growth [5, Section 9.2]. We remark that both $\mathcal{A}_\ell$ and $\mathcal{A}_r$ from Figure 1 generate finite semigroups.

The setting of the paper. The general motivation behind the paper is to shed light on the structure of finite semigroups of $n \times n$ rational matrices and to develop new tools for analyzing them. The concrete question we pursue here is about the maximum size of such semigroups in terms of n. In general, there is no upper bound because for each $m \in \mathbb{N}$ the set $S_m \coloneqq \left\{ \left(\begin{smallmatrix} 0 & i \\ 0 & 0 \end{smallmatrix}\right) \mid 0 \le i < m \right\}$ forms a semigroup of size m. However, intuitively, this example is in a sense degenerate, since only the same one-dimensional subspace is affected by the corresponding linear operators. Matrix semigroups where such degenerate behaviour does not occur are called irreducible (see the next section for the formal definition). They are actively studied in representation theory of finite monoids [25, Chapter 5] and can be viewed as a generalization of the concept of strongly connected finite automata to the case of weighted automata, as argued in [16].

In some applications, such as matrix mortality, a matrix semigroup can be directly analyzed by decomposing it into irreducible semigroups of smaller dimension, see e.g. [2, Section 5] and the proof of Theorem 29 in the appendix of [14]. More generally, it is announced in [16] that every rational matrix semigroup can be decomposed into irreducible parts, resembling a simultaneous Jordan normal form (which does not always exist for multiple matrices), and that irreducibility can be decided in polynomial time. We note that the results announced in [16] do not improve the old bound of Schützenberger [24, 5] on the size of *irreducible* finite matrix semigroups, and thus do not overlap with the present paper.

2 Existing results and our contributions

As usual, we denote by $\mathbb{N}$, $\mathbb{Z}$ and $\mathbb{Q}$ the sets of natural, integer and rational numbers, respectively. We write $\mathrm{GL}_n(\mathbb{Q})$ and $\mathrm{GL}_n(\mathbb{Z})$ for the multiplicative group of all invertible $n \times n$ matrices over $\mathbb{Q}$ and $\mathbb{Z}$, respectively. We denote by $\vec{v}$ a column vector of appropriate dimension, by $\vec{0}$ a zero column vector, by $A^\top$ the transpose of a matrix A, and by O_n and I_n the $n \times n$ zero and identity matrix, respectively. We assume that all vector spaces are over the field $\mathbb{Q}$ and in particular that all matrices have rational entries, unless explicitly stated otherwise.

The maximal order of finite matrix groups. Let us first highlight the importance of rational entries in our setting. Indeed, every cyclic group is isomorphic to a group generated by a 2×2 real rotation matrix, so there is no hope of bounding the size of finite matrix groups with real entries. The case of rational entries is however very different, and the maximal order of rational finite matrix groups is well understood. By a folklore result (see, e.g., [15, Theorem 1.6]), any finite subgroup of $\mathrm{GL}_n(\mathbb{Q})$ is conjugate to a finite subgroup of $\mathrm{GL}_n(\mathbb{Z})$. An elementary proof shows that the order of any finite subgroup of $\mathrm{GL}_n(\mathbb{Z})$ divides $(2n)!$; see, e.g., [18, Chapter IX]. Thus, denoting the order of the largest finite subgroup of $\mathrm{GL}_n(\mathbb{Q})$ by $g(n)$, we have $g(n) \leq (2n)!$. It is shown in a paper by Friedland [11] that $g(n) = 2^n n!$ holds for all sufficiently large n. This bound is attained by the group of signed permutation matrices (that is, matrices with entries in $\{-1, 0, 1\}$ with exactly one nonzero entry in each row and each column). Friedland's proof rests on an article by Weisfeiler [27] which in turn is based on the classification of finite simple groups. Feit showed in an unpublished manuscript [10] that $g(n) = 2^n n!$ holds if and only if $n \in \mathbb{N} \setminus \{2, 4, 6, 7, 8, 9, 10\}$; see also [4, Table 1] for a list of the maximal-order finite subgroups of $\mathrm{GL}_n(\mathbb{Q})$ for $n \in \{2, 4, 6, 7, 8, 9, 10\}$. Feit's proof relies on an unpublished manuscript [26] (also based on the classification of finite simple groups), which Weisfeiler left behind before his tragic disappearance.

The maximal size of finite rational matrix semigroups. In view of the set S_m from the previous section, bounds on the size of finite rational matrix semigroups either need to involve the number of generators (see, e.g., [7] and the references therein, as well as its announced improvement in [16]) or an irreducibility assumption. A semigroup $S \subseteq \mathbb{Q}^{n \times n}$ is called *irreducible* if the only vector spaces $\mathcal{V} \subseteq \mathbb{Q}^n$ such that $X\mathcal{V} \subseteq \mathcal{V}$ for all $X \in S$ are $\mathcal{V} = \mathbb{Q}^n$ and $\mathcal{V} = \{\vec{0}\}$. The semigroup S_m from above is not irreducible because for the vector space $\mathcal{V} = \{\left(\begin{smallmatrix} x \\ 0 \end{smallmatrix}\right) \mid x \in \mathbb{Q}\}$ we have $X\mathcal{V} = \{\left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right)\} \subseteq \mathcal{V}$ for all $X \in S_m$.

Let us mention that the notion of irreducible matrices from nonnegative matrix theory, as in, e.g., [17], is weaker. Following [17], a square matrix with nonnegative entries is called irreducible if permuting its rows and columns cannot result in a matrix of the shape $\left(\begin{smallmatrix} A & B \\ 0 & C \end{smallmatrix}\right)$, where A and C are square matrices. A digraph is strongly connected if and only if its adjacency matrix is irreducible in this sense. Clearly, a matrix generating an irreducible semigroup in our sense must be irreducible in the sense of [17], but the converse is not always true.

In the book by Berstel and Reutenauer [5, Lemma IX.1.2] it is shown that if a semigroup $S \subseteq \mathbb{Q}^{n\times n}$ is finite and irreducible then $|S| \le (2n+1)^{n^2} \in 2^{\mathcal{O}(n^2 \log n)}$. The technique, due to Schützenberger [24], is based on the analysis of the coefficients of characteristic polynomials of the matrices in S, and in particular of their traces. In fact, the quantity $2n+1$ in the bound $(2n+1)^{n^2}$ corresponds to the possible traces in the set $\{-n, -n+1, \ldots, n\}$.

The diameter and the mortality threshold. Let $S \subseteq \mathbb{Q}^{n\times n}$ be a finite semigroup generated by a subset $S_0 \subseteq S$. The length of a shortest product of elements from S_0 resulting in $X \in S$ is called the *depth* of X. The *diameter* of S is the maximum depth among all $X \in S$. Both the depth and the diameter are implicitly defined with respect to the set S_0 of generators. Intuitively, the diameter indicates how fast one can reach any matrix. It is easy to see that the diameter of a finite semigroup cannot exceed its size.

In 2020 it was shown by Bumpus et al. [7], without assuming that S is irreducible, that the diameter of S with respect to any generating set is at most $2^{n(2n+3)}g(n)^{n+1} \in 2^{\mathcal{O}(n^2 \log n)}$, where g is the above-mentioned group-bound function with $g(n) \le (2n)!$. The technique used in [7] is not based on traces but on exterior algebra. Nevertheless, their bound of $2^{\mathcal{O}(n^2 \log n)}$ is strikingly similar to the aforementioned bound on semigroup cardinality. Panteleev [19] showed that for every n there exists a semigroup of diameter $2^{n+\Theta(\sqrt{n \log n})}$ with respect to some generating set. This semigroup is actually constructed as the transformation monoid of a deterministic finite automaton, and thus consists of matrices with entries in $\{0, 1\}$ and exactly one nonzero entry in every row. As far as the authors know, no better lower bound is known for the maximum diameter of finite rational matrix semigroups.

The depth of the zero matrix (again, with respect to a set S_0 of generators) is called *the mortality threshold* of S. Intuitively, it indicates how fast one can reach the zero matrix. Using a variation of the aforementioned technique due to Schützenberger [24, 5], Almeida and Steinberg [2] showed that the mortality threshold of any finite rational matrix semigroup containing the zero matrix is at most $(2n-1)^{n^2}$ for $n \ge 2$. This bound is once again $2^{\mathcal{O}(n^2 \log n)}$, as in [24, 5, 7]. The best known lower bound of $\Theta(n^2)$ on the mortality threshold of finite semigroups of rational matrices is due to Rystsov [21]. He conjectured that $\mathcal{O}(n^2)$ is also the upper bound, which to the best of our knowledge has not been disproved. It is noteworthy that the lower bound again comes from the transition monoid of a DFA.

Our contributions and organization of the paper. Our main result, Theorem 28, is that any finite irreducible semigroup $S \subseteq \mathbb{Q}^{n\times n}$ has at most 3^{n^2} elements, thus “breaking” the $2^{\mathcal{O}(n^2 \log n)}$ barrier in previous results about both the cardinality and the diameter [24, 5, 2, 7]. This is in a sense tight: as we show in Proposition 30, any such bound needs to be at least $3^{\lfloor n^2/4 \rfloor}$. Recall that $|S|$ cannot be bounded purely in n without assuming irreducibility. To showcase our technique, early on we give a relatively direct proof of the fact that if S is also aperiodic (i.e., every subgroup of S has only one element), then $|S| \le 2^{n^2}$ (Theorem 13). This result follows already from the proof of [2, Theorem 5.8], which used a different technique. We also provide a lower bound of $2^{\lfloor n^2/4 \rfloor}$ for the aperiodic case (Proposition 31). Finally, as an application we show that if a finite, not necessarily irreducible, matrix semigroup contains the zero matrix, then its mortality threshold is at most 3^{n^2} (Theorem 29). This improves the result by Almeida and Steinberg [2].

The paper is structured as follows. In Section 3 we establish basic facts about finite irreducible matrix semigroups and their (0-)minimal ideals. In Section 4 we explain the construction of a group "at" (i.e., corresponding to) an idempotent from the (0-)minimal ideal. The results of Sections 3 and 4 are mostly known. Sections 5 and 6 are dedicated to the proof of our main result (which we outline in the next paragraph below). Its application to matrix mortality can be found at the end of Section 6. In Section 7 we provide lower bounds. We conclude the paper by highlighting some open problems in Section 8. Missing proofs can be found in the appendix of [14].

Technique. In contrast to [24, 2, 7], our technique is based neither on traces nor on exterior algebra. In fact, although the overall proof is non-trivial, it does not use anything outside of basic (semi)group theory and linear algebra. We outline our approach in the following.

Let S be a finite rational $n \times n$ matrix semigroup, and let T be a (0-)minimal ideal of S. One can show that all matrices in $T \setminus \{O_n\}$ have the same rank, say r, which is the minimum nonzero rank in S. Given an idempotent $E \in T \setminus \{O_n\}$, fundamental semigroup theory (see [14, Appendix A] for the background we need) describes a finite subgroup G of $\mathrm{GL}_r(\mathbb{Q})$ (often, e.g., in [1], called the maximal subgroup at E), which reflects the symmetries in T (Sections 3 and 4). As discussed above, the asymptotic size of such matrix groups is well understood.

We then construct an injective map Ψ from S to tuples of elements of $G \cup \{O_r\}$ (Section 5.1). Thus, we have $|S| = |\Psi(S)|$ and so it suffices to bound the number of distinct tuples over $G \cup \{O_r\}$. This immediately leads to $|S| \leq 3^{r^2 n^2}$, a bound that does not improve on $2^{\mathcal{O}(n^2 \log n)}$ unless the minimum rank r is constant. However, in this way we already obtain a near-optimal bound on the cardinality of aperiodic semigroups (Section 5.2).

To strengthen the bound in the general case, we then show that the tuple elements are in a sense "coupled" via small matrix groups. This part (Sections 6.1–6.3) is the technical core of the paper and the most delicate aspect of the proof, even though it uses only elementary linear algebra. Finally, the overall bound of 3^{n^2} is obtained by carefully counting the coupled tuples $\Psi(X)$ within a two-dimensional grid (Section 6.4).

3 Irreducible matrix semigroups

Let $n \in \mathbb{N}$ and let $S \subseteq \mathbb{Q}^{n \times n}$ be a semigroup. A vector space $\mathcal{V} \subseteq \mathbb{Q}^n$ is called S-*invariant* if $S\mathcal{V} \subseteq \mathcal{V}$, i.e., $X\mathcal{V} \subseteq \mathcal{V}$ for all $X \in S$. The semigroup S is called *irreducible* if the only S-invariant subspaces of $\mathbb{Q}^n$ are $\mathbb{Q}^n$ and $\{\vec{0}\}$. The definition of irreducibility means that there are only trivial S-invariant "column" subspaces. But it implies that there are also only trivial S-invariant "row" subspaces:

▶ **Proposition 1.** *Let $\mathcal{U} \subseteq \mathbb{Q}^{1 \times n}$ be a (row) vector space such that $\mathcal{U}S \subseteq \mathcal{U}$. Then $\mathcal{U} = \mathbb{Q}^{1 \times n}$ or $\mathcal{U} = \{\vec{0}^\top\}$.*

Proof. Suppose that $\mathcal{U} \neq \mathbb{Q}^{1 \times n}$, i.e., $\mathcal{U}$ is a proper subspace. Define $\mathcal{U}^\circ := \{\vec{v} \in \mathbb{Q}^n \mid \mathcal{U}\vec{v} = \{\vec{0}\}\}$. Then $\dim \mathcal{U} + \dim \mathcal{U}^\circ = n$. Since $\dim \mathcal{U} < n$, we have $\dim \mathcal{U}^\circ > 0$, i.e., $\mathcal{U}^\circ \neq \{\vec{0}\}$. Let $X \in S$ and let $\vec{v} \in \mathcal{U}^\circ$. Since $\mathcal{U}X \subseteq \mathcal{U}$, we have $\mathcal{U}X\vec{v} \subseteq \mathcal{U}\vec{v} = \{\vec{0}\}$, as $\vec{v} \in \mathcal{U}^\circ$. Thus, $X\vec{v} \in \mathcal{U}^\circ$. Since $\vec{v} \in \mathcal{U}^\circ$ was arbitrary, it follows that $X\mathcal{U}^\circ \subseteq \mathcal{U}^\circ$. Since $X \in S$ was arbitrary, $\mathcal{U}^\circ$ is S-invariant. Since $\mathcal{U}^\circ \neq \{\vec{0}\}$ and S is irreducible, we have $\mathcal{U}^\circ = \mathbb{Q}^n$. Since $\dim \mathcal{U} + \dim \mathcal{U}^\circ = n$, it follows that $\mathcal{U} = \{\vec{0}^\top\}$. ◀

In the following we assume that S is finite, irreducible and nonzero, i.e., $S \neq \{O_n\}$. We write $S^1 := S \cup \{I_n\}$.

3.1 The (0-)minimal ideal

A *minimal* ideal of a semigroup is an ideal that is minimal within the set of all ideals. A *0-minimal* ideal of a semigroup with zero is an ideal that is minimal within the set of all nonzero ideals. Every finite semigroup has a minimal ideal, and every non-trivial finite semigroup with zero also has a 0-minimal ideal. Hence, S has a (0-)minimal ideal, say $T \neq \{O_n\}$. We show the following two lemmas.

▶ **Lemma 2.** *We have $T^2 \neq \{O_n\}$.*

Proof. Let $Z \in T \setminus \{O_n\}$ and choose $\vec{v} \in \mathbb{Q}^n$ such that $Z\vec{v} \neq \vec{0}$. Let $\mathcal{V} \subseteq \mathbb{Q}^n$ be the vector space spanned by all $YZ\vec{v}$, where $Y \in S^1$, i.e.,

$$\mathcal{V} \ := \ \left\{ \sum_{Y \in S^1} \lambda_Y Y Z \vec{v} \ \middle| \ \text{all } \lambda_Y \in \mathbb{Q} \right\}.$$

Since $I_n Z\vec{v} = Z\vec{v} \neq \vec{0}$, we have $\mathcal{V} \neq \{\vec{0}\}$. To show that $\mathcal{V}$ is S-invariant, consider an arbitrary spanning vector of $\mathcal{V}$, say $YZ\vec{v} \in \mathcal{V}$ with $Y \in S^1$. Then, for all $X \in S$ we have $XY \in S$ and hence $XYZ\vec{v} \in \mathcal{V}$. Thus, by linearity, $\mathcal{V}$ is S-invariant. Since S is irreducible, it follows that $\mathcal{V} = \mathbb{Q}^n$.

Since $\mathcal{V} = \mathbb{Q}^n$ and $Z \neq 0$, there is $\vec{w} \in \mathcal{V} \setminus \ker Z$. Write $\vec{w} = \sum_{Y \in S^1} \lambda_Y Y Z\vec{v}$ with all $\lambda_Y \in \mathbb{Q}$. Since $\vec{w} \notin \ker Z$, we have $Z\vec{w} \neq \vec{0}$. Thus, there is $Y \in S^1$ with $\lambda_Y ZYZ\vec{v} \neq \vec{0}$. Hence, $ZYZ \neq 0$. Since $Z \in T$ and T is an ideal, we have $ZY \in T$. It follows that $(ZY)Z \in T^2 \setminus \{O_n\}$. ◀

▶ **Lemma 3.** *All matrices in $T \setminus \{O_n\}$ have the same rank $r \in \{1, \ldots, n\}$.*

Proof. Pick $X \in T \setminus \{O_n\}$ of minimal nonzero rank. Since $X \in T$ and T is an ideal, we have $S^1XS^1 \subseteq T$. Moreover, S^1XS^1 is an ideal of S and this ideal is nonzero, as it contains $X \neq 0$. From the (0-)minimality of T we obtain $S^1XS^1 = T$. Hence, for any $Y \in T$ there exist $A, B \in S^1$ with $Y = AXB$. For any $Y \in T \setminus \{O_n\}$, $\operatorname{rk} Y \ = \ \operatorname{rk}(AXB) \ \leq \ \operatorname{rk}(XB) \ \leq \ \operatorname{rk} X \ \leq \ \operatorname{rk} Y$, using $\operatorname{rk}(CD) \leq \min\{\operatorname{rk} C, \operatorname{rk} D\}$ and the minimality of $\operatorname{rk}(X)$ among nonzero elements of T. Hence all nonzero elements of T have rank $\operatorname{rk} X$. ◀

For the remainder, let us write r for this common rank, i.e., $\operatorname{rk} X = r$ for all $X \in T \setminus \{O_n\}$.

4 The idempotent E and its group G

Using machinery from basic semigroup theory (see [14, Appendix A]), one can obtain the following lemmas.

▶ **Lemma 4.** *The ideal $T \subseteq S$ has an idempotent $E \in T \setminus \{O_n\}$ such that $ETE \setminus \{O_n\}$ is a finite group with identity E.*

▶ **Lemma 5.** *We have $ESE = ETE$ (which may contain O_n). Hence, by Lemma 4, $ESE \setminus \{O_n\}$ is a finite group with identity E.*

Proof. Since $E \in T$ and T is an ideal, for all $X \in S$ we have $EX \in T$ and, hence, $EXE = EEXE \in ETE$. Thus, $ESE \subseteq ETE$. The converse inclusion is immediate from $T \subseteq S$. ◀

Fix the idempotent $E \in T \setminus \{O_n\}$ from Lemma 4. The following lemma follows from the idempotence of E.

▶ **Lemma 6.** *There are matrices $D \in \mathbb{Q}^{n\times r}$ and $C \in \mathbb{Q}^{r\times n}$ with $E = DC$ and $CD = I_r$.*

Proof of Lemma 6. Let $D \in \mathbb{Q}^{n\times r}$ be a matrix consisting of columns of E that form a basis of $\operatorname{im} E$. Since $EE = E$ and the columns of D are columns of E, we have $ED = D$. By the rank-nullity theorem, we have $\dim(\ker E) = n - r$. Let $W \in \mathbb{Q}^{n\times(n-r)}$ be a matrix whose columns form a basis of $\ker E$. Thus, $EW = 0$. Since $EE = E$, we have $\operatorname{im} E \cap \ker E = \{\vec{0}\}$, so the columns of the matrix $Q \in \mathbb{Q}^{n\times n}$ with $Q = \begin{pmatrix} D & W \end{pmatrix}$ are linearly independent. Thus, Q is invertible. We have

$$EQ \;=\; \begin{pmatrix} ED & EW \end{pmatrix} \;=\; \begin{pmatrix} D & 0 \end{pmatrix} \;=\; Q\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}.$$

Hence,

$$E \;=\; Q\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} Q^{-1}.$$

Define $C \coloneqq \begin{pmatrix} I_r & 0 \end{pmatrix} Q^{-1}$ and recall that $D = Q\begin{pmatrix} I_r \\ 0 \end{pmatrix}$. Then, as required, we have

$$CD \;=\; \begin{pmatrix} I_r & 0 \end{pmatrix} Q^{-1} Q \begin{pmatrix} I_r \\ 0 \end{pmatrix} \;=\; I_r \quad \text{and}$$

$$DC \;=\; Q\begin{pmatrix} I_r \\ 0 \end{pmatrix}\begin{pmatrix} I_r & 0 \end{pmatrix} Q^{-1} \;=\; Q\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} Q^{-1} \;=\; E. \qquad ◀$$

The factorization $E = DC$ from Lemma 6 allows us to put the group from Lemma 5 in a more succinct form, which will be useful when invoking bounds on the size of matrix groups. To this end, fix $D \in \mathbb{Q}^{n\times r}$ and $C \in \mathbb{Q}^{r\times n}$ from Lemma 6, so that $DC = E$ and $CD = I_r$. Define

$$G \;\coloneqq\; CSD \setminus \{O_r\} \;\subseteq\; \mathbb{Q}^{r\times r}.$$

We have the following lemma.

▶ **Lemma 7.** *The set G is a finite group, i.e., a finite subgroup of $\mathrm{GL}_r(\mathbb{Q})$. Moreover, the finite group $ESE \setminus \{O_n\}$ from Lemma 5 is isomorphic to G via the isomorphism*

$$\phi : ESE \setminus \{O_n\} \to G \quad \textit{with} \quad \phi(X) \;\coloneqq\; CXD.$$

Proof. Let us first consider a generalization of ϕ, namely the map $\Phi : E\mathbb{Q}^{n\times n}E \to \mathbb{Q}^{r\times r}$ with $\Phi(X) \coloneqq CXD$. Note that Φ is a linear map and that

$$\Phi(EXE) \;=\; CDCXDCD \;=\; CXD \qquad \text{for all } X \in \mathbb{Q}^{n\times n}.$$

The map Φ has a trivial kernel, since if $CXD = \Phi(EXE) = O_r$ then $EXE = DCXDC = O_n$. Thus, Φ and hence ϕ are injective. It also follows that $\phi(ESE \setminus \{O_n\}) \subseteq G$.

Towards surjectivity of ϕ, let $X \in S$ with $CXD \neq 0$. Then $\Phi(EXE) = CXD$. If $EXE = O_n$ then $\Phi(EXE) = \Phi(O_n) = O_r \neq CXD$, a contradiction; hence $EXE \neq O_n$. Thus, we also have $\phi(EXE) = \Phi(EXE) = CXD$. It follows that $\phi(ESE \setminus \{O_n\}) = G$; i.e., ϕ is surjective.

It remains to show that ϕ is a homomorphism. Using $CE = CDC = C$ and $ED = DCD = D$, we obtain $\phi(E) = CED = CDCD = I_r$ and

$$\phi(EXE \cdot EYE) \;=\; CEXEYED \;=\; CXEYD \;=\; CXDCYD \;=\; \phi(EXE)\cdot\phi(EYE). \qquad ◀$$

▶ **Example 8.** Set

$$C_1 \;\coloneqq\; \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \qquad C_2 \;\coloneqq\; \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad D_1 \;\coloneqq\; \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad D_2 \;\coloneqq\; \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Then

$$C_1 D_1 \;=\; \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad C_1 D_2 \;=\; \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \;=\; C_2 D_1, \qquad C_2 D_2 \;=\; \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Let $G \subseteq \mathrm{GL}_2(\mathbb{Q})$ be the group of signed 2×2 permutation matrices (order 8). Define

$$S \;\coloneqq\; \{D_i g C_j \mid i, j \in \{1,2\},\ g \in G\}\,.$$

The set S forms a semigroup, as for any $i, j, k, \ell \in \{1,2\}$ and $g, h \in G$,

$$(D_i g C_j)(D_k h C_\ell) \;=\; D_i (g C_j D_k h) C_\ell\,,$$

and each $C_j D_k$ is in G, as computed above. The following facts about S can be checked: (i) S has $8 \cdot 2 \cdot 2 = 32$ distinct elements, none of which is the zero matrix, (ii) all elements have rank $r = 2$, (iii) S is irreducible, (iv) S is its own minimal ideal – equivalently, S is simple, (v) $E \coloneqq D_1 C_1 \in S$ is an idempotent (recall that $C_1 D_1 = I_2$), and (vi) $C_1 S D_1 = G$ (a group, as also implied by Lemma 7). ⌟

5 Upper bound: setting the stage

5.1 The injective map Ψ

To explain our general approach, consider for the moment a map $\mu : S \to G \cup \{O_r\}$ with $\mu(X) = CXD$, a generalization of the map ϕ from Lemma 7. We can use bounds on the group size mentioned in Section 2 to estimate $|G|$. But since μ is not in general injective, $|\mu(S)|$ does not bound $|S|$. Nevertheless, in the following we define a map Ψ with multiple components ψ_{ij}, each of which is a variant of μ. More concretely, we have $\psi_{ij} : S \to G \cup \{O_r\}$ with $\psi_{ij}(X) = CU_i X V_j D$ for some matrices $U_i, V_j \in S^1$. The matrices U_i, V_j are chosen so that $\Psi : X \mapsto (\psi_{ij}(X))_{ij}$ *is* injective. Intuitively, the different ψ_{ij} exhibit different "group aspects" of a semigroup element X. Since Ψ is injective, we have $|S| = |\Psi(S)|$. The known group bounds then help to estimate $|\Psi(S)|$. We provide further intuition of our approach at the end of this subsection.

Since $\sum_{X \in S^1} \mathrm{im}(XD)$ is S-invariant and nonzero (it contains $\mathrm{im}\, D$) and S is irreducible, we have $\sum_{X \in S^1} \mathrm{im}(XD) = \mathbb{Q}^n$. Thus, there exist $V_1, \ldots, V_v \in S^1$ $(v \geq 1)$ such that

$$\mathrm{im}(V_1 D) + \cdots + \mathrm{im}(V_v D) \;=\; \mathbb{Q}^n\,.$$

As we will see later, this sum need not be direct.

Dually (cf. Proposition 1), there exist $U_1, \ldots, U_u \in S^1$ $(u \geq 1)$ such that

$$\mathrm{row}(CU_1) + \cdots + \mathrm{row}(CU_u) \;=\; \mathbb{Q}^{1 \times n}\,.$$

For $0 \leq a \leq u$ and $0 \leq b \leq v$ define the vector spaces

$$\mathcal{U}_a \;\coloneqq\; \mathrm{row}(CU_1) + \cdots + \mathrm{row}(CU_a) \;\subseteq\; \mathbb{Q}^{1 \times n} \qquad \text{and}$$
$$\mathcal{V}_b \;\coloneqq\; \mathrm{im}(V_1 D) + \cdots + \mathrm{im}(V_b D) \;\subseteq\; \mathbb{Q}^n\,.$$

By convention, $\mathcal{U}_0 = \{\vec{0}^\top\}$ and $\mathcal{V}_0 = \{\vec{0}\}$. Without loss of generality, we can assume for all $1 \le a \le u$ that $\mathrm{row}(CU_a) \not\subseteq \mathcal{U}_{a-1}$. Similarly, we also assume for all $1 \le b \le v$ that $\mathrm{im}(V_b D) \not\subseteq \mathcal{V}_{b-1}$. Thus, the vector space inclusions $\{\vec{0}\} = \mathcal{U}_0 \subset \mathcal{U}_1 \subset \ldots \subset \mathcal{U}_u = \mathbb{Q}^{1\times n}$ are strict, and similarly for the $\mathcal{V}_j$. It follows that $u, v \le n$. We note the following lemma.

▶ **Lemma 9.** *For all $1 \le j \le v$ we have $\mathrm{rk}(V_j D) = r$.*

Proof. Recall that $E = DC$. Thus $\mathrm{im}\, E \subseteq \mathrm{im}\, D$. Since $\mathrm{rk}\, E = r = \mathrm{rk}\, D$, we have $\mathrm{im}\, D = \mathrm{im}\, E$. Hence, $\mathrm{im}(V_j D) = \mathrm{im}(V_j E)$, and so $\mathrm{rk}(V_j D) = \mathrm{rk}(V_j E)$. Since $\mathrm{im}(V_j D) \not\subseteq \mathcal{V}_{j-1}$, we have $V_j D \ne O_n$. Thus, $V_j E \ne O_n$. Since $E \in T$ and T is an ideal, we also have $V_j E \in T$. Since r is the common rank among nonzero elements of T, it follows that $\mathrm{rk}(V_j E) = r$. Hence, $\mathrm{rk}(V_j D) = \mathrm{rk}(V_j E) = r$. ◀

For $1 \le i \le u$ and $1 \le j \le v$ and $X \in S$ define

$$\psi_{ij}(X) \;\coloneqq\; CU_i X V_j D \;\in\; G \cup \{O_r\}\,.$$

Also define

$$\Psi(X) \;\coloneqq\; \begin{pmatrix} \psi_{11}(X) & \cdots & \psi_{1v}(X) \\ \vdots & \ddots & \vdots \\ \psi_{u1}(X) & \cdots & \psi_{uv}(X) \end{pmatrix} \;=\; \begin{pmatrix} CU_1XV_1D & \cdots & CU_1XV_vD \\ \vdots & \ddots & \vdots \\ CU_uXV_1D & \cdots & CU_uXV_vD \end{pmatrix} \;\in\; \mathbb{Q}^{ur\times vr}\,.$$

We will primarily view $\Psi(X)$ not as a large matrix, but as a grid (or array) of smaller $\mathbb{Q}^{r\times r}$ matrices.

▶ **Lemma 10.** *The map Ψ is injective.*

Proof. Let $V \;\coloneqq\; \begin{pmatrix} V_1 D & \cdots & V_v D \end{pmatrix} \;\in\; \mathbb{Q}^{n\times vr}$.

It follows from the definition of $V_1, \ldots, V_v$ that $\mathrm{im}\, V = \sum_{j=1}^{v} \mathrm{im}(V_j D) = \mathbb{Q}^n$. Thus, $\mathrm{rk}\, V = n$ and so V has a right inverse $V' \in \mathbb{Q}^{vr\times n}$ with $VV' = I_n$. Dually, the matrix

$$U \;\coloneqq\; \begin{pmatrix} CU_1 \\ \vdots \\ CU_u \end{pmatrix} \;\in\; \mathbb{Q}^{ur\times n}$$

has a left inverse $U' \in \mathbb{Q}^{n\times ur}$ with $U'U = I_n$. Noting that $\Psi(X) = UXV$, we have

$$U'\Psi(X)V' \;=\; U'UXVV' \;=\; I_n X I_n \;=\; X\,.$$

It follows that Ψ is injective. ◀

▶ **Example 11.** We continue Example 8. Choose

$$U_1 \;\coloneqq\; I_3 \text{ (recall that } I_3 \in S^1), \quad U_2 \;\coloneqq\; D_1 C_2\,, \quad V_1 \;\coloneqq\; I_3\,, \quad V_2 \;\coloneqq\; D_2 C_1\,, \quad \text{so that}$$

$$C_1U_1 = C_1 = \left(\begin{smallmatrix} 1&0&0 \\ 0&1&0 \end{smallmatrix}\right),\ C_1U_2 = C_2 = \left(\begin{smallmatrix} 0&1&0 \\ 0&0&1 \end{smallmatrix}\right),\ V_1D_1 = D_1 = \begin{pmatrix} \begin{smallmatrix} 1&0 \\ 0&1 \\ 1&0 \end{smallmatrix} \end{pmatrix},\ V_2D_1 = D_2 = \begin{pmatrix} \begin{smallmatrix} 0&1 \\ 1&0 \\ 0&-1 \end{smallmatrix} \end{pmatrix}.$$

Thus, $\mathrm{row}(C_1U_1) + \mathrm{row}(C_1U_2) = \mathbb{Q}^{1\times 3}$ and $\mathrm{im}(V_1D_1) + \mathrm{im}(V_2D_1) = \mathbb{Q}^3$. The ranks of these four matrices are all 2, consistent with Lemma 9 and its analogue for C_1U_1, C_1U_2. We have $u = v = 2$. ⌟

In the following we will be interested in $|S|$. By Lemma 10, we have $|S| = |\Psi(S)|$; i.e., it suffices to estimate the number of different $\Psi(X)$ for $X \in S$. Taking this further, we have

$$\begin{aligned} |S| \;&=\; |\Psi(S)| \;\le\; \prod_{i=1}^{u}\prod_{j=1}^{v}|\psi_{ij}(S)| \;=\; \prod_{i=1}^{u}\prod_{j=1}^{v}|CU_iSV_jD| \;\le\; \prod_{i=1}^{u}\prod_{j=1}^{v}(|G|+1) \\ &=\; (|G|+1)^{uv} \;\le\; 3^{r^2uv} \qquad \text{using Lemmas 7 and 17.} \end{aligned}$$

Suppose for a moment that the vector spaces $(\mathrm{row}(CU_i))_i$ were independent and the vector spaces $(\mathrm{im}(V_jD))_j$ were independent, i.e., suppose that $\bigoplus_{i=1}^{u}\mathrm{row}(CU_i) = \mathbb{Q}^{1\times n}$ and $\bigoplus_{j=1}^{v}\mathrm{im}(V_jD) = \mathbb{Q}^n$. Then $ur = n$ and $vr = n$ (in particular, $r \mid n$ and $u = v = n/r$), and using the inequality above we would obtain $|S| \le 3^{n^2}$. However, in general this independence does not hold and we have to estimate $u, v \le n$, giving only a weaker bound $|S| \le 3^{r^2n^2}$. Therefore, we pursue a different avenue, based on the idea that if, say, $\mathrm{im}(V_1D)$ and $\mathrm{im}(V_2D)$ overlap nontrivially then $\psi_{i1}(X)$ and $\psi_{i2}(X)$ are "coupled" across $X \in S$, i.e., $\psi_{i1}(X)$ and $\psi_{i2}(X)$ do not vary independently. Formalizing this idea and making it work is the key technical contribution of this paper. Before doing that, we treat a much easier case of aperiodic semigroups.

5.2 The aperiodic case

A semigroup is called *aperiodic* if every subsemigroup which is also a group is trivial, i.e., has only one element. In this subsection we analyze the size of S, assuming it is aperiodic. Showcasing the use of our injective map Ψ, Theorem 13 below recovers a result from [2].

▶ **Lemma 12.** *If S is aperiodic, we have $r = 1$ and $G = \{I_1\}$.*

Proof. By Lemma 5, $ESE \setminus \{O_n\}$ is a group. Thus, it is a subgroup of the semigroup S. Since S is aperiodic, $ESE \setminus \{O_n\} = \{E\}$. By Lemma 7 it follows that $G = \{I_r\}$.

Let $\vec{d} \in \mathbb{Q}^n$ be a (necessarily nonzero) column of D. Since $ED = DCD = D$, we have $E\vec{d} = \vec{d}$. Write $\langle\cdot\rangle$ for the $\mathbb{Q}$-span. The vector space $\langle S\vec{d}\rangle$ is S-invariant and nonzero, as it contains $E\vec{d} = \vec{d}$. Thus, irreducibility of S implies that $\langle S\vec{d}\rangle = \mathbb{Q}^n$. Therefore,

$$\mathrm{im}\,E \;=\; E\mathbb{Q}^n \;=\; E\langle S\vec{d}\rangle \overset{E\vec{d}=\vec{d}}{=} \langle ESE\vec{d}\rangle \overset{ESE\setminus\{O_n\}=\{E\}}{=} \langle E\vec{d}\rangle \overset{E\vec{d}=\vec{d}}{=} \langle\vec{d}\rangle\,.$$

Hence, $r = \mathrm{rk}\,E = \dim(\mathrm{im}\,E) = 1$, and so $G = \{I_1\}$. ◀

Using the injective map Ψ we obtain the following result, which essentially follows from the proof of [2, Theorem 5.8]. This is already a good illustration how our approach differs from the approach of [24, 5, 2]: instead of bounding the possible values that the traces of matrices in S can take, we consider a family Ψ of "linear" maps from S to the group G.

▶ **Theorem 13.** *Let $S \subseteq \mathbb{Q}^{n\times n}$ be an aperiodic finite irreducible semigroup. Then $|S| \le 2^{n^2}$.*

Proof. By Lemma 12, we have $r = 1$. Using the assumption that the spaces $\mathcal{U}_i$ for $0 \le i \le u$ are strictly increasing, and similarly for the $\mathcal{V}_j$, it follows that $u = n = v$. Also by Lemma 12, we have $\psi_{ij}(X) \in \{O_1, I_1\}$ for all $1 \le i, j \le n$ and all $X \in S$; i.e., $\Psi(S) \subseteq \{O_1, I_1\}^{n\times n}$. Since Ψ is injective, $|S| = |\Psi(S)| \le |\{O_1, I_1\}^{n\times n}| = 2^{n^2}$. ◀

6 Upper bound: doing the main work

6.1 The width w_b of a column

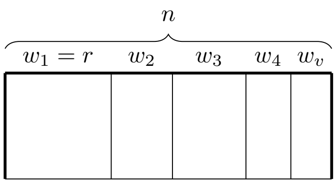
For this and the next two subsections, we fix an arbitrary column index $b \in \{1, \ldots, v\}$. Define the *width* w_b of block column b:

$$w_b \;\coloneqq\; \dim \mathcal{V}_b - \dim \mathcal{V}_{b-1} \;=\; \dim(\mathcal{V}_{b-1} + \operatorname{im}(V_b D)) - \dim \mathcal{V}_{b-1}\,.$$

Intuitively, $\operatorname{im}(V_b D)$ adds $w_b \le r$ independent dimensions to $\mathcal{V}_{b-1}$. We note that

$$\begin{aligned} w_1 \;&=\; \dim \mathcal{V}_1 \;=\; \operatorname{rk}(V_1 D) \;=\; r \quad && \text{by Lemma 9 and} \\ w_1 + \cdots + w_v \;&=\; \dim \mathcal{V}_v \;=\; n && \text{from the definition of the } \mathcal{V}_j. \end{aligned} \tag{1}$$

The following picture illustrates these widths.



Also define

$$\mathcal{L}_b \;\coloneqq\; \{\vec{y} \in \mathbb{Q}^r \mid V_b D \vec{y} \in \mathcal{V}_{b-1}\} \qquad \text{and} \qquad \ell_b \;\coloneqq\; \dim \mathcal{L}_b\,.$$

In words, $\mathcal{L}_b$ is the vector space consisting of the vectors $\vec{y} \in \mathbb{Q}^r$ that the matrix $V_b D$ maps into the intersection of $\mathcal{V}_{b-1}$ and $\operatorname{im}(V_b D)$; i.e., we have $V_b D \mathcal{L}_b = \mathcal{V}_{b-1} \cap \operatorname{im}(V_b D)$. The following lemma connects w_b and ℓ_b.

▶ **Lemma 14.** *We have $w_b = r - \ell_b > 0$.*

Proof. Consider the map $V_b D : \mathbb{Q}^r \to \mathbb{Q}^n$. Its domain is r-dimensional and we have $\operatorname{rk}(V_b D) = r$ by Lemma 9. It follows that the map $V_b D$ is injective. Hence its restriction to any subspace is injective. Thus, $\dim(V_b D \mathcal{L}_b) = \dim \mathcal{L}_b = \ell_b$. Hence,

$$\begin{aligned} w_b \;&=\; \dim(\mathcal{V}_{b-1} + \operatorname{im}(V_b D)) - \dim \mathcal{V}_{b-1} \\ &=\; \dim \mathcal{V}_{b-1} + \operatorname{rk}(V_b D) - \dim(\mathcal{V}_{b-1} \cap \operatorname{im}(V_b D)) - \dim \mathcal{V}_{b-1} \\ &=\; r - \dim(\mathcal{V}_{b-1} \cap \operatorname{im}(V_b D)) && \text{by Lemma 9} \\ &=\; r - \dim(V_b D \mathcal{L}_b) \;=\; r - \ell_b\,. \end{aligned}$$

Since $\mathcal{L}_b \subseteq \mathbb{Q}^r$, we have $\ell_b \le r$. If $\ell_b = r$ then $\mathcal{L}_b = \mathbb{Q}^r$, implying that $\operatorname{im}(V_b D) = V_b D \mathcal{L}_b \subseteq \mathcal{V}_{b-1}$, contradicting the assumption made after the definition of $\mathcal{V}_b$. Hence, $\ell_b < r$. ◀

▶ **Example 15.** Continuing Example 11, we have

$$\mathcal{V}_1 \;=\; \operatorname{im}(V_1 D_1) \;=\; \left\{ \begin{pmatrix} p \\ q \\ p \end{pmatrix} \;\middle|\; p, q \in \mathbb{Q} \right\}, \qquad \operatorname{im}(V_2 D_1) \;=\; \left\{ \begin{pmatrix} p \\ q \\ -p \end{pmatrix} \;\middle|\; p, q \in \mathbb{Q} \right\}.$$

Thus, $\mathcal{V}_2 = \mathcal{V}_1 + \operatorname{im}(V_2 D_1) = \mathbb{Q}^3$. Hence,

$$v = 2, \quad w_1 = \dim \mathcal{V}_1 = 2 = r, \quad w_2 = \dim \mathcal{V}_2 - \dim \mathcal{V}_1 = 3 - 2 = 1.$$

We also have

$$\begin{aligned} \mathcal{L}_2 \;&=\; \{\vec{y} \in \mathbb{Q}^2 \mid V_2 D_1 \vec{y} \in \mathcal{V}_1\} \;=\; \left\{ \begin{pmatrix} p \\ q \end{pmatrix} \in \mathbb{Q}^2 \;\middle|\; \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix} \in \mathcal{V}_1 \right\} \\ &=\; \left\{ \begin{pmatrix} p \\ q \end{pmatrix} \in \mathbb{Q}^2 \;\middle|\; \begin{pmatrix} q \\ p \\ -q \end{pmatrix} \in \mathcal{V}_1 \right\} \;=\; \left\{ \begin{pmatrix} p \\ 0 \end{pmatrix} \;\middle|\; p \in \mathbb{Q} \right\}. \end{aligned}$$

Thus, $\ell_2 = \dim \mathcal{L}_2 = 1$, matching Lemma 14: $w_2 = r - \ell_2 = 2 - 1 = 1$. ⌟

6.2 The coupling group H_b

In this subsection we introduce H_b, a subgroup of G. Later we will see that this "coupling group" H_b restricts the possibilities for $\psi_{ab}(X)$ (for some block row a) once the "prefix" $\psi_{a1}(X), \ldots, \psi_{a(b-1)}(X)$ has been fixed. The smaller the width w_b, the smaller H_b becomes and the fewer possibilities are there for $\psi_{ab}(X)$. Define

$$H_b \;\coloneqq\; \{g \in G \mid g\vec{y} = \vec{y} \text{ for all } \vec{y} \in \mathcal{L}_b\}\,;$$

i.e., H_b consists of those matrices $g \in G$ that fix $\mathcal{L}_b$.

▶ **Example 16.** Continuing Example 15, we have

$$H_2 \;=\; \{g \in G \mid g\vec{y} = \vec{y} \;\; \forall \vec{y} \in \mathcal{L}_2\} \;=\; \{g \in G \mid g\left(\begin{smallmatrix} p \\ 0 \end{smallmatrix}\right) = \left(\begin{smallmatrix} p \\ 0 \end{smallmatrix}\right) \;\; \forall p \in \mathbb{Q}\} \;=\; \left\{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right),\; \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)\right\}.$$

Thus, $|H_2| + 1 = 3 = 3^{1^2} = 3^{w_2^2}$, realizing the upper bound of the following lemma. ⌟

Our analysis of the semigroup size is based on known bounds on the size of finite matrix groups. Concretely, we will use the following lemma. Its proof follows classical lines; see, e.g., [15]. For completeness, we provide a proof in the appendix of [14].

▶ **Lemma 17.** *Let $n \geq 1$. Any finite subgroup of $\mathrm{GL}_n(\mathbb{Q})$ has at most $3^{n^2} - 1$ elements.*

▶ Remark 18. Lemma 17 is not tight. As mentioned in the introduction, it is known (via an elementary proof not based on the classification of finite simple groups) that the order of any finite subgroup, say H, of $\mathrm{GL}_n(\mathbb{Q})$ divides $(2n)!$ (see, e.g., [18, Chapter IX]); so $|H| \leq (2n)! = 3^{\Theta(n \log n)}$. It is not difficult to prove Lemma 17 by showing that $(2n)! \leq 3^{n^2} - 1$, but the more fundamental proof of Lemma 17 in [14] might give more insight.

We will use the group bound from Lemma 17 to bound the size of finite irreducible matrix semigroups $S \subseteq \mathbb{Q}^{n \times n}$ in terms of n. An important role will be played by a certain group that is isomorphic to a finite subgroup of $\mathrm{GL}_r(\mathbb{Q})$, where r is the minimum nonzero rank of the matrices in S. The bottleneck (for our semigroup bound in terms of n) will turn out to be the case $r = 1$, where we have $3^{1^2} - 1 = 2 = (2 \cdot 1)!$. Therefore, the mentioned asymptotically tighter results for groups do not improve our main result on semigroups. ⌟

▶ **Lemma 19.** *We have $|H_b| + 1 \;\leq\; 3^{w_b^2}$.*

Proof sketch. For any $h_1, h_2 \in H_b$ we have $h_1 h_2 \in H_b$, as $h_1 h_2 \vec{y} = h_1 \vec{y} = \vec{y}$ holds for all $\vec{y} \in \mathcal{L}_b$. Let $h \in H_b$. We show that $h^{-1} \in H_b$. Indeed, for all $\vec{y} \in \mathcal{L}_b$ we have $h^{-1}\vec{y} = h^{-1}(h\vec{y}) = (h^{-1}h)\vec{y} = \vec{y}$. We conclude that H_b is a group. It is finite, as $G \supseteq H_b$ is finite. We show the bound on $|H_b|$ in the appendix of [14], using Lemmas 14 and 17. ◀

6.3 A block row prefix

For this subsection, we fix an arbitrary block row index $a \in \{1, \ldots, u\}$. We consider the (number of) possible first b blocks of the ath block row of $\Psi(X)$ when X ranges over S, i.e., the possible

$$\begin{pmatrix} \psi_{a1}(X) & \cdots & \psi_{ab}(X) \end{pmatrix} \qquad \text{where } X \in S.$$

The following lemma states in particular that the action of $\psi_{ab}(X)$ on $\mathcal{L}_b$ is determined by the actions of $\psi_{a1}(X), \ldots, \psi_{a(b-1)}(X)$ on $\mathcal{L}_b$.

▶ **Lemma 20.** *There exist linear maps* $\Theta_1, \dots, \Theta_{b-1} : \mathcal{V}_{b-1} \to \mathbb{Q}^r$ *such that*

$$\psi_{ab}(X)\vec{y} \;=\; \sum_{j=1}^{b-1} \psi_{aj}(X)\Theta_j(V_b D\vec{y}) \qquad \text{for all } X \in S \text{ and all } \vec{y} \in \mathcal{L}_b.$$

Proof. Consider the linear map

$$\Omega : (\mathbb{Q}^r)^{b-1} \to \mathcal{V}_{b-1} \qquad \text{with} \qquad \Omega(\vec{y}_1, \dots, \vec{y}_{b-1}) \;\coloneqq\; \sum_{j=1}^{b-1} V_j D\vec{y}_j\,.$$

Since Ω is surjective, it has a linear right inverse $\Sigma : \mathcal{V}_{b-1} \to (\mathbb{Q}^r)^{b-1}$ with $\vec{z} = \Omega(\Sigma(\vec{z}))$ for all $\vec{z} \in \mathcal{V}_{b-1}$. Write $\Sigma(\vec{z}) =: (\Theta_1(\vec{z}), \dots, \Theta_{b-1}(\vec{z}))$. Thus, $\vec{z} = \sum_{j=1}^{b-1} V_j D\Theta_j(\vec{z})$ for all $\vec{z} \in \mathcal{V}_{b-1}$. In particular, for all $\vec{y} \in \mathcal{L}_b$, since $V_b D\vec{y} \in \mathcal{V}_{b-1}$,

$$V_b D\vec{y} \;=\; \sum_{j=1}^{b-1} V_j D\Theta_j(V_b D\vec{y})\,.$$

Left-multiplying by CU_aX yields the claimed equality. ◀

The following lemma says that if two matrices $\widehat{X}, X \in S$ have the same Ψ-values in the first $b-1$ blocks of block row a, then their ψ_{ab}-values are related by an element of the group H_b. This lemma motivates our term “coupling group” for H_b.

▶ **Lemma 21.** *Suppose that* $\widehat{X}, X \in S$ *satisfy* $\psi_{aj}(\widehat{X}) = \psi_{aj}(X)$ *for all* $1 \le j \le b-1$*. If* $\psi_{ab}(\widehat{X}), \psi_{ab}(X) \in G$ *(i.e., are nonzero), then there is an* $h \in H_b$ *such that* $\psi_{ab}(\widehat{X})h = \psi_{ab}(X)$*.*

Proof. Write $\widehat{g} \coloneqq \psi_{ab}(\widehat{X}) \in G$ and $g \coloneqq \psi_{ab}(X) \in G$. By Lemma 20,

$$\widehat{g}\vec{y} \;=\; \sum_{j=1}^{b-1} \psi_{aj}(\widehat{X})\Theta_j(V_b D\vec{y}) \;=\; \sum_{j=1}^{b-1} \psi_{aj}(X)\Theta_j(V_b D\vec{y}) \;=\; g\vec{y} \qquad \text{for all } \vec{y} \in \mathcal{L}_b\,;$$

i.e., $\widehat{g}$ and g agree on $\mathcal{L}_b$. It follows that $h \coloneqq \widehat{g}^{-1}g$ (where $h \in G$, as G is a group) fixes $\mathcal{L}_b$; i.e., $h\vec{y} = \vec{y}$ for all $\vec{y} \in \mathcal{L}_b$. Thus, $h \in H_b$ and $\widehat{g}h = g$. ◀

▶ **Example 22.** We continue Example 16. Using the expressions for C_iD_j $(i, j \in \{1, 2\})$ from Example 8 and the fact that G is a group one can show that

$$\{(\psi_{11}(X) \quad \psi_{12}(X)) \mid X \in S\} \;=\; \{(g \quad g\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)) \mid g \in G\} \;\cup\; \{(g \quad g\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)) \mid g \in G\}\,.$$

Therefore, for any $\widehat{X}, X \in S$ with $\psi_{11}(\widehat{X}) = \psi_{11}(X)$ we have

$$\psi_{12}(\widehat{X})\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right) \;=\; \psi_{12}(X) \qquad \text{or} \qquad \psi_{12}(\widehat{X})\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right) \;=\; \psi_{12}(X)\,.$$

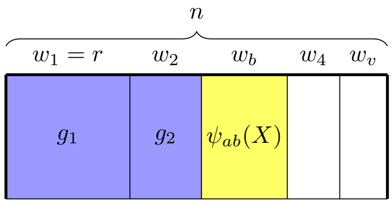
Since we have $H_2 = \{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)\}$ from Example 16, this matches Lemma 21. ⌟

The following lemma bounds the number of different $\psi_{ab}(X)$ when the $\psi_{aj}(X)$ for $j < b$ have been fixed.

▶ **Lemma 23.** *Let* $g_1, \dots, g_{b-1} \in G \cup \{O_r\}$*. Then*

$$|\{\psi_{ab}(X) \mid X \in S,\ \psi_{aj}(X) = g_j \text{ for all } 1 \le j \le b-1\}| \;\le\; 3^{w_b^2}\,.$$

Here is an illustration of the lemma.



Proof of Lemma 23. Set

$$R \coloneqq \{\psi_{ab}(X) \mid X \in S,\ \psi_{aj}(X) = g_j \text{ for } 1 \le j \le b-1\} \subseteq G \cup \{O_r\}.$$

Suppose that $R \setminus \{O_r\}$ is nonempty; i.e., there is $\widehat{X} \in S$ with $\psi_{aj}(\widehat{X}) = g_j$ for all $1 \le j \le b-1$ and $\psi_{ab}(\widehat{X}) \ne O_r$. Then, by Lemma 21, $R \setminus \{O_r\} \subseteq \psi_{ab}(\widehat{X})H_b$. It follows that $|R \setminus \{O_r\}| \le |H_b|$. Thus, $|R| \le |H_b| + 1$. Clearly, this bound also holds when $R \setminus \{O_r\}$ is empty. Hence, Lemma 19 implies $|R| \le 3^{w_b^2}$. ◀

The following proposition bounds the number of length-b prefixes of the ath block row of Ψ. It will not be used later; it serves as "warm-up" for the "2-dimensional" Lemma 27 in Section 6.4 below.

▶ **Proposition 24.** *Let* $Y_b \coloneqq \{(\psi_{a1}(X) \ \cdots \ \psi_{ab}(X)) \mid X \in S\}$. *We have* $|Y_b| \le 3^{w_1^2 + \cdots + w_b^2}$.

Proof. The value of $b \in \{1, \ldots, v\}$ was fixed at the beginning of Section 6.1. In the following we let b vary, and prove the proposition by induction on $b \in \{1, \ldots, v\}$. The induction base, $b = 1$, follows immediately from Lemma 23. For the induction step, suppose $|Y_{b-1}| \le 3^{w_1^2 + \cdots + w_{b-1}^2}$ holds for some $1 < b \le v$. We have

$$\begin{aligned}
|Y_b| &= \sum_{(g_1 \ \cdots \ g_{b-1}) \in Y_{b-1}} |\{\psi_{ab}(X) \mid X \in S,\ \psi_{aj}(X) = g_j \text{ for all } 1 \le j \le b-1\}| \\
&\le \sum_{(g_1 \ \cdots \ g_{b-1}) \in Y_{b-1}} 3^{w_b^2} = |Y_{b-1}| \cdot 3^{w_b^2} && \text{by Lemma 23} \\
&\le 3^{w_1^2 + \cdots + w_{b-1}^2} \cdot 3^{w_b^2} = 3^{w_1^2 + \cdots + w_b^2} && \text{by the induction hypothesis.}
\end{aligned}$$

◀

Using Proposition 24, we can improve the bound $|S| \le 3^{r^2 n^2}$ obtained at the end of Section 5.1. Let us write $Y_{ab} \coloneqq Y_b$ for the set Y_b from Proposition 24, to make its implicit dependence on $a \in \{1, \ldots, u\}$ (fixed at the beginning of the subsection) explicit. Since $w_j \le r$ and $w_1 + \cdots + w_v = n$ by Equation (1), we have $w_1^2 + \cdots + w_v^2 \le rn$. Then Proposition 24 gives $|Y_{av}| \le 3^{rn}$, and we obtain, using $u \le n$,

$$|S| = |\Psi(S)| \le \prod_{a=1}^{u} |Y_{av}| \le \prod_{a=1}^{u} 3^{rn} = 3^{rnu} \le 3^{rn^2},$$

improving on the earlier bound by a factor of r in the exponent.

In order to improve this bound down to $|S| \le 3^{n^2}$, we need to exploit dependencies between the block rows, in addition to the dependencies within block row a explored thus far. Column dependencies are, of course, completely analogous to row dependencies; the remaining challenge is to find a way to couple each block (a, b) both within its row and its column.

6.4 The overall count

In the previous subsection we considered the length-b prefix of the ath block row of Ψ

$$\begin{pmatrix}\psi_{a1}(X) & \cdots & \psi_{ab}(X)\end{pmatrix} \qquad \text{where } X \in S.$$

Next we wish to formulate the column analogue of Lemma 23. Analogously to the width w_b of block column b, we define the *height*, h_a, of block row a, i.e.,

$$h_a \;:=\; \dim \mathcal{U}_a - \dim \mathcal{U}_{a-1} \qquad \text{where } 1 \le a \le u.$$

We have $h_a > 0$, analogously to $w_b > 0$ from Lemma 14. The following equalities are exactly analogous to Equation (1) for w_b in Section 6.1:

$$\begin{aligned} h_1 \;&=\; \dim \mathcal{U}_1 \;=\; \mathrm{rk}(CU_1) \;=\; r \\ h_1 + \cdots + h_u \;&=\; \dim \mathcal{U}_u \;=\; n\,. \end{aligned} \tag{2}$$

In particular, $h_1 = \mathrm{rk}(CU_1) = r$ follows from the analogue of Lemma 9. The following lemma considers a length-a prefix of the bth block column of Ψ.

▶ **Lemma 25.** *Let $1 \le a \le u$ and $1 \le b \le v$. Let $g_1, \ldots, g_{a-1} \in G \cup \{O_r\}$. Then*

$$|\{\psi_{ab}(X) \mid X \in S,\ \psi_{ib}(X) = g_i \text{ for all } 1 \le i \le a-1\}| \;\le\; 3^{h_a^2}\,.$$

The proof follows from transposing the row argument from the last two subsections; we omit the proof, as it is fully analogous to the proof of Lemma 23.

Towards the overall count, define the *grid*

$$\Gamma \;:=\; \{1, \ldots, u\} \times \{1, \ldots, v\}\,.$$

Let $\prec$ be the row-major order on Γ, i.e.,

$$(i,j) \prec (i',j') \quad\Longleftrightarrow\quad i < i' \quad \text{or} \quad (i = i' \text{ and } j < j').$$

Figure 2(a) visualizes the order $\prec$.

The following lemma is a grid analogue to Lemmas 23 and 25; in fact, the proof is based on these lemmas.

▶ **Lemma 26.** *Let $(a,b) \in \Gamma$. For all $(i,j) \prec (a,b)$ fix $g_{ij} \in G \cup \{O_r\}$. Let*

$$R \;:=\; \{\psi_{ab}(X) \mid X \in S,\ \psi_{ij}(X) = g_{ij} \text{ for all } (i,j) \prec (a,b)\}\,.$$

Then $|R| \le 3^{h_a w_b}$.

Proof. Note that $(a,j) \prec (a,b)$ and $(i,b) \prec (a,b)$ for all $1 \le j \le b-1$ and all $1 \le i \le a-1$. Figure 2(b) shows these grid elements in dark-blue. We have

$$\begin{aligned} R \;&\subseteq\; \{\psi_{ab}(X) \mid X \in S,\ \psi_{aj}(X) = g_{aj} \text{ for all } 1 \le j \le b-1\} \qquad \text{and} \\ R \;&\subseteq\; \{\psi_{ab}(X) \mid X \in S,\ \psi_{ib}(X) = g_{ib} \text{ for all } 1 \le i \le a-1\}\,. \end{aligned}$$

Using Lemmas 23 and 25 respectively, we obtain $|R| \le 3^{w_b^2}$ and $|R| \le 3^{h_a^2}$. Since $h_a, w_b > 0$,

$$|R| \;\le\; \min\left\{3^{h_a^2}, 3^{w_b^2}\right\} \;=\; 3^{(\min\{h_a, w_b\})^2} \;\le\; 3^{h_a w_b}\,. \qquad ◀$$

The following lemma is the grid analogue to Proposition 24.

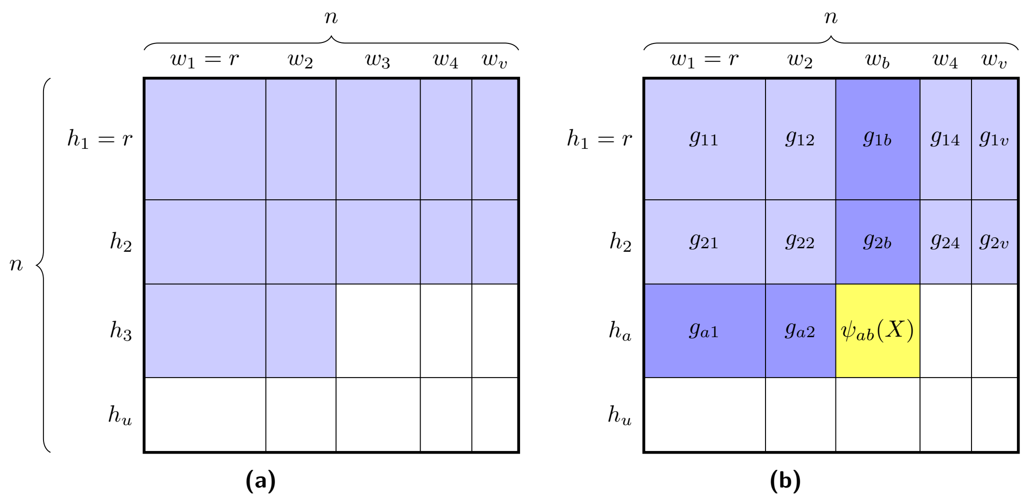


Figure 2 (a) Grid Γ with the cells $\prec (3,3)$ shown in blue. (b) Illustration of the proof of Lemma 26; the number of possible values $\psi_{ab}(X)$ in the yellow cell is limited by the possible combinations of values in the dark-blue cells.

▶ **Lemma 27.** *Let $1 \le k \le uv$ and let $(a,b) \in \Gamma$ be the kth pair in the order $\prec$. Define*

$$Z_k \;\coloneqq\; \left\{ \left(\psi_{ij}(X)\right)_{(i,j)\preceq(a,b)} \;\middle|\; X \in S \right\}.$$

Set $s \coloneqq \sum_{(i,j)\preceq(a,b)} h_i w_j$. Then we have $|Z_k| \le 3^s$.

Proof. We prove the lemma by induction on $k \in \{1,\dots,uv\}$. The induction base, $k=1$, follows immediately from Lemma 26. For the induction step, let $1 < k \le uv$, and suppose that $|Z_{k-1}| \le 3^{s_0}$ holds for $s_0 \coloneqq \sum_{(i,j)\prec(a,b)} h_i w_j$, where $(a,b)\in\Gamma$ is the kth pair in the order $\prec$. We have

$$\begin{aligned}
|Z_k| \;&=\; \sum_{(g_{ij})_{(i,j)\prec(a,b)} \,\in\, Z_{k-1}} |\{\psi_{ab}(X) \mid X \in S,\ \psi_{ij}(X) = g_{ij} \text{ for all } (i,j)\prec(a,b)\}| \\
&\le\; \sum_{(g_{ij})_{(i,j)\prec(a,b)} \,\in\, Z_{k-1}} 3^{h_a w_b} \;=\; |Z_{k-1}|\cdot 3^{h_a w_b} && \text{by Lemma 26} \\
&\le\; 3^{s_0}\cdot 3^{h_a w_b} \;=\; 3^s && \text{by the induction hypothesis.}
\end{aligned}$$ ◀

Now the main theorem follows.

▶ **Theorem 28.** *Let $S \subseteq \mathbb{Q}^{n\times n}$ be a finite irreducible semigroup. Then $|S| \le 3^{n^2}$.*

Proof. Recall from Equations (1) and (2) that $h_1 + \cdots + h_u = n = w_1 + \cdots + w_v$. We have

$$\begin{aligned}
|S| \;&=\; |\Psi(S)| && \text{as } \Psi \text{ is injective by Lemma 10} \\
&=\; |Z_{uv}| \;\le\; 3^s && \text{by Lemma 27},
\end{aligned}$$

where $s = \sum_{(i,j)\in\Gamma} h_i w_j = \sum_{i=1}^{u} h_i \sum_{j=1}^{v} w_j = n\cdot n = n^2$. ◀

From Theorem 28, by decomposing S into irreducible "parts", it is not difficult to prove:

▶ **Theorem 29.** *Let $S \subseteq \mathbb{Q}^{n\times n}$ be a finite, not necessarily irreducible, semigroup, generated by $S_0 \subseteq S$. If S contains the zero matrix, then its mortality threshold is at most 3^{n^2}.*

7 Lower bound

Fix $n \geq 2$ and write $n = p + q$ with

$$p := \lfloor \tfrac{n}{2} \rfloor, \qquad q := \lceil \tfrac{n}{2} \rceil, \qquad P := \{1, \ldots, p\}, \qquad Q := \{p+1, \ldots, n\}.$$

We view P as the "north-west" index set and Q as the "south-east" one; note $P \cap Q = \emptyset$. For $X \in \mathbb{Z}^{n \times n}$ we write

$$\operatorname{supp} X := \{(i,j) \in \{1, \ldots, n\}^2 \mid X_{ij} \neq 0\}$$

for the support of X. We define four families of matrices with entries in $\{-1, 0, 1\}$:

$$\begin{aligned}
\mathsf{NE} &:= \{X \in \{-1,0,1\}^{n \times n} \mid \operatorname{supp} X \subseteq P \times Q\}, \\
\mathsf{COL} &:= \{X \in \{-1,0,1\}^{n \times n} \mid \exists b \in P : \operatorname{supp} X \subseteq P \times \{b\}\}, \\
\mathsf{ROW} &:= \{X \in \{-1,0,1\}^{n \times n} \mid \exists a \in Q : \operatorname{supp} X \subseteq \{a\} \times Q\}, \\
\mathsf{UNIT} &:= \{X \in \{-1,0,1\}^{n \times n} \mid |\operatorname{supp} X| \leq 1\}.
\end{aligned}$$

In words, NE consists of the north-east-supported matrices; COL consists of the north-west-supported matrices with at most one nonzero column; ROW consists of the south-east-supported matrices with at most one nonzero row; and UNIT consists of the signed matrix units and 0. Each family includes the zero matrix. Define $S := \mathsf{NE} \cup \mathsf{COL} \cup \mathsf{ROW} \cup \mathsf{UNIT}$. The following proposition complements Theorem 28.

▶ **Proposition 30.** *The set $S \subseteq \{-1,0,1\}^{n \times n}$ is a finite irreducible integer matrix semigroup with at least $3^{\lfloor n^2/4 \rfloor}$ elements.*

Proof. Clearly, S is finite. To argue that S is closed under multiplication, consider the following multiplication table.

$\cdot$	NE	COL	ROW	UNIT
NE	$\{O_n\}$	$\{O_n\}$	NE	NE ∪ COL
COL	NE	COL	$\{O_n\}$	COL ∪ NE
ROW	$\{O_n\}$	$\{O_n\}$	ROW	UNIT
UNIT	NE ∪ ROW	UNIT	NE ∪ ROW	UNIT

For example, the entry in row NE and column ROW is NE, to indicate that NE·ROW $\subseteq$ NE. To show this, let $X \in \mathsf{NE}$ and $Y \in \mathsf{ROW}$. Since $Y \in \mathsf{ROW}$, there is $a \in Q$ with $\operatorname{supp} Y \subseteq \{a\} \times Q$. Moreover, $\operatorname{supp} X \subseteq P \times Q$. It follows that $\operatorname{supp}(XY) \subseteq P \times Q$; i.e., for $(i,j) \notin P \times Q$ we have $(XY)_{ij} = 0$. For any $(i,j) \in P \times Q$,

$$(XY)_{ij} = \sum_{k=1}^{n} X_{ik} Y_{kj} = X_{ia} Y_{aj} \in \{-1, 0, 1\}.$$

It follows that $XY \in \mathsf{NE}$. The rest of the multiplication table above is shown similarly. In particular, in every product the support constraints force the summation $(XY)_{ij} = \sum_k X_{ik} Y_{kj}$ to have at most one nonzero term; so the entries remain in $\{-1, 0, +1\}$.

For irreducibility, let $\{\vec{0}\} \neq \mathcal{V} \subseteq \mathbb{Q}^n$ be S-invariant. Since $\mathcal{V} \neq \{\vec{0}\}$ and $\mathcal{V}$ is closed under scalar multiplication, there is $\vec{v} \in \mathcal{V}$ with $\vec{v}_j = 1$ for some $1 \leq j \leq n$. Let $1 \leq i \leq n$. It suffices to show that $\vec{e}_i \in \mathcal{V}$, where $\vec{e}_i \in \{0,1\}^n$ denotes the ith coordinate vector. To that end, let $E_{ij} \in \mathsf{UNIT} \subseteq S$ be the matrix whose only nonzero entry is a 1 at position (i,j). Then $\vec{e}_i = E_{ij} \vec{v} \in \mathcal{V}$, as $\mathcal{V}$ is S-invariant.

Finally, $|S| \geq |\mathsf{NE}| = \left|\{-1,0,1\}^{P \times Q}\right| = 3^{|P||Q|} = 3^{\lfloor n^2/4 \rfloor}$. ◀

The following proposition complements Theorem 13.

▶ **Proposition 31.** *The set $S \cap \{0,1\}^{n\times n}$ is an aperiodic irreducible semigroup of matrices with entries in $\{0,1\}$. It has at least $2^{\lfloor n^2/4 \rfloor}$ elements.*

Proof. Define $S_{\geq 0} \coloneqq S \cap \{0,1\}^{n\times n}$. Since S is closed, it follows that $S_{\geq 0}$ is closed; i.e., $S_{\geq 0}$ is a semigroup. The element count and the irreducibility argument from Proposition 30 carry over to $S_{\geq 0}$ analogously. It remains to show that $S_{\geq 0}$ is aperiodic.

Let K be a subgroup of $S_{\geq 0}$ with identity E. Then $EE = E$; i.e., E is idempotent. We need to show that $K = \{E\}$. If $O_n \in K$ then $K = \{O_n\} = \{E\}$. So we assume that $O_n \notin K$; in particular, $E \neq O_n$.

If $E \in \mathsf{NE}$, then $E = EE = O_n$, contradicting our assumption.

Suppose $E \in \mathsf{COL} \cup \mathsf{UNIT}$. Then E is supported on some column $b \in \{1,\ldots,n\}$; i.e., writing $\vec{b} \in \{0,1\}^n$ for the bth coordinate vector, we have $E = \vec{e}\vec{b}^\top$ for some $\vec{e} \in \{0,1\}^n$. Let $X \in K$. Since E is the identity in K, we have

$$X = EX = EXE = (\vec{e}\vec{b}^\top)X(\vec{e}\vec{b}^\top) = \vec{e}(\vec{b}^\top X\vec{e})\vec{b}^\top = (\vec{b}^\top X\vec{e})\vec{e}\vec{b}^\top = (\vec{b}^\top X\vec{e})E\,;$$

i.e., $X \in \{0,1\}^{n\times n} \setminus \{O_n\}$ is a nonnegative integer multiple of E. It follows that $X = E$. Since $X \in K$ was arbitrary, we conclude that $K = \{E\}$.

If $E \in \mathsf{ROW}$, the argument is similar. ◀

8 Conclusions and open problems

Our 3^{n^2} bound on the cardinality of finite irreducible rational matrix semigroups (Theorem 28) breaks the barrier of $2^{\mathcal{O}(n^2 \log n)}$ suggested in previous works [24, 5, 2, 7]. Up to a constant in the exponent our bound is tight (Proposition 30). As discussed in the introduction, the largest finite rational $n \times n$ matrix groups are known explicitly, using the classification of finite simple groups. It would be similarly intriguing to identify the largest finite irreducible rational $n \times n$ matrix semigroups. By our results, they have $2^{\Theta(n^2)}$ elements.

While we now have a good understanding of the maximal cardinality of finite irreducible matrix semigroups, for their diameter there is still a gap between the best known lower bound of $2^{n+\Theta(\sqrt{n \log n})}$ [19] and the upper bound of $2^{\Theta(n^2)}$ implied by our work. As mentioned in the introduction, the gap for the mortality threshold is even bigger, since only polynomial lower bounds are known.

It would be also interesting to understand if the upper bound from Theorem 28 can be made more precise if it is also allowed to depend on the number of generators. The examples from our lower bounds have exponentially many generators. For transformation semigroups (equivalently, semigroups of matrices with entries in $\{0,1\}$ with exactly one nonzero entry in every row) this question was studied in [12]. The diameter of transformation semigroups with a bounded number of generators was studied in [23]. The cardinality of aperiodic transformation semigroups was studied in [6].

References

1 Jorge Almeida, Stuart Margolis, Benjamin Steinberg, and Mikhail Volkov. Representation theory of finite semigroups, semigroup radicals and formal language theory. *Transactions of the American Mathematical Society*, 361(3):1429–1461, 2009.

2 Jorge Almeida and Benjamin Steinberg. Matrix mortality and the Černý-Pin conjecture. In Volker Diekert and Dirk Nowotka, editors, *Developments in Language Theory, 13th International Conference, DLT 2009, Stuttgart, Germany, June 30 - July 3, 2009. Proceedings*, volume 5583 of *Lecture Notes in Computer Science*, pages 67–80. Springer, 2009. doi:10.1007/978-3-642-02737-6_5.

3 Paul C. Bell, Igor Potapov, and Pavel Semukhin. On the mortality problem: From multiplicative matrix equations to linear recurrence sequences and beyond. *Information and Computation*, 281:104736, 2021. doi:10.1016/J.IC.2021.104736.

4 Neil Berry, Artūras Dubickas, Noam D. Elkies, Bjorn Poonen, and Chris Smyth. The conjugate dimension of algebraic numbers. *Quarterly Journal of Mathematics*, 55(3):237–252, 2004.

5 Jean Berstel and Christophe Reutenauer. *Rational Series and Their Languages*. Revised electronic edition, 2008. Electronic edition of the 1988 English translation (Springer-Verlag), preface dated January 8, 2008.

6 Janusz A. Brzozowski and Marek Szykula. Large aperiodic semigroups. *International Journal of Foundations of Computer Science*, 26(7):913–932, 2015. doi:10.1142/S0129054115400067.

7 Georgina Bumpus, Christoph Haase, Stefan Kiefer, Paul-Ioan Stoienescu, and Jonathan Tanner. On the size of finite rational matrix semigroups. In Artur Czumaj, Anuj Dawar, and Emanuela Merelli, editors, *47th International Colloquium on Automata, Languages, and Programming, ICALP 2020, July 8-11, 2020, Saarbrücken, Germany (Virtual Conference)*, volume 168 of *LIPIcs*, pages 115:1–115:13. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2020. doi:10.4230/LIPIcs.ICALP.2020.115.

8 Julien Cassaigne, Vesa Halava, Tero Harju, and François Nicolas. Tighter undecidability bounds for matrix mortality, zero-in-the-corner problems, and more. *CoRR*, abs/1404.0644, 2014. arXiv:1404.0644.

9 Thomas Colcombet, Joël Ouaknine, Pavel Semukhin, and James Worrell. On reachability problems for low-dimensional matrix semigroups. In Christel Baier, Ioannis Chatzigiannakis, Paola Flocchini, and Stefano Leonardi, editors, *46th International Colloquium on Automata, Languages, and Programming, ICALP 2019, July 9-12, 2019, Patras, Greece*, volume 132 of *LIPIcs*, pages 44:1–44:15. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2019. doi:10.4230/LIPIcs.ICALP.2019.44.

10 Walter Feit. The orders of finite linear groups. Unpublished preprint.

11 Shmuel Friedland. The maximal orders of finite subgroups in $\mathrm{GL}_n(\mathbb{Q})$. *Proceedings of the American Mathematical Society*, 125(12):3519–3526, 1997.

12 Markus Holzer and Barbara König. On deterministic finite automata and syntactic monoid size. *Theor. Comput. Sci.*, 327(3):319–347, 2004. doi:10.1016/J.TCS.2004.04.010.

13 Stefan Kiefer and Andrew Ryzhikov. Efficiently computing the minimum rank of a matrix in a monoid of zero-one matrices. In Olaf Beyersdorff, Michal Pilipczuk, Elaine Pimentel, and Kim Thang Nguyen, editors, *42nd International Symposium on Theoretical Aspects of Computer Science, STACS 2025, March 4-7, 2025, Jena, Germany*, volume 327 of *LIPIcs*, pages 61:1–61:22. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2025. doi:10.4230/LIPIcs.STACS.2025.61.

14 Stefan Kiefer and Andrew Ryzhikov. The asymptotic size of finite irreducible semigroups of rational matrices. *CoRR*, abs/2601.01236, 2026. URL: https://arxiv.org/abs/2601.01236.

15 James Kuzmanovich and Andrey Pavlichenkov. Finite groups of matrices whose entries are integers. *The American Mathematical Monthly*, 109(2):173–186, 2002. URL: http://www.jstor.org/stable/2695329.

16 Rida Ait El Manssour, Roland Guttenberg, Nathan Lhote, Mahsa Shirmohammadi, and James Worrell. "Finite Rational Matrix Semigroups have at most Exponential Size", a talk at the Workshop on Loop Invariants and Algebraic Reasoning, 7 July 2025. Accessed: 21 September 2025. URL: https://georgekenison.github.io/uploads/guttenberg-liar25.pdf.

17 Henryk Minc. *Nonnegative matrices*. Wiley-Interscience, 1988.

18 Morris Newman. *Integral Matrices*, volume 45 of *Pure and Applied Mathematics*. Academic Press, New York and London, 1972.

19 Pavel Panteleev. Preset distinguishing sequences and diameter of transformation semigroups. In Adrian-Horia Dediu, Enrico Formenti, Carlos Martín-Vide, and Bianca Truthe, editors, *Language and Automata Theory and Applications - 9th International Conference, LATA 2015, Nice, France, March 2-6, 2015, Proceedings*, volume 8977 of *Lecture Notes in Computer Science*, pages 353–364. Springer, 2015. doi:10.1007/978-3-319-15579-1_27.

20 Mike Paterson. Unsolvability in 3×3 matrices. *Studies in Applied Mathematics*, 49:105–107, 1970.

21 Igor K. Rystsov. Reset words for commutative and solvable automata. *Theoretical Computer Science*, 172(1-2):273–279, 1997. doi:10.1016/S0304-3975(96)00136-3.

22 Andrew Ryzhikov. On shortest products for nonnegative matrix mortality. In Laura Kovács and Ana Sokolova, editors, *Reachability Problems – 18th International Conference, RP 2024, Vienna, Austria, September 25-27, 2024, Proceedings*, volume 15050 of *Lecture Notes in Computer Science*, pages 104–119. Springer, 2024. doi:10.1007/978-3-031-72621-7_8.

23 Andrew Ryzhikov. Careful synchronisation and the diameter of transformation semigroups with few generators. *CoRR*, abs/2506.13962, 2025. doi:10.48550/arXiv.2506.13962.

24 Marcel Paul Schützenberger. Finite counting automata. *Information and Control*, 5(2):91–107, 1962. doi:10.1016/S0019-9958(62)90244-9.

25 Benjamin Steinberg. *Representation theory of finite monoids.* Springer, 2016.

26 Boris Weisfeiler. On the size and structure of finite linear groups. Unpublished preprint, see https://arxiv.org/abs/1203.1960.

27 Boris Weisfeiler. Post-classification version of Jordan's theorem on finite linear groups. *Proceedings of the National Academy of Sciences of the United States of America*, 81:5278–5279, 1984.