A Ramsey Theorem for Finite Monoids

Ismaël Jecker
Institute of Science and Technology, Klosterneuburg, Austria

Abstract

Repeated idempotent elements are commonly used to characterise iterable behaviours in abstract models of computation. Therefore, given a monoid M, it is natural to ask how long a sequence of elements of M needs to be to ensure the presence of consecutive idempotent factors. This question is formalised through the notion of the *Ramsey function* R_M associated to M, obtained by mapping every $k \in \mathbb{N}$ to the minimal integer $\mathsf{R}_M(k)$ such that every word $u \in M^*$ of length $\mathsf{R}_M(k)$ contains k consecutive non-empty factors that correspond to the same idempotent element of M.

In this work, we study the behaviour of the Ramsey function R_M by investigating the *regular $\mathcal{D}$-length* of M, defined as the largest size $L(M)$ of a submonoid of M isomorphic to the set of natural numbers $\{1, 2, \ldots, L(M)\}$ equipped with the max operation. We show that the regular $\mathcal{D}$-length of M determines the degree of R_M, by proving that $k^{L(M)} \leq \mathsf{R}_M(k) \leq (k|M|^4)^{L(M)}$.

To allow applications of this result, we provide the value of the regular $\mathcal{D}$-length of diverse monoids. In particular, we prove that the full monoid of $n \times n$ Boolean matrices, which is used to express transition monoids of non-deterministic automata, has a regular $\mathcal{D}$-length of $\frac{n^2+n+2}{2}$.

2012 ACM Subject Classification Theory of computation → Algebraic language theory

Keywords and phrases Semigroup, monoid, idempotent, automaton

Digital Object Identifier 10.4230/LIPIcs.STACS.2021.44

Related Version *Full Version*: https://arxiv.org/abs/2101.05895

Funding This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 754411.

Acknowledgements I wish to thank Michaël Cadilhac, Emmanuel Filiot and Charles Paperman for their valuable insights concerning Green's relations.

1 Introduction

The algebraic approach to language theory was initiated by Schützenberger with the definition of the syntactic monoid associated to a formal language [14]. This led to several parallels being drawn between classes of languages and varieties of monoids, the most famous being that rational languages are characterised by finite syntactic monoids [12], and that star-free languages are characterised by finite aperiodic syntactic monoids [15]. These characterisations motivate the study of finite monoids as a way to gain some insight about automata. In this work, we focus on the following problem:

> *Given a finite monoid M and $k \in \mathbb{N}$, what is the minimal integer $R_M(k)$ such that every word $u \in M^*$ of length $R_M(k)$ contains k consecutive factors corresponding to the same idempotent element of M?*

The interest of this problem lies in the fact that when we model the behaviours of an abstract machine as elements of a monoid, repeated idempotent factors often characterise the behaviours that have good properties with respect to iteration. This can be used, for instance, to obtain pumping lemmas, as seen in [8] for weighted automata.

A partial answer to this problem is obtained by using Ramsey's Theorem [13] or Simon's Factorisation Forest Theorem [16] (these techniques are detailed in the full version), as both approaches provide upper bounds for $\mathsf{R}_M(k)$. However, neither approximation is precise:

38th International Symposium on Theoretical Aspects of Computer Science (STACS 2021).
Editors: Markus Bläser and Benjamin Monmege; Article No. 44; pp. 44:1–44:13
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Ramsey's theorem disregards the monoid structure, and the Factorisation Forest Theorem guarantees much more than what is required here. We prove a version of Ramsey's Theorem adapted to monoids, or, equivalently, a weaker version of the Forest Factorisation Theorem, that yields an improved bound relying on a parameter of monoids called the regular $\mathcal{D}$-length. We now present some examples, followed with an overview of the main concepts studied in this paper: the Ramsey function associated to a monoid and the regular $\mathcal{D}$-length.

1.1 Examples

We describe three families of monoids, along with the corresponding idempotent elements.

Max monoid. The max monoid H_n is the set $\{1, 2, \ldots, n\}$, equipped with the max operation. In this monoid, every element i is idempotent since $\max(i, i) = i$.

Transformation monoid. The (full) transformation monoid T_n is the set of all (partial) functions from a set of n elements into itself, equipped with the composition. See [3] for a detailed definition of T_n and its properties. Transformation monoids contain a wide range of idempotent elements. For instance, the identity function, mapping each element to itself, or the constant function f_i, mapping all elements to one fixed element i, are idempotent. In general, a function f is idempotent if and only if each element i of its range satisfies $f(i) = i$. Transformations are commonly used to express transition monoids of deterministic finite state automata, as in this setting each input letter acts as a function over the set of states.

Relation monoid. For non-deterministic automata, transition monoids are more complex: functions fail to model the behaviour of the input letters since a single state can transition towards several distinct states. We use the (full) relation monoid B_n of all $n \times n$ Boolean matrices (matrices with values in $\{0, 1\}$), equipped with the usual matrix composition (considering that $1 + 1 = 1$). There are plenty of idempotent matrices, for instance every diagonal matrix, or the full upper triangular matrix. Idempotent Boolean matrices are characterised in [9], they correspond to specific orders over the subsets of $\{1, 2, \ldots, n\}$.

1.2 Ramsey function

Given a finite monoid M, the *Ramsey function* R_M associated to M maps each $k \in \mathbb{N}$ to the minimal integer $\mathsf{R}_M(k)$ such that every sequence of elements of M of length $\mathsf{R}_M(k)$ contains k non-empty consecutive factors that all correspond to the same idempotent element of M.

Related work. There are several known methods to approximate the Ramsey function R_M of a monoid M. Ramsey's Theorem and Simon's Factorisation Forest Theorem are commonly used, however, as stated before, these approaches are too general to obtain a precise bound. The value of $\mathsf{R}_M(k)$ is studied in [4] in the particular case $k = 1$. The authors prove that for a monoid M that contains N non-idempotent elements, $\mathsf{R}_M(1) \leq 2^N - 1$. No general related lower bound is proved, but they show that for every $N \in \mathbb{N}$, there exists a monoid M_N with N non-idempotent elements that actually reaches the upper bound: $\mathsf{R}_{M_N}(1) = 2^N - 1$.

Our contributions. We prove new bounds for R_M by following a different approach: instead of focusing on the non-idempotent elements of M, we study its idempotent elements, and the way in which they interact. In Section 3, we start by considering two specific cases where the exact value of the Ramsey function is easily obtained. First, for a group $\mathcal{G}$, the

Ramsey function is polynomial with respect to the size: $\mathsf{R}_{\mathcal{G}}(k) = k|\mathcal{G}|$. Second, we call *max monoid* H_n the set $\{1, 2, \ldots, n\}$ equipped with the max operation, and we show that here the Ramsey function is exponential with respect to the size: $\mathsf{R}_{H_n}(k) = k^n$. The later result implies that k^n is a lower bound for the Ramsey function of every monoid M that has H_n as a submonoid. We prove a related upper bound: We define the *regular $\mathcal{D}$-length* $L(M)$ of M as the size of the largest max monoid $H_{L(M)}$ embedded in M, and show the following:

▶ **Theorem 1.** *Every monoid M of regular $\mathcal{D}$-length L satisfies $k^L \leq \mathsf{R}_M(k) \leq (k|M|^4)^L$.*

Stated differently: every word $u \in M^*$ of length $(k|M|^4)^L$ contains k consecutive non-empty factors corresponding to the same idempotent element of M, and, conversely, there exists a word $u_M \in M^*$ of length $k^L - 1$ that does not contain k consecutive non-empty factors corresponding to the same idempotent element. Note that while the gap between the lower and upper bound is still wide, this shows that the degree of the Ramsey function R_M is determined by the regular $\mathcal{D}$-length of M.

1.3 Regular $\mathcal{D}$-length

Theorem 1 states that the degree of the Ramsey function of a monoid M is determined by the regular $\mathcal{D}$-length of M, which is the size of the largest max monoid embedded in M. We now show that for transformation monoids and relation monoids, the regular $\mathcal{D}$-length is exponentially shorter than the size. Let us begin by mentioning an equivalent definition of the regular $\mathcal{D}$-length in terms of Green's relations. While this alternative definition is not used in the proofs presented in this paper, it allows us to immediately obtain the regular $\mathcal{D}$-length of monoids whose Green's relations are known.

Alternative definition. The regular $\mathcal{D}$-length of a monoid M is the size of its largest chain of regular $\mathcal{D}$-classes. A $\mathcal{D}$-class of M is an equivalence class of the preorder $\leq_{\mathcal{D}}$ defined by $m \leq_{\mathcal{D}} m'$ if $m = s \cdot m' \cdot t$ for some $s, t \in M$, and it is called regular if it contains at least one idempotent element (see [11] for more details). The equivalence between both definitions is proved in the full version.

Computing the regular $\mathcal{D}$-length. The following table compares the size and the regular $\mathcal{D}$-length of the monoids mentioned earlier. The entries corresponding to the sizes are considered to be general knowledge. We detail below the row listing the regular $\mathcal{D}$-lengths.

Monoid	G	H_n	T_n	B_n
Size	$\|G\|$	n	$(n+1)^n$	$2^{(n^2)}$
Regular $\mathcal{D}$-length	1	n	$n+1$	$\frac{n^2+n+2}{2}$

First, every group $\mathcal{G}$ contains a single idempotent element (the neutral element), hence its regular $\mathcal{D}$-length is 1. Then, using the definition of the regular $\mathcal{D}$-length in terms of embedded max monoid, we immediately obtain that $L(H_n)$ is equal to n. We get the next entry using the definition of the regular $\mathcal{D}$-length in terms of chain of $\mathcal{D}$-classes: The transformation monoid T_n is composed of a single chain of $n + 1$ $\mathcal{D}$-classes that are all regular [3], hence its regular $\mathcal{D}$-length is $n + 1$.

Finally, for the relation monoid B_n, the situation is not as clear: the $\mathcal{D}$-classes do not form a single chain, and some of them are not regular. Determining the exact size of the largest chain of $\mathcal{D}$-classes (note the absence of "regular") is still an open question, yet it is known to grow exponentially with respect to n: a chain of $\mathcal{D}$-classes whose size is the Fibonacci number $F_{n+3} - 1$ is constructed in [2], and, conversely, the upper bound $2^{n-1} + n - 1$ is proved in [6] (and slightly improved in [7, 17, 5]). Our second main result is that, as long as we only consider chains of *regular* $\mathcal{D}$-classes, we can obtain the precise value of the maximal length, and, somewhat surprisingly, it is only quadratic in n:

▶ **Theorem 2.** *The regular $\mathcal{D}$-length of the monoid of $n \times n$ Boolean matrices is $\frac{n^2+n+2}{2}$.*

Therefore, the regular $\mathcal{D}$-length of a transformation monoid is exponentially smaller than its size, and the regular $\mathcal{D}$-length of a relation monoid is even exponentially smaller than its largest chain of $\mathcal{D}$-classes. For such kind of monoids, Theorem 1 performs considerably better than previously known methods to find idempotent factors. For instance, it was used in [10] to close the complexity gap left in [1] for the problem of deciding whether the function defined by a given two-way word transducer is definable by a one-way transducer.

2 Definitions and notations

We define in this section the notions that are used throughout the paper. We denote by $\mathbb{N}$ the set $\{0, 1, 2, \ldots\}$, and for all $i \leq j \in \mathbb{N}$ we denote by $[i, j]$ the interval $\{i, i+1, \ldots, j\}$.

Monoids. A (finite) *semigroup* $(S, \cdot)$ is a finite set S equipped with a binary operation $\cdot : S \times S \to S$ that is *associative*: $(s_1 \cdot s_2) \cdot s_3 = s_1 \cdot (s_2 \cdot s_3)$ for every $s_1, s_2, s_3 \in S$. A *monoid* is a semigroup $(M, \cdot)$ that contains a *neutral element* 1_M: $m \cdot 1_M = m = 1_M \cdot m$ for all $m \in M$. A *group* is a monoid $(\mathcal{G}, \cdot)$ in which every element $g \in \mathcal{G}$ has an *inverse element* $g^{-1} \in \mathcal{G}$: $g \cdot g^{-1} = 1_{\mathcal{G}} = g^{-1} \cdot g$. We always denote the semigroup operation with the symbol $\cdot$. As a consequence, we identify a semigroup $(S, \cdot)$ with its set of elements S.

An element e of a semigroup S is called *idempotent* if it satisfies $e \cdot e = e$. Note that whereas a finite semigroup does not necessarily contain a neutral element, it always contains at least one idempotent element: iterating any element $s \in S$ eventually yields an idempotent element, called the *idempotent power* of s, and denoted $s^{\#} \in S$.

A *homomorphism* between two monoids M and M' is a function $\varphi : M \to M'$ preserving the monoid structure: $\varphi(m_1 \cdot m_2) = \varphi(m_1) \cdot \varphi(m_2)$ for all $m_1, m_2 \in M$ and $\varphi(1_M) = \varphi(1_{M'})$. A *monomorphism* is an injective homomorphism, an *isomorphism* is a bijective homomorphism.

Ramsey decomposition. Let M be a monoid. A *word* over M is a finite sequence $u = m_1 m_2 \ldots m_n \in M^*$ of elements of M. The *length* of u is its number of symbols $|u| = n \in \mathbb{N}$. We enumerate the positions between the letters of u starting from 0 before the first letter, until $|u|$ after the last letter. A *factor* of u is a subsequence of u composed of the letters between two such positions i and j: $u[i, j] = m_{i+1} m_{i+2} \ldots m_j \in M^*$ for some $0 \leq i \leq j \leq |u|$ (where $u[i, j] = \varepsilon$ if $i = j$). We denote by $\pi(u)$ the element $1_M \cdot m_1 \cdot m_2 \cdot \ldots \cdot m_n \in M$, and we say that u *reduces* to $\pi(u)$. For every integer $k \in \mathbb{N}$, a *k-decomposition* of u is a decomposition of u in $k + 2$ factors such that the k middle ones are non-empty:

$$u = x y_1 y_2 \ldots y_k z, \text{ where } x, z \in M^*, \text{ and } y_i \in M^+ \text{ for every } 1 \leq i \leq k.$$

A k-decomposition is called *Ramsey* if all the middle factors $y_1, y_2, \ldots, y_k$ reduce to the same idempotent element $e \in M$. For instance, a word has a Ramsey 1-decomposition if and only if it contains a factor that reduces to an idempotent element. The *Ramsey function* $\mathsf{R}_M : \mathbb{N} \to \mathbb{N}$ associated to M is the function mapping each $k \in \mathbb{N}$ to the minimal $\mathsf{R}_M(k) \in \mathbb{N}$ such that every word $u \in M^*$ of length $\mathsf{R}_M(k)$ has a Ramsey k-decomposition.

3 Ramsey decompositions

In this section, we bound the Ramsey function R_M associated to a monoid M. As a first step we consider two basic cases for which the exact value of the Ramsey function is obtained: in Subsection 3.1 we show that every group $\mathcal{G}$ satisfies $\mathsf{R}_{\mathcal{G}}(k) = k|\mathcal{G}|$, and in Subsection 3.2 we show that every max monoid H_n (obtained by equipping the first n positive integers with the max operation) satisfies $\mathsf{R}_{H_n}(k) = k^n$. Finally, in Subsection 3.3, we prove bounds in the general case by studying the submonoids of M isomorphic to a max monoid.

3.1 Group: prefix sequence algorithm

We show that in a group, the Ramsey function is polynomial with respect to the size.

▶ **Proposition 3.** *For every group $\mathcal{G}$, $\mathsf{R}_{\mathcal{G}}(k) = k|\mathcal{G}|$ for all $k \in \mathbb{N}$.*

We fix for this subsection a group $\mathcal{G}$ and $k \in \mathbb{N}$. We begin by proving an auxiliary lemma, which we then apply to prove matching bounds for $\mathsf{R}_{\mathcal{G}}(k)$: First, we define an algorithm that extracts a Ramsey k-decomposition out of every word of length $k|\mathcal{G}|$. Then, we present the construction of a witness $u_{\mathcal{G}} \in \mathcal{G}^*$ of length $k|\mathcal{G}| - 1$ that has no Ramsey k-decompositions.

Key lemma. In a group, the presence of inverse elements allows us to establish a correspondence between the factors of a word $u \in \mathcal{G}^*$ that reduce to the neutral element, and the pairs of prefixes of u that both reduce to the same element.

▶ **Lemma 4.** *Two prefixes $u[0,i]$ and $u[0,j]$ of a word $u \in \mathcal{G}^*$ reduce to the same element if and only if $u[i,j]$ reduces to the neutral element of $\mathcal{G}$.*

Proof. Let $u \in \mathcal{G}^*$ be a word. The statement is a direct consequence of the fact that for every $0 \leq i \leq j \leq |u|$, $\pi(u[0,i]) \cdot \pi(u[i,j]) = \pi(u[0,j])$: If $\pi(u[0,i]) = \pi(u[0,j])$, then

$$\pi(u[i,j]) = \pi(u[0,i])^{-1} \cdot \pi(u[0,j]) = \pi(u[0,i])^{-1} \cdot \pi(u[0,i]) = 1_{\mathcal{G}}.$$

Conversely, if $\pi(u[i,j]) = 1_{\mathcal{G}}$, then

$$\pi(u[0,i]) = \pi(u[0,i]) \cdot 1_{\mathcal{G}} = \pi(u[0,i]) \cdot \pi(u[i,j]) = \pi(u[0,j]).$$ ◀

Algorithm. We define an algorithm constructing Ramsey k-decompositions.

```
ALG_1: Start with $u \in \mathcal{G}^*$ of length $k|\mathcal{G}|$;
    a. Compute the $k|\mathcal{G}|+1$ prefixes $\pi(u[0,0])$, $\pi(u[0,1])$, ..., $\pi(u[0,|u|])$ of $u$;
    b. Find $k+1$ indices $i_0$, $i_1$, ..., $i_k$ such that all the $\pi(u[0,i_j])$ are equal;
    c. Return the Ramsey $k$-decomposition $u = u[i_0,i_1]u[i_1,i_2]\dots u[i_{k-1},i_k]$.
```

Since Lemma 4 ensures that every pair of elements i_j, i_{j+1} identified at step 2 satisfies $\pi(u[i_j,i_{j+1}]) = 1_{\mathcal{G}}$, we are guaranteed that the returned k-decomposition is Ramsey.

Witness. We build a word $u_{\mathcal{G}} \in \mathcal{G}^*$ of length $k|\mathcal{G}| - 1$ that has no Ramsey k-decompositions. Let $v = a_1 a_2 \dots a_{k|\mathcal{G}|} \in \mathcal{G}^*$ be a word of length $k|\mathcal{G}|$, starting with the letter $1_{\mathcal{G}}$, and containing exactly k times each element of $\mathcal{G}$. For instance, given an enumeration $g_1, g_2, \dots, g_{|\mathcal{G}|}$ of the elements of $\mathcal{G}$ starting with $g_1 = 1_{\mathcal{G}}$, we can simply pick $v = g_1^k g_2^k \dots g_{|\mathcal{G}|}^k$. Now let $u_{\mathcal{G}} = b_1 b_2 \dots b_{k|\mathcal{G}|-1}$ be the word whose sequence of reduced prefixes is v: for every $1 \leq i \leq k|\mathcal{G}| - 1$, the letter b_i is equal to $a_i^{-1} \cdot a_{i+1}$. Then for every k-decomposition of $u_{\mathcal{G}}$, at least one of the factors do not reduce to the neutral element of $\mathcal{G}$, since otherwise Lemma 4 would imply the existence of $k+1$ identical letters in v, which is not possible by construction. As a consequence, $u_{\mathcal{G}}$ has no Ramsey k-decompositions.

3.2 Max monoid: divide and conquer algorithm

Given an integer $n \in \mathbb{N}$, the *max monoid*, denoted H_n, is the monoid over the set $\{1, 2, \ldots, n\}$ with the associative operation $i \cdot j = \max(i, j)$. Whereas in a group only the neutral element is idempotent, each element i of the max monoid H_n is idempotent since $\max(i, i) = i$. As a result of this abundance of idempotent elements, an exponential bound is required to ensure the presence of consecutive factors reducing to the same idempotent element.

▶ **Proposition 5.** *For every max monoid H_n, $R_{H_n}(k) = k^n$ for all $k \in \mathbb{N}$.*

The proof is done in two steps: we first define an algorithm that extracts a Ramsey k-decomposition out of every word of length k^n, and then we present the construction of a witness u_n of length $k^n - 1$ that has no Ramsey k-decompositions.

Algorithm. We define an algorithm that extracts a Ramsey k-decompositions out of each word $u \in H_n^*$ of length k^n. It is a basic divide and conquer algorithm: we divide the initial word u into k equal parts. If each of the k parts reduces to n, they form a Ramsey k-decomposition since n is an idempotent element. Otherwise, one part does not contain the maximal element $n \in H_n$, and we start over with it. Formally,

```
ALG2: Start with $u \in H_n^*$ of length $k^n$, initialize $j$ to $n$. While $j > 0$, repeat the following:
    a. Split $u$ into $k$ factors $u_1$, $u_2$, $\ldots$, $u_k$ of length $k^{j-1}$;
    b. If every $u_i$ contains the letter $j$, return the Ramsey $k$-decomposition $u = u_1 u_2 \ldots u_k$;
    c. If $u_i$ does not contain $j$ for some $1 \leq i \leq j$, decrement $j$ by 1 and set $u := u_i \in H_{j-1}^*$.
```

The algorithm is guaranteed to eventually return a Ramsey k-decomposition: if the n^{th} cycle of the algorithm is reached, it starts with a word of length k whose letters are in the monoid H_1, which only contains the letter 1, hence the algorithm will go to step b.

Witness. We construct an infinite sequence of words $u_1, u_2, \ldots \in \mathbb{N}^*$ such that for all $n \in \mathbb{N}$,
(a) $u_n \in H_n$ satisfies $|u_n| = k^n - 1$ and
(b) u_n has no Ramsey k-decompositions.
Let

$$\begin{array}{lcl} u_1 & = & 1^{k-1} \in H_1^*, \\ u_n & = & (u_{n-1}n)^{k-1}u_{n-1} \in H_n^* \quad \text{for every } n > 1. \end{array}$$

For every $n > 1$, the word u_n is defined as k copies of u_{n-1} separated by the letter n. We prove by induction that the two conditions are satisfied by each word of the sequence. The base case is immediate: the word u_1 has length $k - 1$, and as a consequence has no decomposition into k nonempty factors. Now suppose that $n > 1$, and that u_{n-1} satisfies the two properties. Then u_n has the required length:

$$|u_n| = (k-1)(|u_{n-1}| + 1) + |u_{n-1}| = (k-1)k^{n-1} + k^{n-1} - 1 = k^n - 1.$$

To conclude, we show that every k-decomposition

$$u_n = x y_1 y_2 \ldots y_k z, \text{ with } y_i \in H_n^+ \text{ for all } 1 \leq i \leq k \tag{1}$$

is not Ramsey. Let y be the factor $y_1 y_2 \ldots y_k$ of u_n, and consider the two following cases:

- If $\pi(y) \neq n$, none of the y_i contains the letter n, hence y is factor of one of the factors u_{n-1} of u_n. Therefore, by the induction hypothesis, Decomposition (1) is not Ramsey.
- If $\pi(y) = n$, since u_n contains only $k - 1$ copies of the letter n, one of the factors y_i does not contain n for $1 \leq i \leq k$. Then $\pi(y) \neq \pi(y_i)$, hence Decomposition (1) is not Ramsey.

▶ **Example 6.** Here are the first three words of the sequence in the cases $k = 2$ and $k = 3$:

$$\begin{array}{llll} k = 2: & u_1 = 1 & u_2 = 121 & u_3 = 1213121, \\ k = 3: & u_1 = 11 & u_2 = 11211211 & u_3 = 11211211311211211311211211. \end{array}$$

3.3 General setting

We saw in the previous subsection that for the max monoid H_n, words of length exponential with respect to n are required to guarantee the presence of Ramsey decompositions (Proposition 5). Note that the same lower bound applies to every monoid M that contains a copy of H_n as submonoid. We now show that we can also obtain an upper bound for $\mathsf{R}_M(k)$ by studying the submonoids of M isomorphic to a max monoid. We formalise this idea through the notion of regular $\mathcal{D}$-length of a monoid.

Regular $\mathcal{D}$-length. The *regular $\mathcal{D}$-length* of a monoid M, denoted $L(M)$, is the size of the largest max monoid embedded in M. Formally, it is the largest $\ell \in \mathbb{N}$ such that there exists a monomorphism (i.e. injective monoid homomorphism) $\varphi : H_\ell \to M$. We now present the main theorem of this section, which states that for every monoid M, the degree of $\mathsf{R}_M(k)$ is determined by the regular $\mathcal{D}$-length of M.

▶ **Theorem 1.** *Every monoid M of regular $\mathcal{D}$-length L satisfies $k^L \leq \mathsf{R}_M(k) \leq (k|M|^4)^L$.*

Let us fix for the whole subsection a monoid M of regular $\mathcal{D}$-length $L(M)$ and an integer $k \in \mathbb{N}$. The lower bound is a corollary of Proposition 5: the max monoid $H_{L(M)}$ has a witness $u_{L(M)}$ of length $k^{L(M)} - 1$ that has no Ramsey k-decompositions (its construction is presented in the previous subsection). Then, by definition of the regular $\mathcal{D}$-length, there exists a monomorphism $\varphi : H_{L(M)} \to M$, and applying φ to $u_{L(M)}$ letter by letter yields a witness $u'_{L(M)} \in M^*$ of length $k^{L(M)} - 1$ that has no Ramsey k-decompositions.

The rest of the subsection is devoted to the proof of the upper bound. We begin by defining an auxiliary algorithm that extracts from each long enough word a decomposition where the prefix and suffix absorb the middle factors. Then, we define our main algorithm which, on input $u \in M^*$ of length $(k|M|^4)^n$ for some $n \in \mathbb{N}$, either returns a Ramsey k-decomposition of u, or a copy of the max monoid H_{n+1} embedded in M. In particular, if n is equal to the regular $\mathcal{D}$-length $L(M)$ of M, we are guaranteed to obtain a Ramsey k-decomposition.

Auxiliary algorithm. We define an algorithm which, on input $u \in M^*$ of length $k|M|^2$, returns a k-decomposition

$$u = xy_1y_2\dots y_kz, \text{ where } x, z \in M^*, \text{ and } y_i \in S^+ \text{ for every } 1 \leq i \leq k$$

such that for every $1 \leq i \leq k$, both x and z are able to absorb the factor y_i: $\pi(xy_i) = \pi(x)$ and $\pi(y_iz) = \pi(z)$. This is done as follows: since u is a word of length $k|M|^2$, it can be split into $k|M|^2 + 1$ distinct prefix-suffix pairs. Then $k + 1$ of these pairs reduce to the same pair of elements of M, which immediately yields the desired decomposition. Formally,

```
ALG3: Start with $u \in M^*$ of length $k|M|^2$;

1. a. Compute the $k|M|^2 + 1$ prefixes $\pi(u[0,0]), \pi(u[0,1]), \dots, \pi(u[0,|u|]) \in M$ of $u$,
   b. Compute the $k|M|^2 + 1$ suffixes $\pi(u[0,|u|]), \pi(u[1,|u|]), \dots, \pi(u[|u|,|u|]) \in M$ of $u$,
   c. Identify $k + 1$ indices $s_0, s_1, \dots, s_k$ such that
      (1) all the $\pi(u[0,s_i])$ are equal,
      (2) all the $\pi(u[s_i,|u|])$ are equal;
2. Set $x = u[0,s_0]$, $z = u[s_k,|u|]$, and $y_i = u[s_{i-1},s_i]$ for every $1 \leq i \leq k$;

3. Return the $k$-decomposition $xy_1y_2\dots y_kz$ of $u$.
```

Main algorithm. We define an algorithm extracting Ramsey k-decompositions. Over an input $u \in M^*$ of length $(k|M|^4)^n$ for $n \in \mathbb{N}$, the algorithm works by defining gradually shorter words $u_n, u_{n-1}, \ldots \in M^*$, where each u_j has length $(k|M|^4)^j$, along with a sequence of idempotent elements $e_{n+1}, e_n, \ldots \in M$. Starting with $u_n = u$, we define e_{n+1} as the idempotent power of some well chosen factors of u_n. We then consider k consecutive factors of u_n. If all of them reduce to e_{n+1}, they form a Ramsey k-decomposition, and we are done. Otherwise, we pick a factor u_{n-1} that does not reduce to e_{n+1}, and we start over. This continues until either a Ramsey k-decomposition is found, or n cycles are completed. In the later case, we show that the function $\varphi : H_{n+1} \to M$ mapping i to e_i is a monomorphism.

```
ALG4: Start with $u \in M^*$ of length $(k|M|^4)^n$. Initialize $u_n$ to $u$ and $j$ to $n$.

- While $j > 0$, repeat the following:
  1. a. Call ALG3 to get an $m$-decomposition $u_j = xy_1y_2 \ldots y_m z$, where $m = k^j|M|^{4j-2}$;
     b. Set $v \coloneqq \pi(y_1)\pi(y_2)\ldots\pi(y_m) \in M^*$;
  2. a. Call ALG3 to get an $m'$-decomposition $v = x'y'_1y'_2 \ldots y'_{m'}z'$, where $m' = k^j|M|^{4j-4}$;
     b. Set $w \coloneqq \pi(y'_1)\pi(y'_2)\ldots\pi(y'_{m'}) \in M^*$, and set $e_{j+1} \coloneqq (\pi(z'x'))^\#$;
  3. a. Split $w$ into $k$ factors $y''_1$, $y''_2$, ..., $y''_k$ of length $(k|M|^4)^{j-1}$;
     b. If every $y''_i$ satisfies $\pi(y_i) = e_{j+1}$, then $w = y''_1y''_2\ldots y''_k$ is a Ramsey decomposition.
        Return the corresponding Ramsey $k$-decomposition of $u$;
     c. If $\pi(y''_i) \neq e_{j+1}$ for some $1 \leq i \leq n$, set $u_{j-1} \coloneqq y''_i$, and decrement $j$ by 1.
- Set $e_1 = 1_M$, and return the idempotent elements $e_1, e_2, \ldots, e_{n+1} \in M$.
```

Step 1. We use the auxiliary algorithm to obtain a decomposition $u_j = xy_1y_2\ldots y_m z$, and we build v by concatenating the reductions of the y_i. Since both x and z absorb each y_i, and in step 2b we define e_{j+1} as the idempotent power of reduced factors of v:

$$\text{The word } u_j\text{, its prefix } x \text{ and its suffix } z \text{ satisfy } \pi(u_j) = \pi(xz) = \pi(x)\cdot e_{j+1}\cdot \pi(z). \quad (1)$$

Step 2. We use the auxiliary algorithm to get a decomposition $u' = x'y'_1y'_2\ldots y'_{m'}z'$, we build w by concatenating the reductions of the y'_i, and we set e_{j+1} as the idempotent power of $\pi(z'x')$. As both x' and z' absorb each y'_i, and in step 3c we define u_{j-1} as a factor of w:

$$\text{For every factor } y \text{ of } u_{j-1},\ e_{j+1}\cdot\pi(y) = e_{j+1} = \pi(y)\cdot e_{j+1}. \quad (2)$$

Step 3. We divide w into k factors of equal length. If each of them reduces to e_{j+1}, they form a Ramsey k-decomposition of w. As w is obtained form u by iteratively reducing factors and dropping prefixes and suffixes, this decomposition can be transferred back to a Ramsey k-decomposition of $u = u_n$. If one factor does not reduce to e_{j+1}, we assign its value to u_{j-1}. Therefore:

$$\text{The word } u_{j-1} \text{ does not reduce to } e_{j+1}. \quad (3)$$

Proof of correctness. To prove that the algorithm behaves as intended, we show that if it completes n cycles without returning a Ramsey k-decomposition, then the function $\varphi : H_{n+1} \to M$ defined by $\varphi(j) = e_j$ is a monomorphism. Since e_j is the idempotent power of reduced factors of u_{j-1} for all $1 \leq j \leq n$, Equation (2) yield that $e_{j+1}\cdot e_j = e_{j+1} = e_j \cdot e_{j+1}$.

Therefore φ is a homomorphism. We conclude by showing that it is injective. Suppose, towards building a contradiction, that $\varphi(j) = e_j = e_i = \varphi(i)$ for some $1 \leq j < i \leq n$. Since φ is a homomorphism, all the intermediate elements collapse: in particular $e_j = e_{j+1}$. Then

$$\pi(u_{j-1}) \underset{(1)}{=} \pi(x) \cdot e_j \cdot \pi(z) = \pi(x) \cdot e_{j+1} \cdot \pi(z) \underset{(2)}{=} e_{j+1},$$

which cannot hold by Equation (3).

4 Regular $\mathcal{D}$-length of the monoid of Boolean matrices

A Boolean matrix is a matrix A whose components are Boolean elements: $A_{ij} \in \{0,1\}$. The (full) *Boolean matrix monoid* B_n is the set of all $n \times n$ Boolean matrices, equipped with the matrix composition defined as follows: $(A \cdot B)_{ik} = 1$ if and only if there exists $j \in [1,n]$ satisfying $A_{ij} = B_{jk} = 1$. This fits the standard matrix multiplication if we consider that $1 + 1 = 1$: addition of Boolean elements is the OR operation, and multiplication is the AND operation. The main contribution of this section is the following theorem.

▶ **Theorem 2.** *The regular $\mathcal{D}$-length of the monoid of $n \times n$ Boolean matrices is $\frac{n^2+n+2}{2}$.*

The proof is split in two parts. We prove the upper bound by studying the structure of the idempotent elements of B_n (Subsection 4.1). Then, we prove the lower bound by constructing a monomorphism from the max monoid of size $\frac{n^2+n+2}{2}$ into B_n (Subsection 4.2). We begin by introducing definitions tailored to help us in the following demonstrations.

Stable matrix. A Boolean matrix $A \in B_n$ is called *stable* if for each component A_{ik} equal to 1, there exists $j \in [1,n]$ satisfying $A_{ij} = A_{jj} = A_{jk} = 1$. Idempotent matrices are stable (see the full version).

Positive set. A (maximal) *positive set* of an idempotent matrix $A \in B_n$ is a maximal set $I \subseteq [1,n]$ such that all the corresponding components of A are 1: $A_{ij} = 1$ for all $i,j \in I$, and for every $k \in [1,n] \setminus I$, there exists $i \in I$ such that $A_{ik} = 0$ or $A_{ki} = 0$. The positive sets of an idempotent matrix are disjoint (see the full version), hence A has at most n positive sets.

Free pair. For each idempotent matrix $A \in B_n$ we define the relation $\rightarrow_A$ on $[1,n]$ as follows: given $i,j \in [1,n]$, we have $i \rightarrow_A j$ if for all $i_2, j_2 \in [1,n]$, $A_{i_2 i} = 1 = A_{j j_2}$ implies $A_{i_2 j_2} = 1$. A *free pair* of A is a set of two distinct elements $i,j \in [1,n]$ incomparable by $\rightarrow_A$: $i \not\rightarrow_A j$ and $j \not\rightarrow_A i$. Note that A has at most $\frac{n(n-1)}{2}$ free pairs (all sets of two distinct elements in $[1,n]$). Let us state some observations concerning $\rightarrow_A$ that follow immediately from the definition (see the full version for the proofs). First, as A is idempotent, $\rightarrow_A$ is reflexive. However, it might not be transitive. Moreover, for every component A_{ij} of A equal to 1, we have that $i \rightarrow_A j$. The converse implication is not true, as shown by the following example. Finally, for every $i \in [1,n]$, if the i^{th} row contains no 1, i.e., $A_{ik} = 0$ for all $k \in [1,n]$, then $i \rightarrow_A j$ for every $j \in [1,n]$. Conversely, if the i^{th} column contains no 1, then $j \rightarrow_A i$ for every $j \in [1,n]$.

▶ **Example 7.** We depict below a submonoid of B_4 generated by two matrices A and B. The six elements of this submonoid, including the identity matrix $D \in B_n$, are all idempotent. Under each matrix, we list its positive sets. We then compute the corresponding free pairs.

$$\begin{array}{cccccc}
\begin{matrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{matrix} &
\begin{matrix}1&0&1&0\\0&1&0&1\\1&0&1&0\\0&1&0&1\end{matrix} &
\begin{matrix}1&1&1&1\\0&0&0&1\\0&0&0&1\\0&0&0&1\end{matrix} &
\begin{matrix}1&1&1&1\\0&0&0&1\\1&1&1&1\\0&0&0&1\end{matrix} &
\begin{matrix}1&1&1&1\\0&1&0&1\\0&1&0&1\\0&1&0&1\end{matrix} &
\begin{matrix}1&1&1&1\\0&1&0&1\\1&1&1&1\\0&1&0&1\end{matrix} \\
D & A & B & A \cdot B & B \cdot A & A \cdot B \cdot A \\
\{1\},\{2\},\{3\},\{4\} & \{1,3\},\{2,4\} & \{1\},\{4\} & \{1,3\},\{4\} & \{1\},\{2,4\} & \{1,3\},\{2,4\}
\end{array}$$

Every pair is free in D since the relation $\rightarrow_D$ is the identity: given two distinct elements $i, j \in [1, n]$, we have $D_{ii} = 1 = D_{jj}$, yet $D_{ij} = 0$, hence $i \not\rightarrow_D j$. On the contrary, the four matrices B, $A \cdot B$, $B \cdot A$ and $A \cdot B \cdot A$ has no free pairs: the relation $\rightarrow_B$ only lacks $(4, 1)$, $\rightarrow_{A \cdot B}$ only lacks $(4, 1)$ and $(4, 3)$, $\rightarrow_{B \cdot A}$ only lacks $(2, 1)$ and $(4, 1)$, $\rightarrow_{A \cdot B \cdot A}$ only lacks $(2, 1)$, $(4, 1)$ and $(4, 3)$. Finally, for A, the relation $\rightarrow_A$ is the union of the identity and the four pairs $\{(1, 3), (3, 1), (2, 4), (4, 2)\}$, which yields the free pairs $\{1, 2\}$, $\{1, 4\}$, $\{2, 3\}$ and $\{3, 4\}$.

4.1 Upper bound

To prove the upper bound of Theorem 2, we show that every monomorphism $\varphi : H_m \to B_n$ satisfies $m \leq \frac{n^2+n+2}{2}$. To this end, we study the sequence of matrices $s_\varphi = A_1, A_2, \dots, A_m$ obtained by listing the elements $\varphi(i) = A_i$ of the image of φ. Note that all the elements of s_φ are distinct as φ is injective, and $A_i \cdot A_{i+1} = A_{i+1} = A_{i+1} \cdot A_i$ for all $1 \leq i < m$ as φ is a homomorphism. We introduce three lemmas that imply interesting properties of every pair A_i, A_{i+1} of successive matrices of s_φ. First, Lemma 8 shows that every positive set of A_{i+1} contains a positive set of A_i. Therefore, since positive sets are disjoint, the number of positive sets can never increase along s_φ. Second, Lemma 9 shows that every free pair of A_{i+1} is also a free pair of A_i. As a consequence, the number of free pairs can never increase along s_φ. Finally, Lemma 10 shows that either the number of positive sets or free pairs differs between A_i and A_{i+1}, as otherwise these two matrices would be equal.

Combining the three lemmas yields that between each pair of successive matrices of s_φ, neither the number of positive sets nor the number of free pairs increases, and at least one decreases. This immediately implies the desired upper bound: as the number of positive sets of matrices of B_n ranges from 0 to n and the number of free pairs ranges from 0 to $\frac{n(n-1)}{2}$, s_φ contains at most $n + \frac{n(n-1)}{2} + 1 = \frac{n^2+n+2}{2}$ matrices. To conclude, we now proceed with the formal statements and the proofs of the three lemmas.

▶ **Lemma 8.** *Let A and B be two idempotent matrices of B_n satisfying $A \cdot B = B = B \cdot A$. Then every positive set of B contains a positive set of A.*

Proof. Let us pick two idempotent matrices $A, B \in B_n$ satisfying $A \cdot B = B = B \cdot A$. If B has no positive sets, the statement is trivially satisfied. Now let us suppose that B has at least one positive set $I \subseteq [1, n]$. We show the existence of a positive set $J \subseteq I$ of A.

Since I is not empty by definition, it contains an element i, and $B_{ii} = 1$. Then, as $B = B \cdot A$, there exists $k \in [1, n]$ satisfying $B_{ik} = A_{ki} = 1$. Moreover, as A is stable, there exists $j \in [1, n]$ satisfying $A_{kj} = A_{jj} = A_{ji} = 1$. In particular, $A_{jj} = 1$, hence A has a positive set J containing j. Then, for every $i_2 \in I$ and every $j_2, j_3 \in J$, we obtain

$$\begin{aligned}
&B_{j_2 i_2} = (A \cdot A \cdot B)_{j_2 i_2} = 1 \text{ since } A_{j_2 j} = A_{ji} = B_{i i_2} = 1,\\
&B_{i_2 j_3} = (B \cdot B \cdot A \cdot A)_{i_2 j_3} = 1 \text{ since } B_{i_2 i} = B_{ik} = A_{kj} = A_{j j_3} = 1,\\
&B_{j_2 j_3} = (B \cdot B)_{j_2 j_3} = 1 \text{ since } B_{j_2 i_2} = B_{i_2 j_3} = 1.
\end{aligned}$$

As a consequence, J is a subset of I since positive sets are maximal by definition. ◀

▶ **Lemma 9.** *Let A and B be two idempotent matrices of B_n satisfying $A \cdot B = B = B \cdot A$. Then every free pair of B is a free pair of A.*

Proof. Let us pick two idempotent matrices $A, B \in B_n$ satisfying $A \cdot B = B = B \cdot A$. We prove the lemma by contraposition: we show that for every pair of elements $i, j \in [1, n]$, $i \rightarrow_A j$ implies $i \rightarrow_B j$ (hence if i and j are incomparable by $\rightarrow_B$, so are they by $\rightarrow_A$).

Let us pick $i, j \in [1, n]$ satisfying $i \rightarrow_A j$, and $i_2, j_2 \in [1, n]$ satisfying $B_{i_2 i} = 1 = B_{j j_2}$. To conclude, we show that $B_{i_2 j_2} = 1$. To this end, we introduce two new elements $i_1, j_1 \in [1, n]$: First, as $(B \cdot A)_{i_2 i} = B_{i_2 i} = 1$, there exists $i_1 \in [1, n]$ such that $B_{i_2 i_1} = 1$ and $A_{i_1 i} = 1$; Second, as $(A \cdot B)_{j j_2} = B_{j j_2} = 1$, there exists $j_1 \in [1, n]$ such that $A_{j j_1} = 1$ and $B_{j_1 j_2} = 1$. Then, as $i \rightarrow_A j$ by supposition, we get that $A_{i_1 j_1} = 1$, which implies

$$B_{i_2 j_2} = (B \cdot A \cdot B)_{i_2 j_2} = 1, \text{ since } B_{i_2 i_1} = A_{i_1 j_1} = B_{j_1 j_2} = 1.$$

Since this holds for every $i_2, j_2 \in [1, n]$ satisfying $B_{i_2 i} = 1 = B_{j j_2}$, we obtain that $i \rightarrow_B j$. ◀

▶ **Lemma 10.** *Let A and B be two idempotent matrices of B_n satisfying $A \cdot B = B = B \cdot A$. If A and B have the same number of positive sets and free pairs, then they are equal.*

Proof. Let us pick two idempotent elements $A, B \in B_n$ such that $A \cdot B = B = B \cdot A$. Suppose that A and B have the same number of positive sets. By Lemma 8, each positive set of B contains at least one positive set of A. Since the positive sets of B are disjoint, the pigeonhole principle yields the two following claims.

▷ Claim 1. Each positive set of A is contained in a positive set of B.

▷ Claim 2. Each positive set of B contains exactly one positive set of A.

Moreover, suppose that A and B have the same number of free pairs. By Lemma 9 every free pair of B is a free pair of A. This yields the following claim.

▷ Claim 3. The free pairs of A and B are identical.

We now prove that $A = B$. First, we show that for every component A_{ik} equal to 1, the corresponding component B_{ik} is also equal to 1. Since A is stable, there exists $j \in [1, n]$ satisfying $A_{ij} = A_{jj} = A_{jk} = 1$. Then j is contained in a positive set of A, which is itself contained in a positive set of B by Claim 1. Therefore we obtain that $B_{jj} = 1$, which yields

$$B_{ik} = (A \cdot B \cdot A)_{ik} = 1, \text{ since } A_{ij} = B_{jj} = A_{jk} = 1.$$

To conclude, we show that for every component B_{ij} equal to 1, the corresponding component A_{ij} is also equal to 1. To this end, we introduce four new elements i_1, i_2, j_1, j_2 in $[1, n]$: First, as $(A \cdot B \cdot A)_{ij} = B_{ij} = 1$, there exist $i_2, j_2 \in [1, n]$ such that $A_{i i_2} = B_{i_2 j_2} = A_{j_2 j} = 1$. Second, as A is stable, there exist $i_1, j_1 \in [1, n]$ such that $A_{i i_1} = A_{i_1 i_1} = A_{i_1 i_2} = 1$ and $A_{j_2 j_1} = A_{j_1 j_1} = A_{j_1 j} = 1$. These definitions ensure that

$$B_{i_1 j_1} = (A \cdot B \cdot A)_{i_1 j_1} = 1, \text{ since } A_{i_1 i_2} = B_{i_2 j_2} = A_{j_2 j_1} = 1.$$

Note that, as observed after the definition of the relation induced by an idempotent matrix, this implies that $i_1 \rightarrow_B j_1$. We derive from this that either $i_1 \rightarrow_A j_1$ or $j_1 \rightarrow_A i_1$: if $i_1 = j_1$ this follows from the fact that $\rightarrow_A$ is reflexive, and if $i_1 \neq j_1$ this follows from Claim 3. We show that both possibilities lead to $A_{ij} = 1$.

- If $i_1 \rightarrow_A j_1$, then we obtain $A_{i_1 j_1} = 1$ as $A_{i_1 i_1} = 1 = A_{j_1 j_1}$. Therefore,

$$A_{ij} = (A \cdot A \cdot A)_{ij} = 1 \text{ since } A_{i i_1} = A_{i_1 j_1} = A_{j_1 j} = 1.$$

- If $j_1 \to_A i_1$, then we obtain $A_{j_1 i_1} = 1$ as $A_{j_1 j_1} = 1 = A_{i_1 i_1}$. Therefore,

  $$B_{j_1 i_1} = (A \cdot B \cdot A)_{j_1 i_1} = 1 \text{ since } A_{j_1 i_1} = B_{i_1 j_1} = A_{j_1 i_1} = 1.$$

  As a consequence, i_1 and j_1 are in the same positive set of B. Moreover, as $A_{i_1 i_1} = A_{j_1 j_1} = 1$, both i_1 and j_1 are elements of positive sets of A. Combining these two statements with Claim 2 yields that i_1 and j_1 are in the same positive set of A. Therefore $A_{i_1 j_1} = 1$, which implies that $i_1 \to_A j_1$, and we can conclude as in the previous point.

Since we successfully showed that every 1 of A corresponds to a 1 of B, and reciprocally, we obtain that $A = B$, which proves the statement. ◀

4.2 Lower bound

We construct a monomorphism φ between the max monoid $H_{f(n)}$, where $f(n) = \frac{n^2+n+2}{2}$, and the monoid of Boolean matrices B_n. The construction is split in two steps. First, we define φ over the domain $[1, g(n)+1]$, where $g(n) = \frac{n(n-1)}{2}$ is the number of pairs of elements $i < j$ in $[1, n]$. Then, we complete the definition over the domain $[g(n)+1, f(n)]$.

Diagonal to triangular. Let us define φ over $[1, g(n)+1]$. We map the neutral element $1 \in H_{f(n)}$ to the neutral element $D_n \in B_n$: the identity matrix. Then, we map $g(n)+1 \in H_{f(n)}$ to the full upper triangular matrix $U_n \in B_n$. Note that U_n contains $g(n)$ more 1's than D_n does. We define the images of the elements between 1 and $g(n)+1$ by gradually adding to D_n the 1's of U_n it lacks. Formally, we order the indices corresponding to the components above the diagonal $p_1 < p_2 < \ldots < p_{g(n)} \in [1, n] \times [1, n]$ according to the lexicographic order: (i, j) comes before (i', j') if either $i < i'$, or $i = i'$ and $j < j'$. Then, for every $m \in [1, g(n)+1]$, we construct the image $\varphi(m) \in B_n$ as follows:

- Every component $(\varphi(m))_{ii}$ of the diagonal is 1;
- Every component $(\varphi(m))_{ij}$ below the diagonal is 0;
- Every component $(\varphi(m))_{ij}$ above the diagonal is 1 if $(i, j) < p_m$, and 0 otherwise.

Triangular to empty. Let us define φ over $[g(n)+1, f(n)]$. To fit the first part of the definition, we map $g(n)+1 \in H_{f(n)}$ to the upper diagonal matrix $U_n \in B_n$. Then, we map the absorbing element $f(n) = g(n)+1+n \in H_{f(n)}$ to the absorbing element $0_n \in B_n$: the null matrix. Finally, for $m \in [0, n]$, we construct $\varphi(g(n)+1+m)$ by replacing the last m rows of U_n with 0's. Formally, we have:

- Every component $(\varphi(g(n)+1+m))_{ij}$ is 1 if $i \leq j$ and $i \leq n-m$, and 0 otherwise.

Proof of correctness. We prove that the function φ just defined is a monomorphism.

We show that φ is a homomorphism: $\varphi(m)\cdot\varphi(m') = \varphi(m') = \varphi(m')\cdot\varphi(m)$ for all $1 \leq m \leq m' \leq f(n)$. First, note that if $(\varphi(m'))_{ij} = 1$, then $(\varphi(m) \cdot \varphi(m'))_{ij} = (\varphi(m') \cdot \varphi(m))_{ij} = 1$: if $m \leq g(n)+1$, this follows from the fact that the diagonal of $\varphi(m)$ is filled with 1's, and if $m > g(n)+1$, since $m \leq m'$ we obtain that $(\varphi(m))_{ii} = (\varphi(m'))_{ij} = 1 = (\varphi(m'))_{ii} = (\varphi(m))_{ij}$. It remains to show that if $(\varphi(m) \cdot \varphi(m'))_{ik} = 1$ or $(\varphi(m') \cdot \varphi(m))_{ik} = 1$, then $(\varphi(m'))_{ik} = 1$. If $m' \leq g(n)+1$, this holds since for every triple $i \leq j \leq k \in [1, n]$, the pair (i, k) is lexicographically smaller than or equal to (j, k). If $m' > g(n)+1$, this holds since for every triple $i \leq j \leq k \in [1, n]$, trivially i is smaller than or equal to both i and j.

We conclude by showing that φ is injective: between $\varphi(1)$ and $\varphi(g(n)+1)$ a new 1 is added at each step, and between $\varphi(g(n)+1)$ and $\varphi(f(n))$ we remove at each step a 1 of the diagonal that was present in all the previous images.

▶ **Example 11.** We depict the monomorphism $\varphi : H_{f(n)} \to B_n$ in the case $n = 4$ by listing the $f(4) = 11$ elements of its image in B_4. Under each element, we state its number of positive sets followed by its number of free pairs.

$$\begin{array}{ccccccccccc}
\begin{matrix}\mathsf{I}&\mathsf{o}&\mathsf{o}&\mathsf{o}\\\mathsf{o}&\mathsf{I}&\mathsf{o}&\mathsf{o}\\\mathsf{o}&\mathsf{o}&\mathsf{I}&\mathsf{o}\\\mathsf{o}&\mathsf{o}&\mathsf{o}&\mathsf{I}\end{matrix} &
\begin{matrix}\mathsf{I}&\mathsf{I}&\mathsf{o}&\mathsf{o}\\\mathsf{o}&\mathsf{I}&\mathsf{o}&\mathsf{o}\\\mathsf{o}&\mathsf{o}&\mathsf{I}&\mathsf{o}\\\mathsf{o}&\mathsf{o}&\mathsf{o}&\mathsf{I}\end{matrix} &
\begin{matrix}\mathsf{I}&\mathsf{I}&\mathsf{I}&\mathsf{o}\\\mathsf{o}&\mathsf{I}&\mathsf{o}&\mathsf{o}\\\mathsf{o}&\mathsf{o}&\mathsf{I}&\mathsf{o}\\\mathsf{o}&\mathsf{o}&\mathsf{o}&\mathsf{I}\end{matrix} &
\begin{matrix}\mathsf{I}&\mathsf{I}&\mathsf{I}&\mathsf{I}\\\mathsf{o}&\mathsf{I}&\mathsf{o}&\mathsf{o}\\\mathsf{o}&\mathsf{o}&\mathsf{I}&\mathsf{o}\\\mathsf{o}&\mathsf{o}&\mathsf{o}&\mathsf{I}\end{matrix} &
\begin{matrix}\mathsf{I}&\mathsf{I}&\mathsf{I}&\mathsf{I}\\\mathsf{o}&\mathsf{I}&\mathsf{I}&\mathsf{o}\\\mathsf{o}&\mathsf{o}&\mathsf{I}&\mathsf{o}\\\mathsf{o}&\mathsf{o}&\mathsf{o}&\mathsf{I}\end{matrix} &
\begin{matrix}\mathsf{I}&\mathsf{I}&\mathsf{I}&\mathsf{I}\\\mathsf{o}&\mathsf{I}&\mathsf{I}&\mathsf{I}\\\mathsf{o}&\mathsf{o}&\mathsf{I}&\mathsf{o}\\\mathsf{o}&\mathsf{o}&\mathsf{o}&\mathsf{I}\end{matrix} &
\begin{matrix}\mathsf{I}&\mathsf{I}&\mathsf{I}&\mathsf{I}\\\mathsf{o}&\mathsf{I}&\mathsf{I}&\mathsf{I}\\\mathsf{o}&\mathsf{o}&\mathsf{I}&\mathsf{I}\\\mathsf{o}&\mathsf{o}&\mathsf{o}&\mathsf{I}\end{matrix} &
\begin{matrix}\mathsf{I}&\mathsf{I}&\mathsf{I}&\mathsf{I}\\\mathsf{o}&\mathsf{I}&\mathsf{I}&\mathsf{I}\\\mathsf{o}&\mathsf{o}&\mathsf{I}&\mathsf{I}\\\mathsf{o}&\mathsf{o}&\mathsf{o}&\mathsf{o}\end{matrix} &
\begin{matrix}\mathsf{I}&\mathsf{I}&\mathsf{I}&\mathsf{I}\\\mathsf{o}&\mathsf{I}&\mathsf{I}&\mathsf{I}\\\mathsf{o}&\mathsf{o}&\mathsf{o}&\mathsf{o}\\\mathsf{o}&\mathsf{o}&\mathsf{o}&\mathsf{o}\end{matrix} &
\begin{matrix}\mathsf{I}&\mathsf{I}&\mathsf{I}&\mathsf{I}\\\mathsf{o}&\mathsf{o}&\mathsf{o}&\mathsf{o}\\\mathsf{o}&\mathsf{o}&\mathsf{o}&\mathsf{o}\\\mathsf{o}&\mathsf{o}&\mathsf{o}&\mathsf{o}\end{matrix} &
\begin{matrix}\mathsf{o}&\mathsf{o}&\mathsf{o}&\mathsf{o}\\\mathsf{o}&\mathsf{o}&\mathsf{o}&\mathsf{o}\\\mathsf{o}&\mathsf{o}&\mathsf{o}&\mathsf{o}\\\mathsf{o}&\mathsf{o}&\mathsf{o}&\mathsf{o}\end{matrix} \\
(4,6) & (4,5) & (4,4) & (4,3) & (4,2) & (4,1) & (4,0) & (3,0) & (2,0) & (1,0) & (0,0)
\end{array}$$

Starting with the identity matrix D_4, we gradually add 1's, reaching the triangular matrix U_4 in $g(4) = 6$ steps. Then, we erase line after line, reaching the null matrix 0_4 in 4 steps.

References

1 Félix Baschenis, Olivier Gauwin, Anca Muscholl, and Gabriele Puppis. One-way definability of two-way word transducers. *Logical Methods in Computer Science*, 14, 2018. `doi:10.23638/LMCS-14(4:22)2018`.

2 Michael Breen. A maximal chain of principal ideals in the semigroup of binary relations on a finite set. In *Semigroup Forum*, volume 43, pages 63–76. Springer, 1991.

3 Olexandr Ganyushkin and Volodymyr Mazorchuk. *Classical finite transformation semigroups: an introduction*, volume 9. Springer Science & Business Media, 2008.

4 T. E. Hall and Mark V. Sapir. Idempotents, regular elements and sequences from finite semigroups. *Discrete Mathematics*, 161:151–160, 1996. `doi:10.1016/0012-365X(95)00223-J`.

5 Shaofang Hong. Distribution of cardinalities of row spaces of boolean matrices of order n. *Southeast Asian Bulletin of Mathematics*, 24:51–64, 2000.

6 Janusz Konieczny. On cardinalities of row spaces of boolean matrices. In *Semigroup Forum*, volume 44, pages 393–402. Springer, 1992.

7 Wen Li and Mou-Cheng Zhang. On konieczny's conjecture of boolean matrices. In *Semigroup Forum*, volume 50, pages 37–58. Springer, 1995.

8 Filip Mazowiecki and Cristian Riveros. Pumping lemmas for weighted automata. In Rolf Niedermeier and Brigitte Vallée, editors, *35th Symposium on Theoretical Aspects of Computer Science, STACS 2018*, volume 96 of *LIPIcs*, pages 50:1–50:14. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018. `doi:10.4230/LIPIcs.STACS.2018.50`.

9 A. Mukherjea and R. Chaudhuri. Idempotent boolean matrices. *Semigroup forum*, 21:273–282, 1980. URL: `http://eudml.org/doc/134452`.

10 Anca Muscholl and Gabriele Puppis. The many facets of string transducers (invited talk). In Rolf Niedermeier and Christophe Paul, editors, *36th International Symposium on Theoretical Aspects of Computer Science, STACS 2019*, volume 126 of *LIPIcs*, pages 2:1–2:21. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019. `doi:10.4230/LIPIcs.STACS.2019.2`.

11 Jean-Éric Pin. Mathematical foundations of automata theory. *Lecture notes LIAFA, Université Paris*, 7, 2010.

12 Michael O. Rabin and Dana S. Scott. Finite automata and their decision problems. *IBM Journal of Research and Development*, 3:114–125, 1959. `doi:10.1147/rd.32.0114`.

13 F. P. Ramsey. On a problem of formal logic. *Proceedings of the London Mathematical Society*, s2-30(1):264–286, 1930. `doi:10.1112/plms/s2-30.1.264`.

14 Marcel-Paul Schützenberger. Une théorie algébrique du codage. *Séminaire Dubreil. Algebre et théorie des nombres*, 9:1–24, 1955.

15 Marcel Paul Schützenberger. On finite monoids having only trivial subgroups. *Information and Control*, 8:190–194, 1965. `doi:10.1016/S0019-9958(65)90108-7`.

16 Imre Simon. Factorization forests of finite height. *Theor. Comput. Sci.*, 72:65–94, 1990. `doi:10.1016/0304-3975(90)90047-L`.

17 M-C Zhang, S-F Hong, and H-B Kan. On the cardinalities of the row spaces of non-full rank boolean matrices. In *Semigroup Forum*, volume 59, pages 152–154. Springer, 1999.