# A Note on Factorization Forests of Finite Height

Jérémie Chalopin
Département d'Informatique
École Normale Supérieure de Lyon
69364 Lyon, France

Hing Leung*
Department of Computer Science
New Mexico State University
Las Cruces, NM 88003, U.S.A.

19th May 2003

**Abstract**

Simon [3] has proved that every morphism from a free semigroup to a finite semigroup $S$ admits a Ramseyan factorization forest of height at most $9|S|$. In this paper, we prove the same result of Simon with an improved bound of $7|S|$. We provide a simple algorithm for constructing a factorization forest. In addition, we show that the algorithm cannot be improved significantly. We give examples of semigroup morphism such that any Ramseyan factorization forest for the morphism would require a height not less than $|S|$.

## 1 Introduction

Factorization forests are introduced by Imre Simon [3] to describe factorizations of words over a given alphabet. Simon has proved that every morphism from a free semigroup to a finite semigroup $S$ admits a Ramseyan factorization forest of height at most $9|S|$. Later, Simon [4] gave a short proof of a weaker version of the result.

The result can be used to prove Brown's lemma [1] on locally finite semigroups in a constructive way. Simon [5] has also used this result to solve the limitedness problem on distance automata.

In this paper, we prove the same result of Simon with an improved bound of $7|S|$. We provide a simple algorithm for constructing a factorization forest. Our algorithm is a simplication of Simon's.

In Section 2, we first give a presentation of the problem. Then we prove the result in the next three sections. In Section 6, we show that the algorithm cannot be improved significantly. We give examples of semigroup morphism such that any Ramseyan factorization forest for the morphism would require a height not less than $|S|$.

*The research is supported in part by NSF MII grants EIA-0220590 and EIA-9810732.

## 2 Presentation of the problem

Given a set $A$, we write $A^+$ or $\mathcal{F}(A)$ to denote the free semigroup generated by $A$. A factorization forest $F$ over an alphabet $A$ can be defined by a function $d$ from $A^+$ into $\mathcal{F}(A^+)$ such that for every $x \in A^+$, $d(x) = (x_1, x_2, \ldots, x_p)$ implies that $x = x_1 x_2 \cdots x_p$. We say that $d(x)$ is a factorization of $x$.

With each word $x \in A^+$, we associate a rooted tree $T(x)$ such that the tree nodes are labelled by words in $A^+$. If $|d(x)| = 1$, $T(x)$ consists just of a root labelled $x$. If $d(x) = (x_1, x_2, \ldots, x_p)$ where $p \geq 2$, the root of $T(x)$ is labelled by $x$ and has $p$ children $T(x_i)$ for $1 \leq i \leq p$.

Given $x \in A^+$, We define the height $h(x)$ to be the height of the tree $T(x)$. Specifically, $h(x) = 0$ if $|d(x)| = 1$, and $h(x) = 1 + \max\{h(x_i) \mid 1 \leq i \leq p\}$ where $d(x) = (x_1, x_2, \ldots, x_p)$ and $p \geq 2$. The height of a factorization forest $F$ is defined by $\sup\{h(x) \mid x \in A^+\}$.

Let $f$ be a morphism from a free semigroup $A^+$ to a finite semigroup $S$. A factorization forest $F$ is Ramseyan modulo $f$ if for every $x$ of degree $p \geq 3$, $d(x) = (x_1, x_2, \ldots, x_p)$ implies that there exists an idempotent $e$ such that $e = f(x) = f(x_1) = f(x_2) = \cdots = f(x_p)$. We say that $f$ admits a Ramseyan factorization forest if it admits a factorization forest $F$ over $A$ which is Ramseyan modulo $f$ and the only words such that $d(x) = x$ are the elements of $A$.

In this paper, we assume that the readers are familiar with the local structure theory of semigroup, which are covered in Chapter 2 (Green's relations) of Lallement [2].

## 3 The group case

We consider in this section a morphism $f$ from a free semigroup $A^+$ to a finite group $G$. Let $e$ be the identity of $G$. Since $e$ is the only idempotent element of $G$, the only nodes in a factorization tree with an outdegree greater than two will have a label $x$ such that $f(x) = e$.

**Theorem 1** *Every morphism $f : A^+ \to G$, where $G$ is a finite group, admits a Ramseyan factorization forest of height at most $3|G|$.*

**Proof.** Given a word $x$, let $\text{PrefixImages}(x) = \{f(u) \mid u \in A^+ \text{ is a proper prefix of } x\}$. For all $x, v \in A^+$ and $u, w \in A^*$ such that $x = uvw$, we have $f(u)\,\text{PrefixImages}(v) \subseteq \text{PrefixImages}(x)$.

We will show by induction on $|\text{PrefixImages}(x)|$ that we can find a tree for $x$ of height at most $3|\text{PrefixImages}(x)|$.

If $|\text{PrefixImages}(x)| = 0$, we have $|x| = 1$. Let $d(x) = x$ which gives a factorization tree of height 0.

Suppose $|\text{PrefixImages}(x)| \geq 1$. Let $b \in \text{PrefixImages}(x)$. Let $x = a_1 a_2 \cdots a_p$ where $a_i \in A$ for $1 \leq i \leq p$. Consider all prefixes of $x$ that are mapped to $b$ under $f$. Let $1 \leq i_1 < i_2 < \cdots < i_k \leq p$ be *all* the indices such that $f(a_1 \cdots a_{i_j}) = b$. Let $u = a_1 \cdots a_{i_1}$, $v = a_{i_k+1} \cdots a_p$ and $y_j = a_{i_j+1} \cdots a_{i_{j+1}}$ for each $1 \leq j \leq k-1$. We also denote $y_1 \cdots y_{k-1}$ by $y$. Since $G$ is a group, we have $e = f(y_1) = f(y_2) = \cdots = f(y_{k-1}) = f(y)$. We construct a factorization tree (see Figure 1) for $x$ by defining $d(x) = (uy, v)$, $d(uy) = (u, y)$ and $d(y) = (y_1, \ldots, y_{k-1})$. A degenerate case occurs when $k = 1$, $y = \varepsilon$ and $d(x) = (u, v)$.

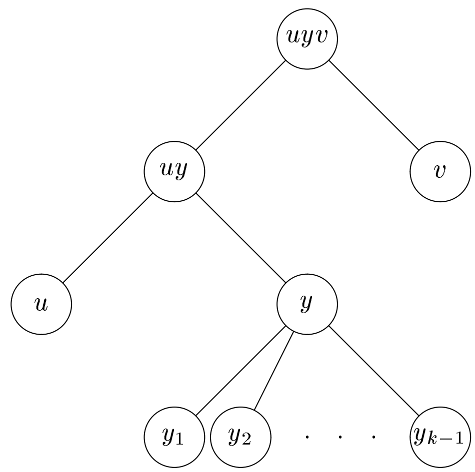


Figure 1: How to build a tree for the group case

We will now show that for each leaf of the tree we have just built, the size of PrefixImages has decreased. We know that $\text{PrefixImages}(u) \subseteq \text{PrefixImages}(x)$ and $b \in \text{PrefixImages}(x) \setminus \text{PrefixImages}(u)$. Thus, $|\text{PrefixImages}(u)| < |\text{PrefixImages}(x)|$. We also know that for each $y_j$, $b\,\text{PrefixImages}(y_j) = f(uy_1 \cdots y_{j-1})\,\text{PrefixImages}(y_i) \subseteq \text{PrefixImages}(x)$. Since $G$ is a group and $b \in \text{PrefixImages}(x) \setminus b\,\text{PrefixImages}(y_j)$, we conclude that $|\text{PrefixImages}(y_j)| < |\text{PrefixImages}(x)|$. With the same argument, we can see that $|\text{PrefixImages}(v)| < |\text{PrefixImages}(x)|$. By the induction hypothesis, there are factorization trees for $u$, $v$ and each $y_i$ of height at most $3(|\text{PrefixImages}(x)|-1)$ and consequently we can construct a tree for $x$ of height at most $3|\text{PrefixImages}(x)|$.

Since $|\text{PrefixImages}(x)| \leq |G|$, we have found a way to build a factorization forest of height at most $3|G|$. □

# 4 The single $\mathcal{D}$-class case

We consider in this section a morphism $f$ from $A^+$ to a finite semigroup that has only one single $\mathcal{D}$-class. Note that the single $\mathcal{D}$-class must be regular.

**Theorem 2** *Every morphism $f : A^+ \to S$, where $S$ is a finite semigroup which has only one single $\mathcal{D}$-class, admits a Ramseyan factorization forest of height at most $5|S|$.*

**Proof.** Let $x = a_1 a_2 \cdots a_p$ where $a_i \in A$ for $1 \leq i \leq p$. We define $\text{int}(x) = \{(L_{f(a_i)}, R_{f(a_{i+1})}) \mid 1 \leq i \leq p-1\}$. Observe that $\text{int}(v) \subseteq \text{int}(x)$ where $v$ is a factor of $x$.

Recall that all $\mathcal{H}$-classes in a $\mathcal{D}$-class are of the same size. Let $q$ be the size of each $\mathcal{H}$-class. We will show by induction on $|\text{int}(x)|$ that we can find for each

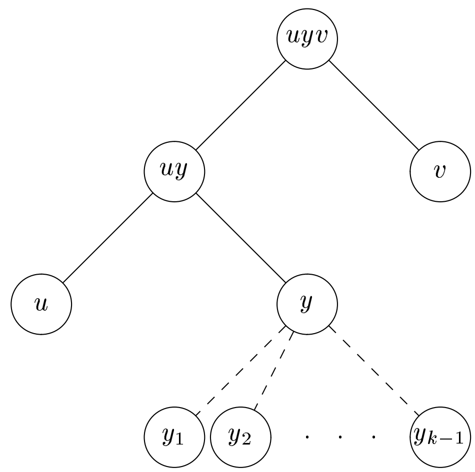


Figure 2: How to build a tree for the single $\mathcal{D}$-class case

word $x$ a factorization tree of height at most $5q|\text{int}(x)|$.

If $\text{int}(x) = \emptyset$, we have $|x| = 1$. Let $d(x) = x$, which gives a factorization tree of height 0.

Suppose $|\text{int}(x)| \geq 1$. Let $(L, R) \in \text{int}(x)$. Let $1 \leq i_1 < i_2 < \cdots < i_k \leq p-1$ be *all* the indices such that $(L_{f(a_{i_j})}, R_{f(a_{i_j+1})}) = (L, R)$. Let $u = a_1 \cdots a_{i_1}$, $v = a_{i_k+1} \cdots a_p$ and $y_j = a_{i_j+1} \cdots a_{i_{j+1}}$ for each $1 \leq j \leq k-1$. We also denote $y_1 \cdots y_{k-1}$ by $y$. By the local structure theory of semigroup, we know that $y$ and each $y_j$ belong to the same $\mathcal{H}$-class $H = R \cap L$, which is a subgroup of $S$. We construct a factorization tree (see Figure 2) for $x$ by defining $d(x) = (uy, v)$ and $d(uy) = (u, y)$. By using the technique from the previous section for the group case, we construct a factorization tree of height at most $3|H| = 3q$ with root $y$ and leaves $y_1, \ldots, y_{k-1}$. It is a degenerate case when $k = 1$ and $y = \epsilon$. In that case, the construction tree for $x$ is defined by $d(x) = (u, v)$.

Since $u$, $v$, and all the $y_j$'s are factors of $x$, we know that $\text{int}(u) \subseteq \text{int}(x)$, $\text{int}(v) \subseteq \text{int}(x)$ and $\text{int}(y_j) \subseteq \text{int}(x)$, for $1 \leq j \leq k-1$. Moreover, by the way $u$, $v$ and the $y_j$'s are defined, we know that $(L, R)$ belongs to $\text{int}(x)$ but not to $\text{int}(u)$, $\text{int}(v)$ or any of the $\text{int}(y_j)$'s. Then, by the induction hypothesis, there are factorization trees for $u$, $v$ and each $y_j$'s of height at most $5q(|\text{int}(x)| - 1)$ and consequently, we can construct a factorization tree for $x$ of height at most $2 + 3q + 5q(|\text{int}(x)| - 1) \leq 5q|\text{int}(x)|$.

Since $|\text{int}(x)|$ is less than or equal to the number of different $\mathcal{H}$-classes in $S$ and $q$ is the size of any $\mathcal{H}$-class, the height of the factorization tree for $x$ is at most $5|S|$. $\square$

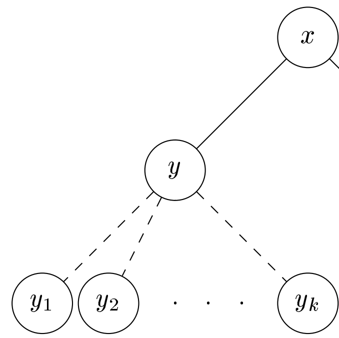


Figure 3: How to build a tree for the general case

## 5 The general case

We consider in this section a morphism $f$ from $A^+$ to a finite semigroup. The following basic knowledge about semigroup is very important in the understanding of the algorithm: if $x = uvw$ then $D_{f(v)} \geq D_{f(x)}$.

**Theorem 3** *Every morphism $f : A^+ \to S$, from a free semigroup to a finite one, admits a Ramseyan factorization forest of height at most $7|S|$.*

**Proof.** Given a word $x$, we consider the position of $D_{f(x)}$ in the partial ordering of the $\mathcal{D}$-classes. We will show by an induction on the $\mathcal{D}$-classes partial ordering that we can construct a factorization tree of height at most $7 \sum_{D \geq_{\mathcal{D}} D_{f(x)}} |D|$.

Firstly suppose that $D_{f(x)}$ is one of the maximal $\mathcal{D}$-class for the partial ordering $\leq_{\mathcal{D}}$. We can apply the technique from the previous section for a single $\mathcal{D}$-class. Thus, there exists a factorization tree for $x$ of height at most $5|D_{f(x)}|$, which is less than $7|D_{f(x)}| = 7 \sum_{D \geq_{\mathcal{D}} D_{f(x)}} |D|$.

Now, for the inductive case, we suppose that $D_{f(x)}$ is not a maximal $\mathcal{D}$-class. We need another induction on the length of $x$. If $|x| = 1$, we put $d(x) = x$ which gives a factorization tree of height 0.

Suppose $|x| \geq 1$. We say that a word $w \in A^+$ is primitive if $w \in A$, or $w$ and $y$ belong to different $\mathcal{D}$-classes where $w = ya$ and $a \in A$. We decompose $x$ into primitive strings $y_1, y_2, \ldots, y_k$ such that $x = y_1 y_2 \cdots y_k v$ where $D_{f(y_j)} = D_{f(x)}$ for $1 \leq j \leq k$ and $D_{f(v)} >_{\mathcal{D}} D_{f(x)}$. Clearly, $k \geq 1$. A degenerate case of the decomposition is when $v$ does not exist. That is, $x$ is decomposed into primitive strings $y_1, y_2, \ldots, y_k$ such that $x = y_1 y_2 \cdots y_k$ where $D_{f(y_j)} = D_{f(x)}$ for $1 \leq j \leq k$. Note that the decomposition of $x$ is unique.

Consider the general case of the decomposition of $x$. (The degenerate case is easier.) We factorize $x$ (see Figure 3) according to the decomposition such that $d(x) = (y, v)$ and $y$ is factorized into $y_1, y_2, \ldots, y_k$ using the same technique given in the previous section for a single $\mathcal{D}$-class. The height of the subtree rooted at $y$ with leaves $y_1, y_2, \ldots, y_k$ (see Figure 3) is at most $5|D_{f(x)}|$.

For each primitive string $y_j$ where $1 \leq j \leq k$, we factorize it (see Figure 4) into $y'_j$ and $a$ where $y_j = y'_j a$ and $a \in A$. That is, $d(y_j) = (y'_j, a)$. There is no

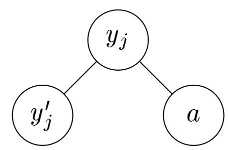

Figure 4: How to build a tree for a primitive string

need to factorize $y_j$ if $y_j \in A$.

By the induction hypothesis, there exist factorization trees for each $y'_j$ and $v$ of height at most $7\sum_{D>_{\mathcal{D}}D_{f(x)}}|D|$. The total height for the factorization of $x$ is at most $2+5|D_{f(x)}|+7\sum_{D>_{\mathcal{D}}D_{f(x)}}|D| \leq 7\sum_{D\geq_{\mathcal{D}}D_{f(x)}}|D|$.

Consequently, for each word $x$ we can construct a factorization tree of height at most $7\sum_{D\geq_{\mathcal{D}}D_{f(x)}}|D| \leq 7|S|$. □

# 6 Lower Bounds

We have shown that for each morphism $f$ from a free semigroup $A^+$ to a finite semigroup $S$, there exists a Ramseyan factorization forest of height at most $7|S|$ which is linear in $|S|$. In this section, we prove that the result cannot be significantly improved. For each of the three steps of the algorithm, we show that there are examples of semigroup morphism such that any Ramseyan factorization forest for the morphism would require a height not less than $|S|$.

## 6.1 The group case

Let $f$ be a morphism from $A^+$ to a finite semigroup. We define a new kind of Ramseyan factorization forest (which we call *Ramseyan group factorization*) as a function $d$ from $A^+$ to $\mathcal{F}(A^+)$ such that $d(x)=(x_1,x_2,\ldots,x_k)$ implies $f(x_i)=e$ for $2\leq i\leq k-1$ where $e$ is an idempotent. As in the former definition, the only words $x$ such that $d(x)=x$ must be the elements of $A$.

One can see that the usual Ramseyan factorization tree is just a special case of the Ramseyan group factorization tree. On the other hand, given a Ramseyan group factorization tree of height $h$, one can easily convert it to a usual Ramseyan factorization tree of height at most $3h$. Thus, the two variants of Ramseyan factorization forests are linearly related in height.

Given $x\in A^+$, we define $h(x)$ to be the minimum height of a Ramseyan group factorization tree for $x$.

**Lemma 1** *For all $x\in A^+$, for all $u,v\in A^*$, $h(x)\leq h(uxv)$.*

**Proof.** Given a group factorization tree for $uxv$, we can prune the tree to obtain a group factorization tree for $x$. The pruning is done by first eliminating symbols from $u$ and $v$ at the leaves level. More pruning may be triggered at the higher levels in order to maintain the properties required for a group factorization tree. The resulting tree has height not exceeding that of the given tree for $uxv$. □

Let $G$ be a group $\{\alpha_1, \ldots, \alpha_n\}$ of size $n$, where $\alpha_1 = e$ is the identity. Let $A = \{a_1, \ldots, a_n\}$ and $f$ be a morphism from $A^+$ to $G$ such that $f(a_i) = \alpha_i$ for $1 \leq i \leq n$. We want to show that there exists $z_n \in A^+$ such that the height of any group factorization tree for $z_n$ is at least $n$.

Let $x = b_1 b_2 \cdots b_k$ where $b_i \in A$ for $1 \leq i \leq k$. Consider the sequence of prefix images $\beta_i = f(b_1 b_2 \cdots b_i)$, for $1 \leq i \leq k$. We define $\beta = \beta_1 \cdots \beta_k$ to be the *prefix images string* of $x$. Suppose we are given with the prefix images string $\beta$ but not the original string $x$. Since $G$ is a group, we can compute $x$ as follows: $b_1 = f^{-1}(\beta_1)$ and $b_i = f^{-1}(\beta_{i-1}^{-1}\beta_i)$ for $2 \leq i \leq k$.

Instead of constructing $z_n$ such that the height of any Ramseyan group factorization tree for $z_n$ is at least $n$, we will construct the prefix images string $\delta_n$ of $z_n$ from which $z_n$ can be recovered uniquely.

Let $\beta$ and $\gamma$ be two prefix images strings of the same length $k$. We say that $\beta$ and $\gamma$ are *structurally similar* if for all $1 \leq i, j \leq k$, $\beta_i = \beta_j$ iff $\gamma_i = \gamma_j$.

**Lemma 2** *Let $x, y \in A^+$ such that the prefix images strings of $x$ and $y$ are structurally similar. Then $h(x) = h(y)$.*

**Proof.** Let $k = |x| = |y|$. Let $x = b_1 \cdots b_k$. Suppose we are given a group factorization tree for $x$. Let $d(x) = (x_1, x_2, \ldots, x_p)$ where $p \geq 2$. Then $f(x_1) = f(x_1 x_2) = f(x_1 x_2 x_3) = \cdots = f(x_1 x_2 \cdots x_{p-1})$. Let $d(y) = (y_1, y_2, \ldots, y_p)$ such that $|y_i| = |x_i|$ for $1 \leq i \leq p$. Since the prefix images strings of $x$ and $y$ are structurally similar, we have $f(y_1) = f(y_1 y_2) = f(y_1 y_2 y_3) = \cdots = f(y_1 y_2 \cdots y_{p-1})$. Thus, $d(y)$ is well-defined. Furthermore, $x_i$ and $y_i$ are again structurally similar for $1 \leq i \leq p$. Therefore, given a group factorization tree for $x$, we can construct a similar group factorization tree for $y$ that is structurally the same, and vice versa. □

To construct $\delta_n$, we construct inductively $\delta_i \in \{\alpha_1, \ldots, \alpha_i\}^+$ for $i = 1, 2, \ldots, n$ such that the height of any group factorization tree for $z_i \in A^+$, which prefix images string is $\delta_i$, is at least $i$. We define $\delta_1 = \alpha_1 \alpha_1$. It is immediate to see that any group factorization tree for $z_1$ has height one. Suppose we have defined $\delta_i \in \{\alpha_1, \ldots, \alpha_i\}^+$. Let $1 \leq j \leq i$. By substituting every occurrence of $\alpha_j$ by $\alpha_{i+1}$ in $\delta_i$, we obtain a string $\delta_i^j \in (\{\alpha_1, \ldots, \alpha_i\} \setminus \{\alpha_j\} \cup \{\alpha_{i+1}\})^+$ which is structurally similar to $\delta_i$. Next, we define $\delta_{i+1} = \delta_i \delta_i^1 \delta_i^2 \cdots \delta_i^i \in \{\alpha_1, \ldots, \alpha_{i+1}\}^+$. We want to show that any group factorization tree for $z_{i+1}$ has height at least $i+1$. Let $d(z_{i+1}) = (x_1, x_2, \ldots, x_p)$ where $p \geq 2$.

Firstly suppose $\alpha_{i+1} = f(x_1) = f(x_1 x_2) = \cdots = f(x_1 x_2 \cdots x_{p-1})$. Since none of the prefix images of $z_i$ is $\alpha_{i+1}$, we deduce that $\delta_i$ is a prefix of the prefix images string for $x_1$ where $f(x_1) = \alpha_{i+1}$. Thus, $z_i$ is a prefix of $x_1$. Since $z_i$ requires a tree of height at least $i$ for factorization, by Lemma 1 any factorization tree for $x_1$ also requires a height of at least $i$. Hence, the total height of this tree for $z_{i+1}$ is at least $i+1$.

Next suppose $\alpha_j = f(x_1) = f(x_1 x_2) = \cdots = f(x_1 x_2 \cdots x_{p-1})$ where $1 \leq j \leq i$. Recall that $\delta_i^j \in (\{\alpha_1, \ldots, \alpha_{i+1}\} \setminus \{\alpha_j\})^+$ is a string that does not consist of the symbol $\alpha_j$. Let $k = |\delta_i|$. Then $|\delta_{i+1}| = k(i+1)$. Since $\delta_{i+1} = \delta_i \delta_i^1 \delta_i^2 \cdots \delta_i^i$, the substring $\delta_i^j$ is located within $\delta_{i+1}$ in positions between $kj+1$ and $k(j+1)$. There must exist $x_m$ where $1 \leq m \leq p$ such that $\delta_i^j$ is 'covered' by $x_m$ in the sense that $|x_1 x_2 \cdots x_{m-1}| < kj+1 < k(j+1) \leq |x_1 x_2 \cdots x_m|$. Since $G$ is a group, there is a substring of the prefix images string of $x_m$ that is structurally similar to $\delta_i^j$, which is also structurally similar to $\delta_i$ since the 'structurally similarity' relation is transitive. By Lemma 1 and Lemma 2, we conclude that $h(x_m) \geq i$. Again, the total height of this alternate group factorization tree is at least $i+1$.

In Section 3, we prove that every morphism from $A^+$ to a group $G$ admits a Ramseyan factorization forest of height at most $3|G|$. A close look at the algorithm shows that one can obtain a Ramseyan group factorization forest of height at most $|G|$. Thus, the existence of $z_n$ shows that the algorithm is indeed tight.

**Theorem 4** *Consider any morphism $f : A^+ \to G$, where $G$ is a finite group. Let $F$ be a Ramseyan factorization forest for $f$. Then the height of $F$ is at least $|G|$.*

**Proof.** It has been shown that there exists a string $z$ such that the height of any Ramseyan group factorization tree is at least $|G|$. Since Ramseyan factorization is a special case of Ramseyan group factorization, we conclude that the height of any Ramseyan factorization tree is also at least $|G|$. □

## 6.2 The case of rectangular bands

In Section 4, we describe a factorization algorithm for the case of a single $\mathcal{D}$-class. The algorithm is done recursively and relies on the results of Section 3 for factorization in the group case. If every subgroup (equivalently, $\mathcal{H}$-class) is trivial, the algorithm given in Section 4 will produce a factorization forest of height at most $3|S|$. In the following, we give a family of examples of single $\mathcal{D}$-class semigroups with trivial $\mathcal{H}$-classes (which are called *rectangular bands*) that require factorization forests of height at least $|S|$.

Given two nonempty sets $I$ and $J$, we define an associative multiplication on the set $I \times J$ as follows:

$$\forall i, i' \in I, \forall j, j' \in J, (i, j)(i', j') = (i, j')$$

A semigroup $S$ is called a rectangular band if there exist $I$ and $J$ such that $S$ is isomorphic to $I \times J$ with the previous multiplication.

Let $f$ be a morphism from $A^+$ to a rectangular band. We define a Ramseyan rectangular factorization forest as a function $d$ from $A^+$ to $\mathcal{F}(A^+)$ such that $d(x) = (x_1, x_2, \ldots, x_q)$ implies $|\text{int}(d(x))| = 1$ where $\text{int}(d(x)) = \text{int}(x_1, x_2, \ldots, x_q) = \{(L_{f(x_i)}, R_{f(x_{i+1})}) \mid 1 \leq i \leq q-1\}$. As in the former definitions, the only words $x$ such that $d(x) = x$ must be the elements of $A$. Observe that every element of a rectangular band is an idempotent and $f(x_2) = f(x_3) = \cdots = f(x_{q-1})$. Thus, Ramseyan factorization forest is a special case of Ramseyan rectangular factorization forest, which in turn is a special case of Ramseyan group factorization forest. Given a Ramseyan rectangular factorization forest of height $h$, we can construct a Ramseyan factorization forest of height at most $3h$. Therefore, the three variants of Ramseyan factorization forests are linearly related in height.

Given $x \in A^+$, we define $h(x)$ to be the minimum height of a Ramseyan rectangular factorization tree for $x$.

**Lemma 3** *For all $x \in A^+$, for all $u, v \in A^*, h(x) \leq h(uxv)$.*

**Proof.** The proof is exactly the same as in the former case for Ramseyan group factorization forest. □

Let $A = \{a_{ij} \mid i \in I, j \in J\}$ and $f$ be the morphism from $A^+$ to the rectangular band $I \times J$ such that for each $(i,j) \in I \times J$, $f(a_{ij}) = (i,j)$. Given a word $x = a_{i_1 j_1} \cdots a_{i_p j_p}$, we define $\mathrm{int}(x) = \{\mathrm{betw}_x(k) \mid 1 \le k \le p-1\}$ where $\mathrm{betw}_x(k) = (j_k, i_{k+1})$. We want to show that there exists a word such that the height of any rectangular factorization tree for this word will be at least the size of the semigroup $|S|$ which is $|I| \times |J|$.

As for the group case, we will construct recursively a word $x_k$ such that $|\mathrm{int}(x_k)| = k$ and $h(x_k) \ge k$. For $k = 1$, pick any specific ordered pair $(i', j')$ from $S$. Let $x_1 = a_{i'j'}a_{i'j'}$ which will need a tree of height 1 and $|\mathrm{int}(x_1)| = |\{(j', i')\}| = 1$.

Suppose $k > 1$. We will construct $x_k$ from the structure of $x_{k-1}$. Pick a specific ordered pair $(j'', i'') \in S \setminus \mathrm{int}(x_{k-1})$. For each $(j,i) \in \mathrm{int}(x_{k-1})$, we construct a word $y_{ji}$ by replacing the letters of $x_{k-1}$ such that $f(y_{ji}) = (i', j')$ and for each $1 \le p \le |x_{k-1}| - 1$, $\mathrm{betw}_{y_{ji}}(p) = (j'', i'')$ if $\mathrm{betw}_{x_{k-1}}(p) = (j, i)$ and $\mathrm{betw}_{y_{ji}}(p) = \mathrm{betw}_{x_{k-1}}(p)$ otherwise. It is easy to see that $\mathrm{int}(y_{ji}) = \{(j'', i'')\} \cup \mathrm{int}(x_{k-1}) \setminus \{(j,i)\}$ and the minimal height of a Ramseyan rectangular factorization tree for each $y_{ij}$ is at least $k-1$. We define $x_k$ to be the concatenation of $x_{k-1}$ and all the $y_{ij}$'s that have been defined previously. Note that $f(x_k) = (i', j')$ and $\mathrm{int}(x_k) = \mathrm{int}(x_{k-1}) \cup \{(j'', i'')\}$. Similar to the proof for the group case, we can show that the minimal height of a Ramseyan rectangular factorization tree for $x_k$ is at least $k$. Thus, there exists a word which Ramseyan rectangular factorization tree requires a height at least $|S|$. Remark: A close look at the algorithm given in Section 4 shows that every morphism from $A^+$ to a rectangular band semigroup $S$ admits a Ramseyan rectangular factorization forest of height $|S|$. Next, since Ramseyan factorization forest is a special case of Ramseyan rectangular factorization forest, we obtain the next theorem.

**Theorem 5** *Consider any morphism $f : A^+ \to S$, where $S$ is a rectangular band. Let $F$ be a Ramseyan factorization forest for $f$. Then the height of $F$ is at least $|S|$.*

### 6.3 $\mathcal{D}$-trivial semigroups

In Section 5, we describe a factorization algorithm for the general case. The algorithm is done recursively and relies on the results of Section 4 for factorization for the single $\mathcal{D}$-class case. If every $\mathcal{D}$-class is trivial, the algorithm given in Section 5 will produce a factorization forest of height at most $3|S|$. In the following, we give a family of examples of $\mathcal{D}$-trivial semigroups that require factorization forests of height more than $|S|$.

For each $n$, let $S_n$ be the semigroup $\{\alpha_1, \ldots, \alpha_n\}$ with the following associative operation: $\alpha_i \alpha_j = \alpha_j \alpha_i = \alpha_{\max\{i,j\}}$. We can easily see that $D_{\alpha_i} = \{\alpha_i\}$ and if $i < j$, $D_j <_{\mathcal{D}} D_i$. Let $A_n = \{a_1, \ldots, a_n\}$. Let $f : A_n^+ \to S_n$ be the morphism such that for each $1 \le i \le n$, $f(a_i) = \alpha_i$. For each word $x \in A_n^+$, we can easily see that $f(x) = \alpha_{i_m}$ where $i_m = \max\{i \mid a_i \text{ is a letter of } x\}$. Let $x_1$ be the word $a_1 a_1$ and $x_{i+1} = (x_i a_{i+1})^2$ for $1 \le i \le n-1$. That is, $x_n = ((a_1^2 a_2)^2 \cdots a_n)^2$.

**Lemma 4** *The minimal height of a Ramseyan factorization tree for $x_n$ is $n+1$, where $n \ge 2$.*

**Proof.** Consider an arbitrary factorization tree for $x_n$ where $n \geq 2$. We want to argue that every internal node is factorized into two nodes. Consequently, the minimal height of a factorization tree is $\lceil \lg |x_n| \rceil = \lceil \lg(2^{n+1} - 2) \rceil = n + 1$. Suppose on the contrary that there is a node labelled by $w$ such that $d(w) = (w_1, w_2, w_3, \ldots, w_p)$ where $p \geq 3$. Since the out-degree is more than two, we have $f(w_1) = f(w_2) = f(w_3) = \cdots = f(w_p)$, which we also denote by $\alpha_k$. Thus, $a_k$ appears as a symbol in $w_1$, $w_2$ and $w_3$. However, by the way $x_n$ is defined, within three occurrences of $a_k$ there must exist $a_{k+1}$. That is, $a_{k+1}$ exists in either $w_1$, $w_2$ or $w_3$. Hence, either $f(w_1) \neq \alpha_k$, $f(w_2) \neq \alpha_k$ or $f(w_3) \neq \alpha_k$, which is a contradiction. $\square$

For the general case, one may wonder if it is possible to show that for any morphism from $A^+$ to a finite semigroup $S$, any Ramseyan factorization forest must have height at least the size of $S$. The following family of examples proves the negative.

For each $n$, let $S_n$ be the $\mathcal{D}$-trivial semigroup $\{\beta, \alpha_1, \ldots, \alpha_n\}$ with the following associative operation: $\beta\beta = \beta$, $\alpha_i\beta = \beta$, $\beta\alpha_i = \beta$, $\alpha_i\alpha_i = \alpha_i$ and $\alpha_i\alpha_j = \beta$ where $i \neq j$. Let $A_n = \{a_1, \ldots, a_n\}$ and $f$ be a morphism from $A_n^+$ to $S_n$ such that $f(a_i) = \alpha_i$ for $1 \leq i \leq n$.

**Lemma 5** *There is a Ramseyan factorization forest for $f$ of height 4.*

**Proof.** Let $z \in A_n^+$. We define $d(z) = (x, y)$ and $d(x) = (x_1, x_2, \ldots, x_q)$ such that $y$ consists of the same symbol from $A_n$ and for $1 \leq k \leq q$, $x_k = a_{k_1} a_{k_1} \cdots a_{k_1} a_{k_2}$ where $1 \leq k_1 \neq k_2 \leq n$. Thus, $f(x_1) = f(x_2) = \cdots = f(x_q) = \beta$. Moreover, $d(x_k) = (x'_k, a_{k_2})$ and $d(x'_k) = (a_{k_1}, a_{k_1}, \ldots, a_{k_1})$. Note that the factorization tree may have a height smaller than 4 for degenerate cases. $\square$

# References


[1] T.C. Brown (1971), 'An Interesting Combinatorial Method in the Theory of Locally Finite Semigroups', *Pacific Journal of Mathematics*, **36** 285–289.

[2] G. Lallement (1979), *Semigroups and Combinatorial Applications*, John Wiley and Sons, New York.

[3] I. Simon (1990), 'Factorization Forests of Finite Height', *Theor. Comput. Sci.*, **72** 65–94.

[4] I. Simon (1992), 'A Short Proof of the Factorization Forest Theorem', in *Tree Automata and Languages*, M. Nivat and A. Podelski (eds.), Elsevier Science Publishers B.V., 433–438.

[5] I. Simon (1994), 'On Semigroups of Matrices over the Tropical Semiring', *RAIRO ITA*, **28** 277–294.