Bull. Aust. Math. Soc. **113** (2026), 323–330
doi:10.1017/S0004972725100385

THE FINITE BASIS PROBLEM FOR FREE TREE MONOIDS

YAN FENG LUO, ZHEN FENG JIN and WEN TING ZHANG✉

(Received 9 February 2025; accepted 21 June 2025; first published online 8 August 2025)

Abstract

For each $n \geq 1$, let FT_n be the free tree monoid of rank n and E_n the full extensive transformation monoid over the finite chain $\{1, 2, \ldots, n\}$. It is shown that the monoids FT_n and E_{n+1} satisfy the same identities. Therefore, FT_n is finitely based if and only if $n \leq 3$.

2020 *Mathematics subject classification*: primary 20M07; secondary 03C05.

Keywords and phrases: free tree monoid, extensive transformation monoid, $\mathscr{R}$-trivial monoid, finite basis problem.

1. Introduction

Ayyer *et al.* [1] developed a general theory of Markov chains realisable as random walks on $\mathscr{R}$-trivial monoids, providing an elegant and uniform treatment of many classical examples and their generalisations. An essential role in [1] is played by a novel series of finite $\mathscr{R}$-trivial monoids, so-called free tree monoids. For a positive integer n, the *free tree monoid of rank n*, denoted by FT_n, is generated by the identity ε and a totally ordered alphabet $\mathcal{A}_n = \{1 < 2 < \cdots < n\}$ with the relations

$$\begin{aligned} i^2 &= i \quad \text{for each } i = 1, 2, \ldots, n, \\ kik &= ki \quad \text{if } 1 \leq i < k \leq n. \end{aligned}$$

The monoid FT_n is called the free tree monoid because there is a bijection between the elements of FT_n and certain trees. In fact, it follows from [1, Proposition 5.1] that all rewriting rules

$$\{i\mathbf{a}i \to i\mathbf{a} \mid i \in \mathcal{A}_n, \mathbf{a} \in \{1, 2, \ldots, i-1\}^*\} \tag{1.1}$$

constitute a complete rewriting system of FT_n. Consequently, we can identify the elements of FT_n with the reduced words in $\mathcal{A}_n^*$. For any $\mathbf{a} \in \mathcal{A}_n^*$, let $\overline{\mathbf{a}}$ be the unique reduced word corresponding to $\mathbf{a}$. Now, take a reduced word $\overline{\mathbf{a}}$. If $\mathcal{A}_n$ (and therefore $\overline{\mathbf{a}}$)

This research was partially supported by the National Natural Science Foundation of China (Grant nos. 12271224, 12171213, 12401017) and the Fundamental Research Funds for the Central University (Grant no. lzujbky-2023-ey06).

© The Author(s), 2025. Published by Cambridge University Press on behalf of Australian Mathematical Publishing Association Inc.

https://doi.org/10.1017/S0004972725100385 Published online by Cambridge University Press

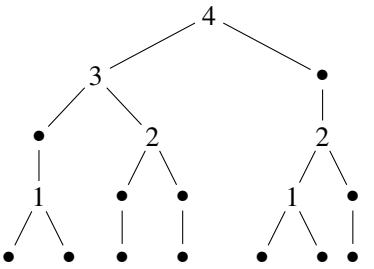


FIGURE 1. The tree $\mathcal{T}_{\mathcal{A}_4}(132412)$.

is empty, let $\mathcal{T}_{\mathcal{A}_n}(\overline{\mathbf{a}})$ be the tree consisting of a single leaf. Otherwise, note that n is the largest letter of $\mathcal{A}_n$. If n appears in $\overline{\mathbf{a}}$, write $\overline{\mathbf{a}} = \mathbf{a}_1 n \mathbf{a}_2$, where $\mathbf{a}_1, \mathbf{a}_2 \in \mathcal{A}_{n-1}^*$ and define $\mathcal{T}_{\mathcal{A}_n}(\overline{\mathbf{a}})$ as the tree whose root has subtrees $\mathcal{T}_{\mathcal{A}_{n-1}}(\mathbf{a}_1)$ and $\mathcal{T}_{\mathcal{A}_{n-1}}(\mathbf{a}_2)$. If n does not appear in $\overline{\mathbf{a}}$, then define $\mathcal{T}_{\mathcal{A}_n}(\overline{\mathbf{a}})$ as the tree whose root has $\mathcal{T}_{\mathcal{A}_{n-1}}(\overline{\mathbf{a}})$ as the single subtree. For example, $\mathcal{T}_{\mathcal{A}_4}(132412)$ is given in Figure 1.

Clearly, the number of leaves of the tree is the length of the word plus one. It follows from [1, Proposition 5.5] that the map $\mathcal{T}_{\mathcal{A}_n}$ is a bijection between the elements of the free tree monoid FT_n and certain trees. By the definition of free tree monoids, FT_n is a finite monoid, $|FT_1| = 2$ and $|FT_{n+1}| = |FT_n|^2 + |FT_n|$.

Let $\mathcal{X}$ be any countably infinite alphabet and $\mathcal{X}^*$ the free monoid over $\mathcal{X}$. Elements of $\mathcal{X}$ are referred to as *variables* and elements of $\mathcal{X}^*$ are referred to as *words*. Let $\mathbf{u} \in \mathcal{X}^*$. The *content* of $\mathbf{u}$, denoted by $\mathsf{con}(\mathbf{u})$, is the set of all variables occurring in $\mathbf{u}$. A semigroup identity is usually written as $\mathbf{u} \approx \mathbf{v}$, where $\mathbf{u}, \mathbf{v} \in \mathcal{X}^*$. A semigroup S *satisfies* $\mathbf{u} \approx \mathbf{v}$ if $\alpha(\mathbf{u}) = \alpha(\mathbf{v})$ in S under all possible substitutions $\alpha : \mathcal{X}^* \to S$. A semigroup is *finitely based* if there is a finite subset of its identities from which all of its identities may be deduced. Otherwise, it is *nonfinitely based*. The variety defined by a given set Σ of identities is the class of all semigroups that satisfy all identities in Σ. The *variety generated* by S, denoted by $\mathrm{var}(S)$, is the variety defined by the set of all identities that S satisfies.

Recently, Sapir and Volkov [7, Proposition 4.5] proved that the free tree monoid FT_n is nonfinitely based for any $n \geq 4$. It follows from [5] that both FT_1 of order 2 and FT_2 of order 6 are finitely based. However, the finite basis property of the monoid FT_3 of order 42 remains unknown.

Let τ be any total transformation on $\{1, 2, \ldots, n\}$. If $\tau(i) \geq i$ for all i, then τ is *extensive*; if $\tau(i) \leq \tau(j)$ whenever $i < j$, then τ is *order-preserving*. The set of all total extensive transformations of the finite chain $\{1, 2, \ldots, n\}$ forms a monoid denoted by E_n. The set of all order-preserving transformations in E_n forms a submonoid, called the Catalan monoid C_n, because the cardinality of C_n coincides with the nth Catalan number $1/(n+1)\binom{2n}{n}$ (see [3, Theorem 3.1]). The equational properties of E_n and C_n have been widely studied. The finite basis problem for Catalan monoids was completely solved and a description of identities holding in Catalan monoids was given in [8].

https://doi.org/10.1017/S0004972725100385 Published online by Cambridge University Press

PROPOSITION 1.1 [8, Theorem 2]. *For every $n \geq 1$, the identity $\mathbf{u} \approx \mathbf{v}$ holds in C_{n+1} if and only if the words $\mathbf{u}$ and $\mathbf{v}$ have the same set of scattered subwords of length less than or equal to n.*

Recall that the word $\mathbf{w} = x_1 x_2 \cdots x_r \in \mathcal{X}^*$ is a *scattered subword* of $\mathbf{u}$ if $\mathbf{u}$ can be decomposed as

$$\mathbf{u} = \mathbf{u}_1 x_1 \mathbf{u}_2 x_2 \cdots \mathbf{u}_r x_r \mathbf{u}_{r+1}$$

for some $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{r+1} \in \mathcal{X}^*$. Further, this decomposition is called *the first occurrence of the scattered subword* $\mathbf{w}$ *in the word* $\mathbf{u}$ if $x_i \notin \mathsf{con}(\mathbf{u}_i)$ for all $i = 1, \ldots, r$. Define the relation $\sim_n$ on $\mathcal{X}^*$ by letting $\mathbf{u} \sim_n \mathbf{v}$ if and only if the following conditions hold:

(I) the words $\mathbf{u}$ and $\mathbf{v}$ have the same sets of scattered subwords of length less than or equal to n;

(II) if $\mathbf{w} = x_1 x_2 \cdots x_r$ $(r \leq n)$ is an arbitrary common scattered subword of $\mathbf{u}$ and $\mathbf{v}$, and

$$\mathbf{u} = \mathbf{u}_1 x_1 \mathbf{u}_2 x_2 \cdots \mathbf{u}_r x_r \mathbf{u}_{r+1},$$
$$\mathbf{v} = \mathbf{v}_1 x_1 \mathbf{v}_2 x_2 \cdots \mathbf{v}_r x_r \mathbf{v}_{r+1},$$

are the first occurrences of the scattered subword $\mathbf{w}$ in $\mathbf{u}$ and $\mathbf{v}$, then $\mathsf{con}(\mathbf{u}_i) = \mathsf{con}(\mathbf{v}_i)$ for $i = 1, 2, \ldots, r+1$.

In [2], Goldberg studied the equational property and the finite basis problem for E_{n+1}.

PROPOSITION 1.2 [2, Proposition 2.2]. *The identity $\mathbf{u} \approx \mathbf{v}$ holds in the monoid E_{n+1} if and only if $\mathbf{u} \sim_{n-1} \mathbf{v}$.*

A combination of [2, Theorem 1.1], [4, Theorem 1.2] and [6, Theorem 1] completely solved the finite basis problem for E_{n+1}.

PROPOSITION 1.3. *The monoid E_{n+1} is finitely based if and only if $n \leq 3$.*

In this paper, we show that the identity $\mathbf{u} \approx \mathbf{v}$ holds in the free tree monoid FT_n if and only if $\mathbf{u} \sim_{n-1} \mathbf{v}$, and so by Proposition 1.2, the monoids FT_n and E_{n+1} satisfy the same identities. Consequently, the finite basis problem for FT_n is completely solved.

THEOREM 1.4. *For each positive integer n, the monoids FT_n and E_{n+1} satisfy the same identities. Consequently, FT_n is finitely based if and only if $n \leq 3$.*

2. Proof of Theorem 1.4

Given $\mathbf{a} \in \mathcal{A}_n^*$, the *support* of $\mathbf{a}$, denoted by $\mathsf{supp}(\mathbf{a})$, is the set of all letters occurring in $\mathbf{a}$. The number of occurrences of $i \in \mathcal{A}_n$ in $\mathbf{a}$ is denoted by $\mathsf{occ}(i, \mathbf{a})$. Let $\mathbf{a}[1, 2, \ldots, n-1]$ denote the word obtained from $\mathbf{a}$ by retaining only the letters $1, 2, \ldots, n-1$.

https://doi.org/10.1017/S0004972725100385 Published online by Cambridge University Press

LEMMA 2.1. *Let $n \geq 2$ and $\mathbf{u}, \mathbf{v} \in \mathcal{X}^*$. If $\mathbf{w} = x_1 x_2 \cdots x_r$ $(r \leq n-1)$ is an arbitrary common scattered subword of $\mathbf{u}$ and $\mathbf{v}$, and*

$$\mathbf{u} = \mathbf{u}_1 x_1 \mathbf{u}_2 x_2 \cdots \mathbf{u}_r x_r \mathbf{u}_{r+1},$$
$$\mathbf{v} = \mathbf{v}_1 x_1 \mathbf{v}_2 x_2 \cdots \mathbf{v}_r x_r \mathbf{v}_{r+1},$$

are the first occurrences of $\mathbf{w}$ in $\mathbf{u}$ and $\mathbf{v}$, and $\mathsf{con}(\mathbf{u}_i) \neq \mathsf{con}(\mathbf{v}_i)$ for some $i \in \{1, 2, \ldots, r+1\}$, then FT_n does not satisfy the identity $\mathbf{u} \approx \mathbf{v}$.

PROOF. Without loss of generality, assume that $y \in \mathsf{con}(\mathbf{u}_i) \setminus \mathsf{con}(\mathbf{v}_i)$ for some $i \in \{1, 2, \ldots, r+1\}$. Then, $x_1 x_2 \cdots x_{i-1} y x_i \cdots x_r$ is a scattered subword of $\mathbf{u}$. Note that x_j may equal x_k for some distinct j and k, and y may equal x_j for some j. We partition the multiset $\{x_1, x_2, \ldots, x_r, y\}$, variables in it possibly equal, into the disjoint union of some multisets such that the same variables belong to the same multiset. Without loss of generality, assume that

$$\{x_1, \ldots, x_r, y\} = \{x_{j_1^1} = \cdots = x_{j_1^{k_1}}\} \sqcup \{x_{j_2^1} = \cdots = x_{j_2^{k_2}}\} \sqcup \cdots \sqcup \{x_{j_\ell^1} = \cdots = x_{j_\ell^{k_\ell}} = y\},$$

where $k_1, \ldots, k_{\ell-1} \geq 1$, $k_\ell \geq 0$ and $x_{j_1^1}, x_{j_2^1}, \ldots, x_{j_{\ell-1}^1}, y$ are pairwise distinct and

$$j_1^1 < \cdots < j_1^{k_1}, \quad j_2^1 < \cdots < j_2^{k_2}, \quad \ldots, \quad j_\ell^1 < \cdots < j_\ell^{k_\ell}.$$

Now, we define the substitution $\beta : \mathcal{X}^* \to FT_n$ as follows:

$$\begin{aligned}
\beta(x_{j_1^1}) = \cdots = \beta(x_{j_1^{k_1}}) &= (n+1-j_1^{k_1}) \cdots (n+1-j_1^2)(n+1-j_1^1),\\
\beta(x_{j_2^1}) = \cdots = \beta(x_{j_2^{k_2}}) &= (n+1-j_2^{k_2}) \cdots (n+1-j_2^2)(n+1-j_2^1),\\
&\vdots\\
\beta(x_{j_\ell^1}) = \cdots = \beta(x_{j_\ell^{k_\ell}}) = \beta(y) &= 1(n+1-j_\ell^{k_\ell}) \cdots (n+1-j_\ell^2)(n+1-j_\ell^1),\\
\beta(z) &= \varepsilon \ \text{ for any } \ z \in \mathcal{X} \setminus \{x_1, x_2, \ldots, x_r, y\}.
\end{aligned}$$

By the definition of β,

$$\beta(\mathbf{u}) = \beta(\mathbf{u}_1)\beta(x_1)\beta(\mathbf{u}_2)\beta(x_2) \cdots \beta(\mathbf{u}_r)\beta(x_r)\beta(\mathbf{u}_{r+1}),$$
$$\beta(\mathbf{v}) = \beta(\mathbf{v}_1)\beta(x_1)\beta(\mathbf{v}_2)\beta(x_2) \cdots \beta(\mathbf{v}_r)\beta(x_r)\beta(\mathbf{v}_{r+1}),$$

where:

(a) $\mathsf{occ}(n, \beta(x_1)) = \mathsf{occ}(n-1, \beta(x_2)) = \cdots = \mathsf{occ}(n+1-r, \beta(x_r)) = 1$;
(b) $n \notin \mathsf{supp}(\beta(\mathbf{u}_1)\beta(\mathbf{v}_1)), n-1 \notin \mathsf{supp}(\beta(\mathbf{u}_2)\beta(\mathbf{v}_2)), \ldots, n+1-r \notin \mathsf{supp}(\beta(\mathbf{u}_r)\beta(\mathbf{v}_r))$.

Since $y \in \mathsf{con}(\mathbf{u}_i) \setminus \mathsf{con}(\mathbf{v}_i)$, it follows that $1 \in \mathsf{supp}(\beta(\mathbf{u}_i)) \setminus \mathsf{supp}(\beta(\mathbf{v}_i))$. By the definition of β, the reduced words $\overline{\beta(\mathbf{u})}$ and $\overline{\beta(\mathbf{v})}$ corresponding respectively to $\beta(\mathbf{u})$ and $\beta(\mathbf{v})$ can be written as

$$\overline{\beta(\mathbf{u})} = \mathbf{a}_1 n \mathbf{a}_2 (n-1) \cdots \mathbf{a}_{i-1}(n-i+2)\mathbf{a}_i(n-i+1) \cdots \mathbf{a}_r(n-r+1)\mathbf{a}_{r+1},$$
$$\overline{\beta(\mathbf{v})} = \mathbf{b}_1 n \mathbf{b}_2 (n-1) \cdots \mathbf{b}_{i-1}(n-i+2)\mathbf{b}_i(n-i+1) \cdots \mathbf{b}_r(n-r+1)\mathbf{b}_{r+1},$$

https://doi.org/10.1017/S0004972725100385 Published online by Cambridge University Press

where:

(c) $\mathbf{a}_1, \dots, \mathbf{a}_{r+1}, \mathbf{b}_1, \dots, \mathbf{b}_{r+1} \in \mathcal{A}_n^*$ and $1 \in \mathsf{supp}(\mathbf{a}_i) \setminus \mathsf{supp}(\mathbf{b}_i)$;
(d) $n \notin \mathsf{supp}(\mathbf{a}_1\mathbf{a}_2 \cdots \mathbf{a}_{r+1}\mathbf{b}_1\mathbf{b}_2 \cdots \mathbf{b}_{r+1}), (n-1) \notin \mathsf{supp}(\mathbf{a}_2 \cdots \mathbf{a}_{r+1}\mathbf{b}_2 \cdots \mathbf{b}_{r+1}), \dots,$ $(n-r+1) \notin \mathsf{supp}(\mathbf{a}_r\mathbf{a}_{r+1}\mathbf{b}_r\mathbf{b}_{r+1})$.

Consequently, $\overline{\beta(\mathbf{u})} \neq \overline{\beta(\mathbf{v})}$, this implies that $\beta(\mathbf{u}) \neq \beta(\mathbf{v})$ in FT_n. □

LEMMA 2.2. *Let* $\mathbf{u}, \mathbf{v} \in \mathcal{X}^*$ *and* $\mathbf{u} \sim_n \mathbf{v}$. *If*

$$\mathbf{u} = \mathbf{u}_1 x \mathbf{u}_2, \tag{2.1}$$

$$\mathbf{v} = \mathbf{v}_1 x \mathbf{v}_2 \tag{2.2}$$

for some $\mathbf{u}_1, \mathbf{u}_2, \mathbf{v}_1, \mathbf{v}_2 \in \mathcal{X}^*$ *such that* $x \notin \mathsf{con}(\mathbf{u}_1\mathbf{v}_1)$, *then* $\mathbf{u}_1 \sim_{n-1} \mathbf{v}_1$ *and* $\mathbf{u}_2 \sim_{n-1} \mathbf{v}_2$.

PROOF. We only need to prove $\mathbf{u}_1 \sim_{n-1} \mathbf{v}_1$ since the result $\mathbf{u}_2 \sim_{n-1} \mathbf{v}_2$ can be obtained similarly. Since $\mathbf{u} \sim_n \mathbf{v}$, the words $\mathbf{u}$ and $\mathbf{v}$ have the same sets of scattered subwords of length less than or equal to n. Assume that $y_1y_2 \cdots y_r$ $(r \leq n-1)$ is any scattered subword of $\mathbf{u}_1$. Since (2.1), (2.2) are the first occurrences of the scattered subword x in $\mathbf{u}$ and $\mathbf{v}$, it follows from $y_1 \in \mathsf{con}(\mathbf{u}_1)$ and condition (II) that $y_1 \in \mathsf{con}(\mathbf{v}_1)$. Let

$$\mathbf{u} = \mathbf{u}_3 y_1 \mathbf{u}_4 x \mathbf{u}_2, \tag{2.3}$$

$$\mathbf{v} = \mathbf{v}_3 y_1 \mathbf{v}_4 x \mathbf{v}_2 \tag{2.4}$$

for some $\mathbf{u}_3, \mathbf{u}_4, \mathbf{v}_3, \mathbf{v}_4 \in \mathcal{X}^*$ such that $y_1 \notin \mathsf{con}(\mathbf{u}_3\mathbf{v}_3)$. Since $y_1y_2 \cdots y_r$ is a scattered subword of $\mathbf{u}_1$, it follows that $y_2 \in \mathsf{con}(\mathbf{u}_4)$. Note that as (2.3), (2.4) are the first occurrences of the scattered subword y_1x in $\mathbf{u}$ and $\mathbf{v}$, it follows from condition (II) that $y_2 \in \mathsf{con}(\mathbf{v}_4)$. Repeating this process shows that $y_1y_2 \cdots y_r$ $(r \leq n-1)$ is a scattered subword of $\mathbf{v}_1$. By symmetry, if $y_1y_2 \cdots y_r$ $(r \leq n-1)$ is a scattered subword of $\mathbf{v}_1$, then it is also a scattered subword of $\mathbf{u}_1$. Hence, the words $\mathbf{u}_1$ and $\mathbf{v}_1$ have the same sets of scattered subwords of length less than or equal to $n-1$.

If $y_1y_2 \cdots y_r$ $(r \leq n-1)$ is an arbitrary common scattered subword of $\mathbf{u}_1$ and $\mathbf{v}_1$, and

$$\mathbf{u}_1 = \mathbf{u}_1' y_1 \mathbf{u}_2' y_2 \cdots \mathbf{u}_r' y_r \mathbf{u}_{r+1}',$$
$$\mathbf{v}_1 = \mathbf{v}_1' y_1 \mathbf{v}_2' y_2 \cdots \mathbf{v}_r' y_r \mathbf{v}_{r+1}',$$

are the first occurrences of the scattered subword $y_1y_2 \cdots y_r$ in $\mathbf{u}_1$ and $\mathbf{v}_1$, then

$$\mathbf{u} = \mathbf{u}_1 x \mathbf{u}_2 = \mathbf{u}_1' y_1 \mathbf{u}_2' y_2 \cdots \mathbf{u}_r' y_r \mathbf{u}_{r+1}' x \mathbf{u}_2,$$
$$\mathbf{v} = \mathbf{v}_1 x \mathbf{v}_2 = \mathbf{v}_1' y_1 \mathbf{v}_2' y_2 \cdots \mathbf{v}_r' y_r \mathbf{v}_{r+1}' x \mathbf{v}_2$$

are the first occurrences of the scattered subword $y_1 \cdots y_r x$ in $\mathbf{u}$ and $\mathbf{v}$. Since $\mathbf{u} \sim_n \mathbf{v}$, it follows from condition (II) that $\mathsf{con}(\mathbf{u}_i') = \mathsf{con}(\mathbf{v}_i')$ for $i = 1, 2, \dots, r+1$. □

PROOF OF THEOREM 1.4. Clearly, FT_1 is isomorphic to E_2. Therefore, we assume that $n \geq 2$. Let $\mathbf{u}, \mathbf{v} \in \mathcal{X}^*$. It suffices to show that the identity $\mathbf{u} \approx \mathbf{v}$ holds in FT_n if and only if $\mathbf{u} \sim_{n-1} \mathbf{v}$. Then, by Proposition 1.2, the monoids FT_n and E_{n+1} satisfy the same identities.

https://doi.org/10.1017/S0004972725100385 Published online by Cambridge University Press

Assume that $\mathbf{u} \approx \mathbf{v}$ holds in FT_n. Then, $\mathbf{u} \approx \mathbf{v}$ holds in the Catalan monoid C_{n+1} since C_{n+1} is a homomorphic image of FT_n [7, Lemma 4.3]. It follows from Proposition 1.1 that the words $\mathbf{u}$ and $\mathbf{v}$ have the same sets of scattered subwords of length less than or equal to n, and so condition (I) holds. Condition (II) holds by Lemma 2.1.

Conversely, assume that $\mathbf{u} \sim_{n-1} \mathbf{v}$. Now, we prove that $\mathbf{u} \approx \mathbf{v}$ holds in FT_n. We proceed by induction on n. If $n = 2$, then $\mathbf{u} \sim_1 \mathbf{v}$. For any substitution $\alpha : \mathcal{X}^* \to FT_2$, we have $\mathsf{supp}(\alpha(\mathbf{u})) = \mathsf{supp}(\alpha(\mathbf{v}))$ by condition (I). If $2 \notin \mathsf{supp}(\alpha(\mathbf{u})) = \mathsf{supp}(\alpha(\mathbf{v}))$, then clearly $\alpha(\mathbf{u}) = \alpha(\mathbf{v}) = 1$ or ε. If $2 \in \mathsf{supp}(\alpha(\mathbf{u})) = \mathsf{supp}(\alpha(\mathbf{v}))$, then let

$$\mathbf{u} = \mathbf{u}_1 x \mathbf{u}_2 \tag{2.5}$$

for some $x \in \mathcal{X}, \mathbf{u}_1, \mathbf{u}_2 \in \mathcal{X}^*$ such that $2 \in \mathsf{supp}(\alpha(x))$ and $2 \notin \mathsf{supp}(\alpha(\mathbf{u}_1))$. Then, $x \in \mathsf{con}(\mathbf{v})$ by condition (I). Denote

$$\mathbf{v} = \mathbf{v}_1 x \mathbf{v}_2, \tag{2.6}$$

where $\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{X}^*$ and $x \notin \mathsf{con}(\mathbf{v}_1)$. Since (2.5), (2.6) are the first occurrences of the scattered subword x in $\mathbf{u}$ and $\mathbf{v}$, it follows from $\mathbf{u} \sim_1 \mathbf{v}$ and condition (II) that $\mathsf{con}(\mathbf{u}_1) = \mathsf{con}(\mathbf{v}_1)$ and $\mathsf{con}(\mathbf{u}_2) = \mathsf{con}(\mathbf{v}_2)$. Hence, $2 \notin \mathsf{supp}(\alpha(\mathbf{u}_1)) = \mathsf{supp}(\alpha(\mathbf{v}_1))$ and $\mathsf{supp}(\alpha(\mathbf{u}_2)) = \mathsf{supp}(\alpha(\mathbf{v}_2))$, and so $\alpha(\mathbf{u}) = \alpha(\mathbf{v})$ by (1.1). Thus, $\mathbf{u} \approx \mathbf{v}$ holds in FT_2.

Suppose that $\mathbf{u} \approx \mathbf{v}$ holds in FT_k if $\mathbf{u} \sim_{k-1} \mathbf{v}$ for $2 \le k \le n-1$. We need to show $\mathbf{u} \sim_{n-1} \mathbf{v}$ implies that $\mathbf{u} \approx \mathbf{v}$ holds in FT_n. Since $\mathbf{u} \sim_{n-1} \mathbf{v}$, it follows from condition (I) that $\mathsf{supp}(\alpha(\mathbf{u})) = \mathsf{supp}(\alpha(\mathbf{v}))$ for any substitution $\alpha : \mathcal{X}^* \to FT_n$. If $n \notin \mathsf{supp}(\alpha(\mathbf{u})) = \mathsf{supp}(\alpha(\mathbf{v}))$, then α can be seen as a substitution $\alpha : \mathcal{X}^* \to FT_{n-1}$. Since $\mathbf{u} \sim_{n-1} \mathbf{v}$, it follows that $\mathbf{u} \sim_{n-2} \mathbf{v}$. According to the inductive hypothesis, we have $\alpha(\mathbf{u}) = \alpha(\mathbf{v})$. If $n \in \mathsf{supp}(\alpha(\mathbf{u})) = \mathsf{supp}(\alpha(\mathbf{v}))$, then let

$$\mathbf{u} = \mathbf{u}_1 x \mathbf{u}_2 \tag{2.7}$$

for some $x \in \mathcal{X}, \mathbf{u}_1, \mathbf{u}_2 \in \mathcal{X}^*$ such that $n \in \mathsf{supp}(\alpha(x))$ and $n \notin \mathsf{supp}(\alpha(\mathbf{u}_1))$. Hence, $x \in \mathsf{con}(\mathbf{v})$ by condition (I). Denote

$$\mathbf{v} = \mathbf{v}_1 x \mathbf{v}_2, \tag{2.8}$$

where $\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{X}^*$ and $x \notin \mathsf{con}(\mathbf{v}_1)$. Since (2.7), (2.8) are the first occurrences of the scattered subword x in $\mathbf{u}$ and $\mathbf{v}$, it follows from $\mathbf{u} \sim_{n-1} \mathbf{v}$ and condition (II) that $\mathsf{con}(\mathbf{u}_1) = \mathsf{con}(\mathbf{v}_1)$ and $\mathsf{con}(\mathbf{u}_2) = \mathsf{con}(\mathbf{v}_2)$. For

$$\alpha(\mathbf{u}) = \alpha(\mathbf{u}_1)\alpha(x)\alpha(\mathbf{u}_2),$$
$$\alpha(\mathbf{v}) = \alpha(\mathbf{v}_1)\alpha(x)\alpha(\mathbf{v}_2),$$

we have $n \in \mathsf{supp}(\alpha(x))$, $n \notin \mathsf{supp}(\alpha(\mathbf{u}_1)\alpha(\mathbf{v}_1))$, and all n in $\alpha(\mathbf{u}_2), \alpha(\mathbf{v}_2)$ can be eliminated by the rewriting rules in (1.1). Hence, there is a substitution $\beta : \mathcal{X}^* \to FT_{n-1}$ given by

$$\mathbf{a} \mapsto \begin{cases} \alpha(\mathbf{a}) & \text{if } n \notin \mathsf{supp}(\alpha(\mathbf{a})), \\ \alpha(\mathbf{a})[1, 2, \ldots, n-1] & \text{if } n \in \mathsf{supp}(\alpha(\mathbf{a})), \end{cases}$$

https://doi.org/10.1017/S0004972725100385 Published online by Cambridge University Press

such that

$$\alpha(\mathbf{u}) = \alpha(\mathbf{u}_1)\alpha(x)\alpha(\mathbf{u}_2) = \beta(\mathbf{u}_1)\alpha(x)\beta(\mathbf{u}_2),$$
$$\alpha(\mathbf{v}) = \alpha(\mathbf{v}_1)\alpha(x)\alpha(\mathbf{v}_2) = \beta(\mathbf{v}_1)\alpha(x)\beta(\mathbf{v}_2).$$

By Lemma 2.2, $\mathbf{u}_1 \sim_{n-2} \mathbf{v}_1$ and $\mathbf{u}_2 \sim_{n-2} \mathbf{v}_2$. According to the inductive hypothesis, $\beta(\mathbf{u}_1) = \beta(\mathbf{v}_1)$ and $\beta(\mathbf{u}_2) = \beta(\mathbf{v}_2)$, and so $\alpha(\mathbf{u}) = \alpha(\mathbf{v})$. That is, $\mathbf{u} \approx \mathbf{v}$ holds in FT_n. □

COROLLARY 2.3. *We have* $\mathrm{var}(C_2) = \mathrm{var}(FT_1) = \mathrm{var}(E_2)$, *and* $\mathrm{var}(C_{n+1})$ *is a proper subvariety of* $\mathrm{var}(FT_n) = \mathrm{var}(E_{n+1})$ *for any* $n \geq 2$.

PROOF. Clearly, C_2, FT_1 and E_2 are isomorphic to each other. For $n \geq 2$, it follows from Theorem 1.4 and [7, Lemma 4.3] that $\mathrm{var}(C_{n+1}) \subseteq \mathrm{var}(FT_n) = \mathrm{var}(E_{n+1})$. Next, we prove that $\mathrm{var}(C_{n+1})$ is a proper subvariety of $\mathrm{var}(FT_n)$. Consider the identity $(xy)^n \approx (yx)^n$. It is routine to check that $(xy)^n$ and $(yx)^n$ have the same set of scattered subwords of length less than or equal to n, and it follows from Proposition 1.1 that C_{n+1} satisfies $(xy)^n \approx (yx)^n$. However,

$$(xy)^n = x \cdot y(xy)^{n-1} \quad \text{and} \quad (yx)^n = y \cdot x \cdot (yx)^{n-1}$$

are the first occurrences of the common scattered subword x in $(xy)^n$ and $(yx)^n$. Clearly, $(xy)^n \approx (yx)^n$ does not satisfy condition (II). Therefore, FT_n does not satisfy $(xy)^n \approx (yx)^n$. □

References

[1] A. Ayyer, A. Schilling, B. Steinberg and N. M. Thiéry, 'Markov chains, $\mathscr{R}$-trivial monoids and representation theory', *Internat. J. Algebra Comput.* **25**(1–2) (2015), 169–231.

[2] I. A. Goldberg, 'On the finite basis problem for the monoids of extensive transformations', in: *Semigroups and Formal Languages* (eds. J. M. André, V. H. Fernandes, G. M. S. Gomes, M. J. J. Branco, J. Fountain and J. C. Meakin) (World Scientific Publishing, Hackensack, NJ, 2007), 101–110.

[3] P. M. Higgins, 'Combinatorial results for semigroups of order-preserving mappings', *Math. Proc. Cambridge Philos. Soc.* **113** (1993), 281–296.

[4] E. W. H. Lee, 'Hereditarily finitely based monoids of extensive transformations', *Algebra Universalis* **61**(1) (2009), 31–58.

[5] E. W. H. Lee and J. R. Li, 'Minimal non-finitely based monoids', *Dissertationes Math.* **475** (2011), 1–65.

[6] J. R. Li and Y. F. Luo, 'Equational property of certain transformation monoids', *Internat. J. Algebra Comput.* **20**(6) (2010), 833–845.

[7] O. B. Sapir and M. V. Volkov, 'Catalan monoids inherently nonfinitely based relative to finite $\mathscr{R}$-trivial semigroups', *J. Algebra* **633** (2023), 138–171.

[8] M. V. Volkov, 'Reflexive relations, extensive transformations and piecewise testable languages of a given height', *Internat. J. Algebra Comput.* **14** (2004), 817–827.

YAN FENG LUO, School of Mathematics and Statistics,
Lanzhou University, Lanzhou, Gansu 730000, PR China
e-mail: jinzhf20@lzu.edu.cn

ZHEN FENG JIN, School of Mathematics and Statistics,
Lanzhou University, Lanzhou, Gansu 730000, PR China
e-mail: luoyf@lzu.edu.cn

https://doi.org/10.1017/S0004972725100385 Published online by Cambridge University Press

WEN TING ZHANG, School of Mathematics and Statistics,
Lanzhou University, Lanzhou, Gansu 730000, PR China
e-mail: zhangwt@lzu.edu.cn

https://doi.org/10.1017/S0004972725100385 Published online by Cambridge University Press