TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 335, Number 1, January 1993

INVERSE MONOIDS, TREES, AND CONTEXT-FREE LANGUAGES

STUART W. MARGOLIS AND JOHN C. MEAKIN

ABSTRACT. This paper is concerned with a study of inverse monoids presented by a set X subject to relations of the form $e_i = f_i$, $i \in I$, where e_i and f_i are Dyck words, i.e. idempotents of the free inverse monoid on X. Some general results of Stephen are used to reduce the word problem for such a presentation to the membership problem for a certain subtree of the Cayley graph of the free group on X. In the finitely presented case the word problem is solved by using Rabin's theorem on the second order monadic logic of the infinite binary tree. Some connections with the theory of rational subsets of the free group and the theory of context-free languages are explored.

1. INTRODUCTION

We refer the reader to Lallement [5] for basic information about semigroups and connections with automata theory and formal language theory. We shall be concerned in this paper primarily with inverse semigroups and inverse monoids. For the convenience of the reader we briefly summarize the basic notions and results about inverse monoids that we will need in the present paper; many more details and results about inverse monoids may be found in the book of Petrich [9].

An *inverse semigroup* is a semigroup S with the property that, for each $a \in S$ there is a unique element $a^{-1} \in S$ such that $a = aa^{-1}a$ and $a^{-1} = a^{-1}aa^{-1}$. If S has an identity 1, we refer to it as an *inverse monoid.* Equivalently, an inverse semigroup is a (von Neumann) regular semigroup in which the idempotents commute. It follows that the set $E(S)$ of idempotents of an inverse semigroup S forms a (lower) semilattice with respect to multiplication as the meet operation. Each inverse semigroup S is equipped with a natural partial order relation $\leq$ on S defined by

$$a \leq b \text{ (for } a, b \in S) \quad \text{iff } a = eb \text{ for some } e \in E(S).$$

There is a smallest congruence $\sigma = \sigma_S$ on S such that S/σ is a group (σ is called "the minimal group congruence on S" and S/σ is called "the maximal group homomorphic image of S"). In fact

$$\sigma = \{(a, b) \in S \times S\colon \exists c \in S \text{ such that } c \leq a \text{ and } c \leq b\}.$$

S is called *E-unitary* if and only if σ_S is *idempotent-pure* (i.e. if $a\sigma = e\sigma$ for some $a \in S$, $e \in E(S)$, then $a \in E(S)$). There are many equivalent

Received by the editors March 28, 1989 and, in revised form, October 5, 1990.
1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 20M05, 20M18, 20M35.
Research supported by NSF Grant No. DMS 8702019.

definitions of E-unitary and this concept plays an important role in inverse semigroup theory (see Petrich [9] for much more information about this).

The standard example of an inverse semigroup to keep in mind is the symmetric inverse monoid $\mathrm{SIM}(X)$ on a set X. The monoid $\mathrm{SIM}(X)$ consists of all partial one-to-one maps of X (i.e. all bijections from subsets of X to subsets of X) under the usual composition of partial functions. Thus if $D(\alpha)$ (resp. $R(\alpha)$) denotes the domain (resp. range) of the partial bijection α, then for all $\alpha, \beta \in \mathrm{SIM}(X)$, $D(\alpha\beta) = (R(\alpha) \cap D(\beta))\alpha^{-1}$ and $R(\alpha\beta) = (R(\alpha) \cap D(\beta))\beta$. Clearly $\mathrm{SIM}(X)$ has a zero, the "empty mapping" from the empty subset of X to itself. It is easy to see that $\mathrm{SIM}(X)$ is an inverse monoid with $\alpha^{-1}: R(\alpha) \to D(\alpha)$ as the inverse of the bijection α. The importance of this example stems from the following result, known as the Preston-Wagner representation theorem.

Theorem 10. *Let S be an inverse semigroup and for each $a \in S$ define the partial map ρ_a by $D(\rho_a) = Sa^{-1}$, $R(\rho_a) = Sa$, and $x\rho_a = xa$ for all $x \in D(\rho_a)$. Then $\rho_a \in \mathrm{SIM}(S)$ and the map $f: a \to \rho_a$ is an embedding of S into $\mathrm{SIM}(S)$.*

Thus inverse semigroup theory is concerned with the study of partial one-to-one transformations. The representation $f: S \to \mathrm{SIM}(S)$ described in Theorem 1.1 is called the *Preston-Wagner representation* of S. Closely related to the Preston-Wagner representation is the *Schützenberger representation* of S relative to an $\mathscr{R}$-class R of S. (Recall that the $\mathscr{R}$-class R_a of an element a in the inverse semigroup S is the set

$$R_a = \{b \in S: aS = bS\} = \{b \in S: aa^{-1} = bb^{-1}\};$$

the $\mathscr{L}$-class L_a is defined dually.) If R is an $\mathscr{R}$-class of S then the Schützenberger representation of S relative to R is the representation $P_R: S \to \mathrm{SIM}(R)$ defined by

$$P_R(a) = \{(x, y) \in R \times R: y = xa\}.$$

It is clear that P_R is a *transitive* representation of S by partial one-to-one transformations of R and that the Preston-Wagner representation f of S is the sum of the Schützenberger representations P_R as R runs over the set of all $\mathscr{R}$-classes of S.

Inverse semigroups form a variety of algebras of type $\langle 2, 1\rangle$ defined by associativity and the laws:

$$x = xx^{-1}x, \quad (x^{-1})^{-1} = x, \quad x^{-1}xy^{-1}y = y^{-1}yx^{-1}x.$$

Inverse monoids form a variety of algebras of type $\langle 2, 1, 0\rangle$ defined by the above laws and $x \cdot 1 = 1 \cdot x = x$. As such, free inverse semigroups (monoids) exist. We denote the free inverse semigroup (resp. monoid) on a set X by $\mathrm{FIS}(X)$ (resp. $\mathrm{FIM}(X)$). To construct $\mathrm{FIS}(X)$ we let X^{-1} be a set disjoint from X and in one-to-one correspondence with X by a map $x \to x^{-1}$ $(x \in X)$. Then $\mathrm{FIS}(X) \cong (X \cup X^{-1})^+/\rho$ where $(X \cup X^{-1})^+$ denotes the free semigroup on $X \cup X^{-1}$ and ρ is the Wagner congruence on $(X \cup X^{-1})^+$; i.e., ρ is the smallest congruence on $(X \cup X^{-1})^+$ that forces the laws above to hold in $(X \cup X^{-1})^+/\rho$. Also, $\mathrm{FIM}(X) \cong (X \cup X^{-1})^*/\rho \cong \mathrm{FIS}(X)^1$. Of course this description of ρ is not effective—we would clearly like some sort of algorithm for deciding when two words $u, v \in (X \cup X^{-1})^*$ are ρ-related. That is, we would like to solve the *word problem* for $\mathrm{FIM}(X)$. The first explicit description of the structure of $\mathrm{FIM}(X)$ was provided by Scheiblich [12]. Much work has been done on this semigroup since then and we refer the reader to Petrich [9] for additional references and results. An elegant solution to the word problem for $\mathrm{FIM}(X)$ was provided by Munn [8] who associated with each word $u \in (X \cup X^{-1})^*$ a certain finite tree, which we may identify with a subtree of the Cayley graph of the free group $\mathrm{FG}(X)$ on X.

Let $G = gp\langle X \colon T\rangle$ be the group presented by the set X of generators and the set T of relations, and let $f_T \colon (X \cup X^{-1})^* \to G$ be the natural map from $(X \cup X^{-1})^*$ onto G. The *Cayley graph* $\Gamma = \Gamma(X, T)$ associated with this presentation has as *vertices* the elements of G and has an *edge* $(g, x, g \cdot x f_T)$ for each $x \in X \cup X^{-1}$ and $g \in G$. Note that we are considering a *graph* here as a digraph with involution in the sense of Serre [13]. We may think of the edge $(g, x, g \cdot x f_T)$ of $\Gamma(X, T)$ as being labelled by x, with initial vertex g and terminal vertex $g \cdot x f_T$. The edge $(g \cdot x f_T, x^{-1}, g)$ may be viewed as the inverse of $(g, x, g \cdot x f_T)$. The pair of edges $(g, x, g \cdot x f_T)$ and $(g \cdot x f_T, x^{-1}, g)$ of $\Gamma(X; T)$ is usually represented by the segment

$$g \xrightarrow{\quad x \quad} g \cdot x f_T$$

or more simply by

$$g \xrightarrow{\quad x \quad} g \cdot x$$

when sketching the Cayley graph. Recall that $\Gamma(X; T)$ is a *tree* if and only if $T = \varnothing$ (i.e. when $G = \mathrm{FG}(X)$ relative to the usual representation). We denote the tree of $\mathrm{FG}(X)$ by $\Gamma(X)$. Note that G acts on $\Gamma(X; T)$ by left translation. The action of $h \in G$ on the edge

$$g \xrightarrow{\quad x \quad} g \cdot x$$

yields the edge

$$hg \xrightarrow{\quad x \quad} hg \cdot x.$$

Associated with each group presentation $G = gp\langle X \colon T\rangle$ we build an inverse monoid $M = M(X; T)$ as follows. Let

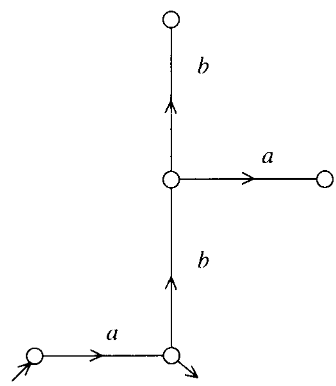
$$M(X; T) = \{(\Gamma, g) \colon \Gamma \text{ is a finite connected subgraph of } \Gamma(X, T) \text{ containing } 1 \text{ and } g \text{ as vertices}\}$$

with multiplication $(\Gamma, g)(\Delta, h) = (\Gamma \cup g \cdot \Delta, gh)$, where $g \cdot \Delta$ denotes the graph obtained from Δ by acting on each vertex (edge) of Δ on the left by g. The following result is proved in Margolis and Meakin [6].

Theorem 1.1. *If $G = gp\langle X \colon T\rangle$ then $M = M(X; T)$ is an E-unitary inverse monoid with maximal group homomorphic image G. In particular, if $T = \varnothing$ then the corresponding inverse monoid $M = M(X; T)$ is isomorphic to the free inverse monoid on X.*

In fact many more interesting properties of the monoids $M(X; T)$ are established in [6]. In particular it is shown that this construction naturally induces a functor $\mathscr{M}$ from the category of X-generated groups to the category of X-generated E-unitary inverse monoids which is left adjoint to the functor σ from X-generated E-unitary inverse monoids to X-generated groups. The construction of the monoids $M(X; T)$ may also be used to construct the relatively free objects in certain varieties of inverse monoids (see [6]).

For the present paper, we are concerned only with case $T = \varnothing$, in which case the corresponding Cayley graph $\Gamma(X)$ is a tree and the corresponding inverse monoid $M(X; \varnothing)$ is $\mathrm{FIM}(X)$. This construction immediately yields a version of Munn's solution to the word problem for $\mathrm{FIM}(X)$, which we now describe. For each word $u \in (X \cup X^{-1})^*$ we denote by $\mathrm{MT}(u)$ the subtree of $\Gamma(X)$ obtained by traversing the path in $\Gamma(X)$ that starts at 1 and is labelled by the word u. Clearly $\mathrm{MT}(u)$ is a finite (birooted) subtree of $\Gamma(X)$ with initial vertex 1 and terminal vertex $r(u)$ (the reduced form of u in $\mathrm{FG}(X)$). $\mathrm{MT}(u)$ is referred to as the *Munn tree* of u. For example, if $u = abbb^{-1}aa^{-1}b^{-1}a^{-1}abb^{-1}$, then $\mathrm{MT}(u)$ is the finite tree shown below (with initial vertex indicated by and terminal vertex indicated by):



The following version of a theorem of Munn [8] provides a solution to the word problem for $\mathrm{FIM}(X)$.

Corollary 1.2. *Let X be a nonempty set and ρ the Wagner congruence on $(X \cup X^{-1})^*$. Then for words $u, v \in (X \cup X^{-1})^*$ we have $u \,\rho\, v$ if and only if $\mathrm{MT}(u) = \mathrm{MT}(v)$ and $r(u) = r(v)$.*

It is possible to view the Munn tree of a word $u \in (X \cup X^{-1})^*$ in several slightly different ways. In particular, we may view it as an *automaton* with vertices corresponding to the states of the automaton, and edges corresponding to transitions. There is one initial state (the initial vertex corresponding to 1) and one terminal state (the terminal vertex corresponding to $r(u)$). The *language* accepted by this automaton is $L(u) = \{w \in (X \cup X^{-1})^* : w$ labels a path from 1 to $r(u)$ in $\mathrm{MT}(u)\}$. This has a natural interpretation relative to the free inverse monoid, namely $L(u) = \{w \in (X \cup X^{-1})^* : w\rho \geq u\rho$ in the natural partial order on $\mathrm{FIM}(X)\}$. We may also view $\mathrm{MT}(u)$ as the graph associated with the Schützenberger representation of $\mathrm{FIM}(X)$ relative to the $\mathscr{R}$-class $R_{u\rho}$. That is, the vertices of $\mathrm{MT}(u)$ are in one-to-one correspondence with the elements of $R_{u\rho}$ (with 1 corresponding to $(uu^{-1})\rho$ and $r(u)$ corresponding to $u\rho$) and the edges are of the form

$$v\rho \xrightarrow{\;x\;} (vx)\rho$$

with $v\rho$, $(vx)\rho \in R_{u\rho}$. This point of view is very useful and has been extended greatly by Stephen [15] to study arbitrary presentations of inverse monoids.

Let X be a nonempty set and let $T = \{(u_i, v_i) : i \in I\}$ be a relation on $(X \cup X^{-1})^*$; i.e., u_i, v_i $(X \cup X^{-1})^*$ for each $i \in I$. We define the inverse monoid presented by the set X of generators and the set T of relations to be the inverse monoid $M = \mathrm{Inv}\langle X : T\rangle = (X \cup X^{-1})^*/\tau$ where τ is the congruence on $(X \cup X^{-1})^*$ generated by $\rho \cup T$. We sometimes also abuse notation slightly and regard the u_i, v_i as elements of $\mathrm{FIM}(X)$ and think of $\mathrm{Inv}\langle X : T\rangle$ as the inverse monoid $\mathrm{Inv}\langle X : T\rangle \cong \mathrm{FIM}(X)/\tau$ (where here we are viewing τ as the congruence on $\mathrm{FIM}(X)$ generated by T). We refer to the pair $P = (X : T)$ as a *presentation* of $M = \mathrm{Inv}\langle X : T\rangle$. If $P_1 = (X : T_1)$ and $P_2 = (X : T_2)$ are two inverse monoid presentations with the same set X of generators, we say that P_1 and P_2 are *equivalent* if T_1 and T_2 induce the same congruence on $\mathrm{FIM}(X)$: clearly this implies that $\mathrm{Inv}\langle X : T_1\rangle \cong \mathrm{Inv}\langle X : T_2\rangle$.

The following simple fact is important in developing an understanding of presentations of inverse monoids.

Lemma 1.3. *If $M = \mathrm{Inv}\langle X : T\rangle$ then $G = GP\langle X : T\rangle$ is isomorphic to M/σ_M; i.e., G is the maximal group homomorphic image of M.*

(For example, $(\mathbb{Z}, +)$ is the maximal group homomorphic image of the bicyclic monoid $B = \mathrm{Inv}\langle a : aa^{-1} = 1\rangle$, $\mathrm{FG}(X)$ is the maximal group homomorphic image of $\mathrm{FIM}(X)$, etc.)

In order to study presentations of inverse monoids we shall associate with each word $u \in (X \cup X^{-1})^*$ an automaton $\mathscr{A}(u)$ that serves as a "canonical form" for the congruence class containing u relative to the given presentation, in the same way as the Munn tree of u serves as a canonical form for u in the free inverse monoid. Let $M = \mathrm{Inv}\langle X : T\rangle = (X \cup X^{-1})^*/\tau$ be an inverse monoid presentation and let u be a word in $(X \cup X^{-1})^*$ (so that $u\tau$ is the image of u in the inverse monoid M). The *Schützenberger graph* $S\Gamma(X ; T ; u)$ (usually denoted by $S\Gamma(u)$ if the presentation is understood) is the labelled graph defined as follows. Its *vertices* are the elements of M that are related via Green's $\mathscr{R}$-relation to $u\tau$ in the monoid M (i.e., $V(S\Gamma(u)) = R_{u\tau}$ in M). There is a *labelled edge* $(v\tau, x, w\tau)$ in $S\Gamma(u)$ whenever $v, w \in (X \cup X^{-1})^*$, $v\tau, w\tau \in R_{u\tau}$ and $w\tau = (vx)\tau$ for some $x \in X \cup X^{-1}$. It is easy to see that this also forces $(wx^{-1})\tau = v\tau$, so there is also an edge $(w\tau, x^{-1}, v\tau)$ in $S\Gamma(u)$. The pair of edges $(v\tau, x, w\tau)$ and $(w\tau, x^{-1}, v\tau)$ is usually represented by a segment

$$v\tau \xrightarrow{\;x\;} w\tau$$

in a sketch of $S\Gamma(u)$. Notice that $S\Gamma(u)$ is the graph of the Schützenberger representation of M relative to the $\mathscr{R}$-class $R_{u\tau}$. The triple $\mathscr{A}(u) = \mathscr{A}(X ; T ; u) = ((uu^{-1})\tau, S\Gamma(u), u\tau)$ may be regarded as an automaton over the alphabet $X \cup X^{-1}$ with set $R_{u\tau}$ of states, initial state $(uu^{-1})\tau$, terminal state $u\tau$, and transitions corresponding to the edges in $S\Gamma(u)$. Clearly $\mathscr{A}(u)$ is an *inverse* automaton (i.e., each $x \in X \cup X^{-1}$ induces a partial one-one function on the states of $\mathscr{A}(u)$, and x^{-1} induces the inverse partial one-one function). Notice that $\mathscr{A}(u) = (1, \mathrm{MT}(u), r(u))$ if $T = \varnothing$ (i.e., if $M = \mathrm{FIM}(X)$). The automata $\mathscr{A}(u)$ have the following important properties, which are not difficult to verify. The reader is referred to Hopcroft and Ullman [4] for undefined notation.

Theorem 1.4 (Stephen [15]). *For all words $u \in (X \cup X^{-1})^*$ we have the following:*

(1) *$\mathscr{A}(u) = \mathscr{A}(X; T; u)$ is deterministic, injective, and trim.*
(2) *The language accepted by $\mathscr{A}(u)$ is $L(\mathscr{A}(u)) = u\uparrow = \{w \in (X \cup X^{-1})^* : w\tau \geq u\tau$ in the natural partial order on $M\}$.*
(3) *$\mathscr{A}(u)$ is the minimal automaton of $u\uparrow$.*
(4) *$u\tau \mathscr{R} v\tau$ if and only if $u \in L(\mathscr{A}(u))$ and $v \in L(\mathscr{A}(u))$.*

In view of this result, we regard $\mathscr{A}(u)$ as a "canonical form" for the τ-class $u\tau$. Thus we can solve the word problem for the monoid $M = \mathrm{Inv}\langle X : T\rangle$ if we can devise an algorithm that will test, for all words $u, v \in (X \cup X^{-1})^*$, whether $v \in \mathscr{A}(u)$ or not. In his paper [12], Stephen provides an iterative technique for constructing the automaton $\mathscr{A}(u)$. The idea basically is to start with $(1, \mathrm{MT}(u), r(u))$ and successively apply "expansions" and "reductions" to intermediate automata, thus building a sequence of injective, deterministic, trim $X \cup X^{-1}$-automata

$$\mathscr{A}_0(u) = (1, \mathrm{MT}(u), r(u)), \mathscr{A}_1(u), \mathscr{A}_2(u), \ldots, \mathscr{A}_n(u), \ldots$$

with $L(\mathscr{A}_i(u)) \subseteq L(\mathscr{A}_{i+1}(u)) \subseteq u\uparrow$ (for all i) and $\bigcup_{n=0}^{\infty} L(\mathscr{A}_n(u)) = u\uparrow$. Briefly an "expansion" consists of adding to an automaton a new path labelled by one side (v_i, say) of one of the relations $u_i = v_i$ in T when there is already a path in the automaton labelled by the other side (u_i). A "reduction" consists of identifying two edges with the same label and the same initial vertex (a "folding" in the sense of Stallings [14]). These ideas are discussed in detail in Stephen's paper [15], so we will not repeat the details here.

If $P = (X : T)$ is a presentation of an inverse monoid $M = \mathrm{Inv}\langle X : T\rangle = (X \cup X^{-1})^*/\tau$, then for each $u \in (X \cup X^{-1})^*$, the natural homomorphism σ from M onto its maximal group homomorphic image $G = gp\langle X : T\rangle$ induces a graph morphism (again denoted by σ) from $S\Gamma(X; t; u)$ into $\Gamma(X : T)$, the Cayley graph of the corresponding group presentation $(X : T)$. The morphism σ simply maps the edge

$$v\tau \xrightarrow{\;x\;} (vx)\tau$$

of $S\Gamma(X; T; u)$ to the corresponding edge

$$(v\tau)\sigma \xrightarrow{\;x\;} ((vx)\tau)\sigma$$

of $\Gamma(X; T)$. Note that the idempotent $(uu^{-1})\tau$ in $R_{u\tau}$ maps to the vertex 1 of $\Gamma(X : T)$ under the morphism σ. It is not difficult to observe the following fact:

Lemma 1.5. *Let $P = (X : T)$ be a presentation of an inverse monoid $M = \mathrm{Inv}\langle X : T\rangle = (X \cup X^{-1})^*/\tau$. Then the natural morphism*

$$\sigma : M \twoheadrightarrow G = gp\langle X : T\rangle$$

induces a graph embedding of each Schützenberger graph $S\Gamma(X\,;T\,;u)$ *(for* $u \in (X \cup X^{-1})^*$*) into* $\Gamma(X\,;T)$ *if and only if* M *is* E*-unitary.*

Proof. This is simply a consequence of the well-known fact that an inverse monoid M is E-unitary if and only if σ induces an embedding of each $\mathscr{R}$-class of M into the maximal group homomorphic image G of M.

Our concern in this paper is with inverse monoids of the form $M = \mathrm{Inv}\langle X : e_i = f_i,\ i \in I\rangle$ where e_i, f_i are idempotents of $\mathrm{FIM}(X)$; that is, e_i and f_i are Dyck words in $(X \cup X^{-1})^*$, i.e. the reduced form of $e_i(f_i)$ (in the usual group-theoretic sense) is 1. We may view $\mathrm{FIM}(X)$ itself as an example of such a presentation (take $I = \varnothing$ or take $e_i = f_i = 1\ \ \forall i \in I$). Other standard examples include the bicyclic monoid $B = \mathrm{Inv}\langle a\colon aa^{-1} = 1\rangle$ or more generally the inverse monoids of the form

$$M = \mathrm{Inv}\langle x_1, \dots, x_n\colon x_i x_i^{-1} = 1,\ x_j^{-1}x_j = 1,\\ i = 1, \dots, n,\ j = 1, \dots, k\rangle$$

that arise naturally in connection with the (generalized) Dyck languages (see, for example, Berstel [2] for a study of these languages). Clearly the free group $\mathrm{FG}(X)$ may also be presented in this form, namely

$$\mathrm{FG}(X) = \mathrm{Inv}\langle X\colon xx^{-1} = x^{-1}x = 1\ \forall x \in X\rangle.$$

We first need a preliminary lemma that relates these presentations to subtrees of $\Gamma(X)$. We recall that the *trace* of a congruence θ on an inverse monoid M is the equivalence relation $\mathrm{tr}(\theta) = \theta|_{E(M)\times E(M)}$ (i.e., $\mathrm{tr}(\theta)$ is the restriction of θ to the semilattice $E(M)$ of idempotents of M): the *kernel* of θ is defined to be $\ker\theta = \{a \in M\colon a\theta a^2\}$. It is well known (Petrich [9]) that every congruence θ on an inverse monoid M is uniquely determined by its trace and its kernel.

In addition, if θ is an arbitrary congruence on an inverse monoid M then there is a smallest congruence $\theta_{\min}$ on M with $\mathrm{tr}(\theta_{\min}) = \mathrm{tr}(\theta)$. If θ is a congruence on $\mathrm{FIM}(X)$ then the natural morphism $\theta^\#$ from $\mathrm{FIM}(X)$ onto $\mathrm{FIM}(X)/\theta$ factors according to the diagram

$$\begin{tikzcd}
\mathrm{FIM}(X) \arrow[r, "\theta_1"] \arrow[dr, "\theta^\#"'] & \mathrm{FIM}(X)/\theta_{\min} \arrow[d, "\theta_2"] \\
 & \mathrm{FIM}(X)/\theta
\end{tikzcd}$$

where θ_1 is idempotent-pure and θ_2 is idempotent-separating (i.e., if $e_1\theta_2 e_2$ for some idempotents e_1 and e_2, then $e_1 = e_2$). Idempotent-separating morphisms are well studied, and behave very much like morphisms between groups. It is thus clearly of interest to study idempotent-pure images of $\mathrm{FIM}(X)$. The next lemma relates these to the class of presentations under consideration in this paper.

Lemma 1.6. *Let* $P = (X : T)$ *be a presentation of an inverse monoid* $M = \mathrm{Inv}\langle X : T\rangle = \mathrm{FIM}(X)/\theta = (X \cup X^{-1})^*/\tau$. *The following are equivalent*:

(a) P *is equivalent to a presentation of the form* $P_1 = (X : T_1)$ *where* $T_1 = \{(e_i, f_i)\colon i \in I\}$ *for some set* I *and idempotents* e_i, f_i *of* $\mathrm{FIM}(X)$.

(b) θ *is an idempotent-pure congruence on* $\mathrm{FIM}(X)$.

(c) *Each Schützenberger graph* $S\Gamma(X; T; u)$ (*for* $u \in (X \cup X^{-1})^*$) *is a* (*labelled*) *tree.*

Proof. Suppose first that $M \cong \mathrm{Inv}\langle X \colon e_i = f_i, i \in I\rangle$ for some idempotents e_i, f_i of $\mathrm{FIM}(X)$. Let θ_1 be the congruence on $\mathrm{FIM}(X)$ generated by $T_1 = \{(e_i, f_i) \colon i \in I\}$. It is evident that $\theta_1 \subseteq \sigma$, the minimum group congruence on $\mathrm{FIM}(X)$, so θ_1 is an idempotent-pure congruence on $\mathrm{FIM}(X)$ since $\mathrm{FIM}(X)$ is E-unitary (see Petrich [9]).

Suppose next that $M \cong \mathrm{FIM}(X)/\theta$ for some idempotent-pure congruence θ on $\mathrm{FIM}(X)$. Since $\theta \subseteq \sigma$, the natural homomorphism from $\mathrm{FIM}(X)$ onto $\mathrm{FG}(X)$ factors through M. Then M has the same maximal group homomorphic image as $\mathrm{FIM}(X)$, and M is E-unitary since $\mathrm{FIM}(X)$ is E-unitary. By Lemma 1.5 it follows that the natural homomorphism from M onto $\mathrm{FG}(X)$ induces a graph embedding of each Schützenberger graph $S\Gamma(X; T; u)$ (for $u \in (X \cup X^{-1})^*$) into $\Gamma(X)$, the Cayley graph of $\mathrm{FG}(X)$, so $\mathrm{SG}(X; T; u)$ is a (labelled) tree since $\Gamma(X)$ is a (labelled) tree.

Now suppose that each Schützenberger graph $S\Gamma(X; T; u)$ relative to the presentation $P = (X : T)$ is a labelled tree. Corresponding to each vertex α of $S\Gamma(X : T : u)$ there is a (unique) geodesic path from $(uu^{-1})\tau$ to α: let $w(\alpha)$ be the word in $(X \cup X^{-1})^*$ that labels the geodesic from $(uu^{-1})\tau$ to α. Clearly $w(\alpha)$ is a reduced word in the usual group-theoretic sense. The map $\phi \colon S\Gamma(X; T; u) \to \Gamma(X)$ that maps each vertex α of $S\Gamma(X; T; u)$ to $w(\alpha)$ and the edge

$$\underset{\alpha}{\circ} \xrightarrow{\;x\;} \underset{\alpha x}{\circ}$$

(for $x \in X \cup X^{-1}$) to

$$\underset{w(\alpha)}{\circ} \xrightarrow{\;\;x\;\;} \underset{w(\alpha x)}{\circ}$$

is clearly a graph embedding of $S\Gamma(X; T; u)$ into $\Gamma(X)$ that maps $(uu^{-1})\tau$ to 1. Recall also that for each $u \in (X \cup X^{-1})^*$, the Munn tree $MT(u)$ is a (labelled) subtree of $\Gamma(X)$. It is evident that the image under ϕ of the path in $S\Gamma(X; T; u)$ from $(uu^{-1})\tau$ to $u\tau$ labelled by the word $u \in (X \cup X^{-1})^*$ is in fact the Munn tree $\mathrm{MT}(u)$, so we may regard $\mathrm{MT}(u)$ as a subtree of $S\Gamma(X; T; u)\phi$ and the initial (resp. terminal) root of $\mathrm{MT}(u)$ coincides with that of $S\Gamma(X; T; u)\phi$. Now let e be an idempotent of $\mathrm{FIM}(X)$ and suppose that $e\ \theta\ u$ for some $u \in \mathrm{FIM}(X)$, where θ is the congruence on $\mathrm{FIM}(X)$ generated by R. Since $e\ \theta\ u$, the graphs $S\Gamma(X : T : e)\phi$ and $S\Gamma(X; T; u)\phi$ coincide and their terminal roots coincide. But the terminal root of $S\Gamma(X; T; e)\phi$ coincides with the terminal root of $\mathrm{MT}(e)$ and the terminal root of $S\Gamma(S; T; u)\phi$ coincides with the terminal root of $\mathrm{MT}(u)$. Hence the terminal roots of $\mathrm{MT}(u)$ and $\mathrm{MT}(e)$ coincide. Since e is an idempotent the terminal root of $\mathrm{MT}(e)$ is 1 and so the terminal root of $\mathrm{MT}(u)$ is 1, so u is also an idempotent of $\mathrm{FIM}(X)$. It follows that θ is an idempotent-pure congruence on $\mathrm{FIM}(X)$.

Finally, suppose that $M \cong \mathrm{FIM}(X)/\theta$ for some idempotent-pure congruence θ on $\mathrm{FIM}(X)$. Let T_1 be the set of pairs of the form (e, f) where e is an idempotent of $\mathrm{FIM}(X)$ and $f\ \theta\ e$: clearly we also have $f^2 = f$ in $\mathrm{FIM}(X)$. Let υ be the congruence on $\mathrm{FIM}(X)$ generated by T_1. By the argument given in the first paragraph of the proof, υ is idempotent-pure, so $\mathrm{Ker}\,\upsilon = \mathrm{Ker}\,\theta$, where $\operatorname{Ker}\upsilon$ (resp. $\operatorname{Ker}\theta$) denotes the set of elements of $\mathrm{FIM}(X)$ that are υ (resp. θ) related to an idempotent of $\mathrm{FIM}(X)$. Clearly υ and θ have the same restriction to the set of idempotents of $\mathrm{FIM}(X)$ since $\upsilon \subseteq \theta$, so υ and θ have the same trace. It follows that $\upsilon = \theta$ and hence M has a presentation of the desired form.

Corollary 1.7. *Every inverse monoid* $M = \mathrm{Inv}\langle X : e_i = f_i,\ i \in I\rangle$ *(where* e_i, f_i *are idempotents of* $\mathrm{FIM}(X)$*) is E-unitary with maximal group image* $\mathrm{FG}(X)$.

In order to provide an effective construction of the Schützenberger graphs corresponding to inverse monoid presentations of the form $M = \mathrm{Inv}\langle X : e_i = f_i,\ i = 1, \ldots, n\rangle$ (for e_i, f_i idempotents of $\mathrm{FIM}(X)$ and X finite) we shall need to make use of some results from logic and formal language theory. We provide a brief discussion of these results in the next section.

2. Second order monadic logic and rational sets

In this section we review material from logic and formal language theory that we need for our solution to the word problem. We will be necessarily brief and we refer the reader to standard references on logic and language theory for more details.

2.1. **Rational sets.** For proofs of theorems in this subsection see Berstel [2].

Let M be a monoid. The set $\mathrm{Rat}(M)$ of *rational subsets* of M is the smallest collection of subsets of M containing the singleton sets and closed under finite union, product of subsets and submonoid generation (usually referred to as the "star" operation). A subset L of M is *recognizable* if $L = P\eta^{-1}$ where $\eta : M \to N$ is a morphism into some finite monoid N and $P \subseteq N$. Let $\mathrm{Rec}(M)$ denote the recognizable subsets of M. The following is a fundamental result of the theory of automata and formal languages.

Theorem 2.1 (Kleene's Theorem). *Let* M *be a finitely generated free monoid. Then* $\mathrm{Rec}(M) = \mathrm{Rat}(M)$.

In general, $\mathrm{Rec}(M) \neq \mathrm{Rat}(M)$. An interesting case occurs when the monoid is a group.

Theorem 2.2. *Let* G *be a group and let* H *be a subgroup of* G. *Then* $H \in \mathrm{Rec}(M)$ *iff* $[G : H] < \infty$, *and* $H \in \mathrm{Rat}(M)$ *iff* H *is finitely generated.*

Our main concern in this paper is with the rational subsets of $\mathrm{FG}(X)$, the free group on a finite set X. Let $\nu : (X \cup X^{-1})^* \twoheadrightarrow \mathrm{FG}(X)$ be the canonical morphism from the free monoid on $X \cup X^{-1}$ to $\mathrm{FG}(X)$. Here, as usual, X^{-1} is a set in bijection with and disjoint from X. There is also a function (not a morphism!) $\iota : \mathrm{FG}(X) \to (X \cup X^{-1})^*$ that assigns $g \in \mathrm{FG}(X)$ to the unique reduced word in $(X \cup X^{-1})^*$ representing g. Note that if $u \in (X \cup X^{-1})^*$ then $u\nu\iota = r(u)$, the reduced from of u in the usual sense.

Theorem 2.3 (Benois). *Let* $L \subseteq \mathrm{FG}(X)$. *Then* $L \in \mathrm{Rat}(\mathrm{FG}(X))$ *iff* $L\iota \in \mathrm{Rat}((X \cup X^{-1})^*)$.

Let $L \in \mathrm{Rat}(\mathrm{FG}(X))$. Since $L\iota \in \mathrm{Rat}(X \cup X^{-1})^*$, it follows from Kleene's Theorem that there is a finite state automaton over $X \cup X^{-1}$ that recognizes $L\iota$. We define $\mathscr{B}(L)$ to be the minimal automaton of $L\iota$.

We need an effective version of Benois' Theorem for our intended applications. We would like to effectively compute $\mathscr{B}(L)$ from a description of $L \subseteq \mathrm{Rat}(\mathrm{FG}(X))$ by a rational expression. This follows from the proof of Lemma 2.11 of Berstel [2], but for completeness we include a version of that lemma here. If $L \subseteq (X \cup X^{-1})^*$ let $\overline{L} = L\nu\iota \subseteq (X \cup X^{-1})^*$, the set of reductions of words in L, i.e., $\overline{L} = \{r(u) \colon u \in L\}$.

Lemma 2.4. *Let $L \subseteq (X \cup X^{-1})^*$. If L is recognizable, then so is $\overline{L}$. Furthermore, an automaton recognizing $\overline{L}$ can be effectively constructed from an automaton recognizing L.*

Proof. Let $\mathscr{A} = (Q, X \cup X^{-1}, \delta, i, T)$ be an automaton recognizing L. Recall (Hopcroft and Ullman [4]) that this notation means Q is a (finite) set of states, $X \cup X^{-1}$ the alphabet, δ the next state function, $i \in Q$, the initial state, and $T \subseteq Q$, the set of terminal states.) Let $D_X = 1\nu^{-1}$, the (two-sided) Dyck set over X, consisting of words in $(X \cup X^{-1})^*$ equal to 1 in the free group.

For $p, q \in Q$, let $A_{p,q} = \{w \in (X \cup X^{-1})^* | \delta(p, w) = q\}$. $A_{p,q}$ is thus an effectively constructible recognizable language.

It is well known that D_X is a context-free language. It follows that xD_X and $D_X x D_X$ are context-free languages for every $x \in X \cup X^{-1}$. Furthermore, we can assume by standard results of formal language theory that we have effectively constructible push-down automata for each of the languages xD_X and $D_X x D_X$, $x \in X \cup X^{-1}$.

Now consider the nondeterministic finite state automaton $\mathscr{B} = (Q \cup \{s\}, X \cup X^{-1}, \delta', s, T')$ where $s \notin Q$ and δ' is given by

$$q \in \delta'(p, x) \quad \text{iff} \quad xD_X \cap A_{p,q} \neq \varnothing, \qquad p, q \in Q,$$

and

$$q \in \delta'(s, x) \quad \text{iff} \quad D_X x D_X \cap A_{i,q} \neq \varnothing.$$

Also

$$T' = \begin{cases} T & \text{if } D_X \cap L \neq \varnothing, \\ T \cup \{s\} & \text{if } D_X \cap L = \varnothing. \end{cases}$$

We claim that $\mathscr{B}$ is effectively constructible. This follows from two important theorems of formal language theory (see Hopcroft and Ullman [4]). The first states that the intersection of a context free language L_1 and a recognizable set R is again context-free. Furthermore, a push-down automaton (p.d.a.) for $L_1 \cap R$ can be effectively constructed from a p.d.a. recognizing L_1 and a finite state automaton recognizing R. The second theorem states that it is decidable whether the language accepted by a p.d.a. is empty or not.

It is easy to see that the language L' accepted by $\mathscr{B}$ is given by

$$L' = \{w \in (X \cup X^{-1})^* | w = x_1 \cdots x_n,\ x_i \in X \cup X^{-1},\ 1 \leq i \leq n \text{ and } \exists d_0, \ldots, d_n \in D_X \text{ such that } d_0 x_1 d_1 x_2 \cdots x_n d_n \in L\}.$$

Therefore L' is recognizable. Finally, the set R of reduced words is a recognizable language and clearly $\overline{L} = L' \cap R$. Since we can effectively construct an automaton for L' and an automaton for R, we can effectively construct an automaton for their intersection $\overline{L}$.

Corollary 2.5. *Let $L \in \mathrm{Rat}(\mathrm{FG}(X))$ be given by a rational expression. Then we can effectively construct the automaton $\mathscr{B}(L)$.*

Proof. Let $L \in \mathrm{Rat}(\mathrm{FG}(X))$. By definition, $\mathscr{B}(L)$ is the minimal automaton of $L\iota \subseteq (X \cup X^{-1})^*$. Since L is given by a rational expression R we can effectively construct an automaton $\mathscr{A}$ accepting the language L_1 defined by R considered as a language over $X \cup X^{-1}$. It is clear that $L_1\nu = L$ and thus $L\iota = L_1\nu\iota = \overline{L}_1$. By the lemma above we can effectively construct an automaton $\mathscr{B}$ recognizing $\overline{L}_1$ and thus we can effectively construct $\mathscr{B}(L)$ by using standard results of the theory of automata.

2.2. **Second-order monadic logic.** We assume some familarity with basic definitions and ideas of (first-order) logic. See, for example, Barwise [1].

In second-order monadic logic, quantifiers refer to sets (i.e. unary or monadic predicates) as well as to individual members of a structure. We review the basic definitions.

Let $M = (A, \{R_i | i \in I\}, \{f_j | j \in J\})$ be a structure. Thus A is a nonempty set, each R_i, $i \in I$, is an n_i-ary relation on A for some $n_i > 0$ and each f_j, $j \in J$, is an m_j-ary function for some $m_j \geq 0$: 0-ary functions are interpreted as constants.

The second-order monadic language $\mathscr{L}$ appropriate for M consists of the following data:

Individual variables— $\{x_n | n \in \mathbb{N}\}$;

Set variables— $\{X_n | n \in \mathbb{N}\}$;

Predicate symbols— $\{R_i | i \in I\}$;

Function symbols— $\{f_j | j \in J\}$;

Logical symbols— $\{\wedge, \vee, \neg, (,), =, \forall, \in\}$ (or any other complete set of connectives).

Thus there is a one-to-one correspondence between function (predicate) symbols of $\mathscr{L}$ and functions (relations) in M. We assume that each function (predicate) has the same arity as the corresponding function (predicate) of M.

The syntax and semantics of terms and well-formed formulae are defined inductively in the usual manner. Atomic formulae include those of the form $t \in X$ where t is a term and X is a set variable. A sentence of the form $\forall X \phi(X)$ where X is a set variable, in particular, is true in M iff $\phi(Y)$ is (inductively) true in M for each $Y \subseteq M$. If a sentence ϕ is true in M we write $M \vDash \phi$ and we define $\mathrm{Th}_2(M) = \{\phi | M \vDash \phi\}$. The (second-order monadic) theory of M is *decidable* if there is an algorithm that tests whether a given sentence ϕ of $\mathscr{L}$ is in $\mathrm{Th}_2(M)$ or not.

We will be interested in two particular structures and their associated theories in this paper.

Let A be a countable set and consider the structure $T_A = (A^*, \{r_a | a \in A\}, \leq)$. Here $r_a \colon A^* \to A^*$ is right multiplication by a, $xr_a = xa$ $\forall x \in A^*$ and $\leq$ is the prefix order $x \leq y$ iff $\exists u \in A^*$, $xu = y$. We call $\mathrm{Th}_2(T_A)$ the theory of A-successor functions. We can now state Rabin's Tree Theorem.

Theorem 2.6 (Rabin [10]). $\mathrm{Th}_2(T_A)$ *is decidable.*

The terminology "Tree Theorem" comes from the usual representation of A^* as a labelled rooted tree. The root is labelled by the empty word, and a node labelled by $x \in A^*$ has a descendant labelled xa for each $a \in A$. The Tree Theorem is one of the deepest decidability results known. The decidability of a number of other theories can be reduced to $\mathrm{Th}_2(T_A)$ (see Barwise [1]).

The second result of Rabin that we need concerns the definable sets in T_A. We say that a set $L \subseteq A^*$ is *definable* if there is a formula $\phi(X)$ of $\mathscr{L}$ with exactly one free (set) variable X such that $M \vDash \exists! X\phi(X)$ and $M \vDash \phi(L)$.

The following theorem summarizes Theorem 2.1, Theorem 2.3, and Corollary 2.4 of Rabin [11]. See also the remark following Theorem 2.3 of [11].

Theorem 2.7 (Rabin [11]). *$L \subseteq A^*$ is definable iff L is rational. Furthermore, if $\exists X\phi(X)$ is true in T_A, then we can effectively find a rational set L such that $T_A \vDash \phi(L)$.*

We now introduce our second structure of interest. Let $G_X = (\mathrm{FG}(X), \{\sigma_x | x \in X \cup X^{-1}\})$ where $\mathrm{FG}(X)$ is the free group on a (countable) set X and if $x \in X \cup X^{-1}$, $\sigma_x \colon \mathrm{FG}(X) \to \mathrm{FG}(X)$ is given by $g\sigma_x = gx$. Notice here that this is right multiplication in $\mathrm{FG}(X)$, *not* concatenation in $(X \cup X^{-1})^*$. We wish to show also that $\mathrm{Th}_2(G_X)$ is decidable. We do this by reducing $\mathrm{Th}_2(G_X)$ to $\mathrm{Th}_2(T(X \cup X^{-1})^*)$. We will, for each sentence ϕ of G_X, give an (effectively constructible) sentence $\tilde{\phi}$ of $T_{X\cup X^{-1}}$ such that $G_X \vDash \phi$ iff $T_{X\cup X^{-1}} \vDash \tilde{\phi}$. We do this by the method of (semantic) interpretation (Barwise [1]). The idea is to represent $\mathrm{FG}(X)$ by (the $\mathscr{L}_2$ definable) set of reduced words considered as a subset of $(X \cup X^{-1})^*$ and to define $\sigma_x \colon \mathrm{FG}(X) \to \mathrm{FG}(X)$ by an $\mathscr{L}_2$ definable relation in $(X \cup X^{-1})^*$: $\tilde{\phi}$ can then be defined in a natural way inductively from the structure of ϕ as a well-formed formula. Again see Barwise [1] for details. We mention that the decidability of $\mathrm{Th}_2(G_X)$ follows also from the work of Müller and Schupp [7], but we include a proof here for the sake of completeness.

Theorem 2.8. $\mathrm{Th}_2(G_X)$ *is decidable.*

Proof. Consider the formula $\phi(w)$ of $\mathscr{L}_2(T_{X\cup X^{-1}})$ given by

$$\bigwedge_{x\in X\cup X^{-1}} \forall v(v \le w \wedge \exists u\, ur_x = v) \Rightarrow \neg\exists z(zr_{x^{-1}} = u)).$$

This statement says that no prefix of w ends with the word $x^{-1}x$, $x \in X \cup X^{-1}$. Clearly, $T_{X\cup X^{-1}} \vDash \phi(g)$ with $g \in (X \cup X^{-1})^*$ iff g is in R, the set of reduced words over $X \cup X^{-1}$ in the sense of group theory.

Now consider for each $x \in (X \cup X^{-1})$ the formula $\psi_x(v, w)$ defined by

$$\phi(v) \wedge \phi(w) \wedge (\exists z(zr_{x^{-1}} = v \Rightarrow z = w) \vee vr_x = w).$$

It is easy to see that a pair (g, h), $g, h \in (X \cup X^{-1})^*$, has $T_{X\cup X^{-1}} \vDash \psi_x(g, h)$ iff g and h are both in R and $g\sigma_x = h$ in $\mathrm{FG}(X)$.

Now given a sentence α of $\mathscr{L}_2(G_X)$ we define a formula $\tilde{\alpha}$ of $\mathscr{L}_2(T_{X\cup X^{-1}})$ by relativizing quantifiers to R and replacing terms involving σ_x by ψ_x. Clearly $G_X \vDash \alpha$ iff $T_{X\cup X^{-1}} \vDash \tilde{\alpha}$ and thus $\mathrm{Th}_2(G_X)$ is reduced to $\mathrm{Th}_2(T_{X\cup X^{-1}})$ and is thus decidable by Rabin's Tree Theorem.

Corollary 2.9. *Let $L \subseteq \mathrm{FG}(X)$. If L is definable then L is rational.*

Proof. Let L be defined by the $\mathscr{L}_2(G_X)$ sentence $\psi = \exists! Y\phi(Y)$. Then the sentence $\tilde{\psi}$ constructed in Theorem 2.8 is $\exists! Y\tilde{\phi}(Y)$. It follows that $T_{X\cup X^{-1}} \vDash \tilde{\psi}$ and $T_{X\cup X^{-1}} \vDash L\iota$. Therefore $L\iota$ is rational by Theorem 2.7 and thus L is rational by Benois' theorem (Theorem 2.3).

3. The word problem

We turn now to a study of the word problem for an inverse monoid presentation of the form

$$(*) \qquad M = \mathrm{Inv}\langle X : e_i = f_i,\ i \in I\rangle$$

where I and X are *finite* sets and e_i, f_i are idempotents of $\mathrm{FIM}(X)$. Note first that if e_i and f_i are idempotents of $\mathrm{FIM}(X)$ then $e_i = f_i$ in M if and only if $e_i = e_i f_i$ and $f_i = e_i f_i$ in M. Since $e_i f_i \le e_i, f_i$ in the natural partial order on $\mathrm{FIM}(X)$, it is evident that each presentation of the form considered above is equivalent to one of the form $M = \mathrm{Inv}\langle X : e_i = f_i,\ i \in I\rangle$ where e_i, f_i are idempotents of $\mathrm{FIM}(X)$ and $e_i < f_i$ in the natural partial order on $\mathrm{FIM}(X)$. Consequently, in this section we shall study presentations of the form $(*)$ with $e_i < f_i$ in $\mathrm{FIM}(X)$. It follows that $\mathrm{MT}(f_i)$ is a subtree of $\mathrm{MT}(e_i)$, both being viewed as subtrees of $\Gamma(X)$. We allow as a special case that some (or all) of the f_i may be 1 (the identity of $\mathrm{FIM}(X)$) or that I may be empty, in which case M is $\mathrm{FIM}(X)$. For each word $u \in (X \cup X^{-1})^*$ we provide an iterative construction of a birooted labelled tree (i.e. automaton) $B\Gamma(u)$ that recognizes the language $u\uparrow$; that is, we provide an iterative construction of the Schützenberger automaton $\mathscr{A}(u)$ relative to this presentation. The iterative construction is essentially that of Stephen [15], the only distinction being that at each stage of the iteration we map the corresponding iterate into the tree $\Gamma(X)$. Thus we iteratively construct, inside the tree $\Gamma(X)$, the image of $\mathscr{A}(u)$ induced by the natural map $\sigma : M \to \mathrm{FG}(X)$. The iterates are constructed as follows.

Let $\Gamma_1(u) = \mathrm{MT}(u)$. Then $\Gamma_1(u)$ may be considered as a birooted subtree of $\Gamma(X)$, the roots being 1 (initial) and $r(u)$ (terminal). Note that since $\Gamma_1(u)$ is a subtree of $\Gamma(X)$ containing the vertex 1, the set V_1 of vertices of $\Gamma_1(u)$ is a *Schreier* subset of $\mathrm{FG}(X)$, that is, each element of V_1 may be considered as a reduced word in $\mathrm{FG}(X)$ and if $v = x_1 \cdots x_k$ is a reduced word in V_1, then $x_1 \cdots x_i \in V_1$ for all i with $0 \le i \le k$. The approximate $\Gamma_n(u)$ is a birooted subtree of $\Gamma(X)$ (with initial root 1 and terminal $r(u)$) constructed by induction from $\Gamma_{n-1}(u)$ in the following way. Let V_{n-1} denote the Schreier subset of $\mathrm{FG}(X)$ consisting of the vertices of the tree $\Gamma_{n-1}(u)$. For each $i \in I$ let E_i denote the set of vertices of $\mathrm{MT}(e_i)$ and let F_i denote the set of vertices of $\mathrm{MT}(f_i)$: thus E_i and F_i are finite Schreier subsets of $\mathrm{FG}(X)$ with $F_i \subseteq E_i$ for each $i \in I$. For each $i \in I$ let $G_i = \{v \in V_{n-1} : v \cdot F_i \subseteq V_{n-1}\}$ where $v \cdot F_i$ denotes the set of all reduced forms of the words $v \cdot x (x \in F_i)$ in $\mathrm{FG}(X)$: also let $H_i = G_i E_i$ (the set of all reduced forms of the word $v \cdot v_1$ with $v \in G_i$ and $v_1 \in E_i$). Then define

$$V_n = V_{n-1} \cup \left(\bigcup_{i \in I} G_i E_i \right).$$

It is easy to see by induction that each set V_n is a (finite) Schreier subset of $\mathrm{FG}(X)$ containing the vertices 1 and $r(u)$. Hence V_n serves as the set of vertices of a (uniquely determined) birooted subtree $\Gamma_n(u)$ of $\Gamma(X)$, the roots again being 1 (initial) and $r(u)$ (terminal). Finally define $B\Gamma(u) = \bigcup_{n=1}^{\infty} \Gamma_n(u)$. Clearly $B\Gamma(u)$ is the (possibly infinite) birooted subtree of $\Gamma(X)$ whose set of vertices is $\bigcup_{n=1}^{\infty} V_n$ and whose roots are 1 (initial) and $r(u)$ (terminal). Each approximate $\Gamma_n(u)$ as defined above is the image in $\Gamma(X)$ of the corresponding "full P-expansion" of Stephen [15], and the tree $B\Gamma(u)$ is actually the image in $\Gamma(X)$ of the "basic graph" $B\Gamma(u)$ defined by Stephen [15]. Since the monoid $M = \mathrm{Inv}\langle X : e_i = f_i,\ i \in I\rangle$ is E-unitary (Corollary 1.5), the Schützenberger graph $S\Gamma(u)$ actually *embeds* in $\Gamma(X)$ (Lemma 1.2). It follows from the results of Stephen [15] that the birooted graph $B\Gamma(u)$ constructed above is actually the image of the (birooted) Schützenberger graph $\mathscr{A}(u) = ((uu^{-1})\tau, S\Gamma(u), u\tau)$ under the natural embedding into $\Gamma(X)$. When viewed as automata, $B\Gamma(u)$ and $\mathscr{A}(u)$ may be identified with the minimal automaton recognizing the language $u\uparrow$. We summarize all of this in the following theorem, which is essentially a very special case of some of the results of Stephen [15].

Theorem 3.1. *Let $M = \mathrm{Inv}\langle X : e_i = f_i, i \in I\rangle$, let $u \in (X \cup X^{-1})^*$, and let $B\Gamma(u)$ be the birooted tree (i.e. automaton) constructed above. Then $B\Gamma(u)$ is (isomorphic to) $\mathscr{A}(u)$, the minimal automaton recognizing the language $u\uparrow$.*

We now turn to a solution to the word problem for these presentations. Note that from our definition of the trees $B\Gamma(u)$ $(u \in (X \cup X^{-1})^*)$ it follows that the set $V(B\Gamma(u))$ of vertices of $B\Gamma(u)$ is the smallest subset V of $\mathrm{FG}(X)$ such that (1) $V_1 = V(\mathrm{MT}(u)) \subseteq V$, and (2) $v \cdot E_i \subseteq V$ whenever $v \cdot F_i \subseteq V$ for some $i \in I$ and $v \in V$. (Here, as before, the multiplication is in $\mathrm{FG}(X)$.) It is straightforward to define V in the monadic logic of the structure $G_X = (\mathrm{FG}(X), \{\sigma_x | x \in X \cup X^{-1}\})$ (see §2 above). For completeness, we include these details here.

Consider the formula:

$$\psi(Y) = (V_1 \subseteq Y) \wedge \bigwedge_{i \in I} \forall v (v \cdot F_i \subseteq Y \Rightarrow v \cdot E_i \subseteq Y).$$

Then $L \subseteq \mathrm{FG}(X)$ satisfies ψ iff L satisfies properties (1) and (2) above. It is easy to see that the containment relation "$\subseteq$" is definable in $\mathscr{L}_2(G_X)$. Furthermore, since $V_1, E_i, F_i,\ i \in I$, are all *finite* subsets of $\mathrm{FG}(X)$, (being vertices of Munn trees) we can construct $\mathscr{L}_2$-formulae representing the conditions $V_1 \subseteq Y$, $v \cdot F_i \subseteq Y$, $v \cdot E_i \subseteq Y$, $i \in I$. Thus we can consider $\psi(Y)$ to be an $\mathscr{L}_2$-formula. Finally, if

$$\phi(Y) = \psi(Y) \wedge \forall Z (\psi(Z) \Rightarrow Y \subseteq Z)$$

then $V = V(B\Gamma(u))$ is the unique subset of $\mathrm{FG}(X)$ that satisfies $G_X \models \phi(V)$. That is, $G_X \models \exists! Y \phi(Y)$ and $G_X \models \phi(V)$. These observations lead to the following theorem.

Theorem 3.2. *Let $M = \mathrm{Inv}\langle X : e_i = f_i,\ i = 1, \ldots, n\rangle$ (where e_i, f_i are idempotents of* $\mathrm{FIM}(X)$*). Let $u \in (X \cup X^{-1})^*$. Then $V = V(B\Gamma(u))$ is an effectively constructible rational Schreier subset of* $\mathrm{FG}(X)$*. In particular, the word problem for M is decidable.*

Proof. We have seen that we can effectively construct a formula in $\mathscr{L}_2(G_X)$ defining V. It follows from Theorem 2.7, Theorem 2.8, and Corollary 2.9 that the Schreier set V is an effectively constructible rational subset of $\mathrm{FG}(X)$. Furthermore, $B\Gamma(u) = (1, V(B\Gamma(u)), r(u))$ is the (effectively constructible) minimal automaton of $u\uparrow$. Thus $u = v$ in M iff $u\uparrow = v\uparrow$ iff $V(B\Gamma(u)) =$

$V(B\Gamma(v))$ and $r(u) = r(v)$. Since these last two conditions can be effectively checked, it follows that the word problem for M is decidable.

Remark 3.3. It is possible to provide a much more direct proof of the fact that each set $V = V(B\Gamma(u))$ in the statement of Theorem 3.2 is a rational Schreier subset of $\mathrm{FG}(X)$. However, in general we need Rabin's theorem to guarantee that V is effectively constructible and hence that the word problem is decidable. While Rabin's theorem does provide us with a tool to solve the word problem, this kind of solution is exceedingly complex and it would be desirable to find a much more direct solution to the word problem. In the case where each relation is of the form $e_i = 1$, i.e. in the case where $M = \mathrm{Inv}\langle X : e = 1\rangle$ for some $e = e^2$ in $\mathrm{FIM}(X)$, it is possible to avoid the use of Rabin's theorem. This case is discussed in a later paper [3].

4. Connections with context-free languages

In this section we establish some connections between the inverse monoid presentations discussed above and the theory of context-free languages. We refer to Hopcroft and Ullman [4] and Berstel [2] for basic results and notation concerning context-free languages. In particular we shall view a deterministic pushdown automaton as a 7-tuple $(Q, \Sigma, \Gamma, \delta, q_0, Z_0, F)$ as defined by Hopcroft and Ullman [4, Chapter 5]. We first establish some basic results concerning rational Schreier subsets of the free group. The next result is a generalization of the well-known fact that the Dyck languages (see Berstel [2]) are deterministic context-free languages.

Lemma 4.1. *Let $L \subseteq \mathrm{FG}(X)$ be a rational Schreier subset of $\mathrm{FG}(X)$ and let $\Gamma(L)$ be the subtree of $\Gamma(X)$ determined by L. Then the language $\widetilde{L} = \{w \in (X \cup X^{-1})^* : w$ labels a path from 1 to 1 in $\Gamma(L)\}$ is a deterministic context-free language.*

Proof. Since L is rational, $L\imath$ is a recognizable subset of $(X \cup X^{-1})^*$ by the theorem of Benois (Theorem 2.3), so the minimal automaton $\mathscr{B}(L)$ of $L\imath$ is a finite state automaton. Thus the tree of $L\imath$ contains only a finite number of distinct (directed) rooted subtrees and in fact the states of the minimal automaton $\mathscr{B}(L)$ of $L\imath$ are the isomorphism classes of directed rooted subtrees of the tree of $L\imath$. Let S be the set of states of the automaton $\mathscr{B}(L)$; then, for $i, j \in S$ and $x \in X \cup X^{-1}$, $i \cdot x = j$ in $\mathscr{B}(L)$ if and only if x labels an edge from the root of a tree in the isomorphism class i to the root of a tree in the isomorphism class j. Denote the start state of $\mathscr{B}(L)$ by s_0. Now define the deterministic pushdown automaton $\mathscr{P} = (Q, \Sigma, \Gamma, \delta, q_0, Z_0, F)$ as follows:

$Q = \{p, q\}$ is the set of states of $\mathscr{P}$,
$\Sigma = X \cup X^{-1}$ is the input alphabet of $\mathscr{P}$,
$\Gamma = S \times [X \cup X^{-1} \cup \{Z\}]$ (where Z is a symbol not in $X \cup X^{-1}$) is the stack alphabet of $\mathscr{P}$,
$q_0 = p$ is the initial state of $\mathscr{P}$,
$Z_0 = (s_0, Z)$ is the start stack symbol of $\mathscr{P}$,
$F = \{p\}$ is the set of final states of $\mathscr{P}$, and
δ is a partial map from $Q \times (\Sigma \cup \{\varepsilon\}) \times \Gamma$ to $Q \times \Gamma^*$ defined by

$$\delta(p, x, (s_0, Z)) = \begin{cases} (q, (s_0, Z)(s_0 \cdot x, x)) & \text{if } s_0 \cdot x \text{ is defined in } \mathscr{B}(L), \\ \text{undefined} & \text{otherwise}, \end{cases}$$

$$\delta(q, x, (i, y)) = \begin{cases} (q, (i, y)(i \cdot x, x)) & \text{if } i \cdot x \text{ is defined in } \mathscr{B}(L) \text{ and } \\ & y \neq x^{-1}, \\ (q, 1) & \text{if } y = x^{-1} \text{ (here 1 is the identity} \\ & \text{of } \Gamma^*), \\ \text{undefined} & \text{otherwise}, \end{cases}$$

$$\delta(q, \varepsilon, (x_0, Z)) = (p, (x_0, Z)).$$

Now let w be a word in $(X \cup X^{-1})^*$. For $(q_i, \alpha_i) \in Q \times \Gamma^*$, $i = 1, 2$, we write $(q_1, \alpha_1) \overset{*}{\underset{w}{\vdash}} (q_2, \alpha_2)$ if $(q_1, w, \alpha_1) \overset{*}{\vdash} (q_2, \varepsilon, \alpha_2)$ in the sense of Hopcroft and Ullman [4, §5.2]. By an easy induction on the length of a word $w \in (X \cup X^{-1})^*$, we can see that w labels a path in $\Gamma(L)$ from 1 to u if and only if $(p, (s_0, Z)) \overset{*}{\underset{w}{\vdash}} (q, \alpha)$ where $\alpha = (s_0, Z)(s_1, x_1) \cdots (s_n, x_n)$ with $s_n = s_0 \cdot (r(w))$ and $r(x_1 \cdots x_n) = r(w) = u$. (Here, as before, $r(w)$ is the reduced form of w in the group-theoretic sense.) It follows that $w \in \widetilde{L}$ if and only if $(p, (s_0, Z)) \overset{*}{\underset{w}{\vdash}} (q, (s_0, Z)) \underset{\varepsilon}{\vdash} (p, (s_0, Z))$, so $w \in \widetilde{L}$ if and only if $w \in L(\mathscr{P})$. Hence $\widetilde{L}$ is a deterministic context-free language.

Remark 4.2. If $L = \mathrm{FG}(X)$ and $X = \{x_1, \dots, x_n\}$ then $\widetilde{L}$ is just the Dyck language D_n^* (in the notation of Berstel [2]). If $L = X^*$ (the positive cone of $\mathrm{FG}(X)$) and $X = \{x_1, \dots, x_n\}$ then $\widetilde{L}$ is just the restricted Dyck language $D_n'^*$ (see Berstel [2]). Thus the languages $\widetilde{L}$ may be viewed as generalizations of the Dyck languages discussed in Berstel [2].

Corollary 4.3. *Let L be a rational Schreier subset of* $\mathrm{FG}(X)$ *and let u be a word in $L\iota$. Let $\widetilde{L}_u = \{w \in (X \cup X^{-1})^* \colon w$ labels a path from 1 to u in $\Gamma(L)\}$. Then $\widetilde{L}_u$ is deterministic context-free language.*

Proof. It is clear that w labels a path from 1 to u in $\Gamma(L)$ if and only if wu^{-1} labels a path from 1 to 1 in $\Gamma(L)$. If follows that $\widetilde{L}_u$ is the right quotient of $\widetilde{L}$ with respect to the rational language $\{u^{-1}\}$, and so $\widetilde{L}_u$ is a deterministic context-free language by Hopcroft and Ullman [4, Theorem 10.2].

Let $M = \mathrm{Inv}\langle X \colon u_i = v_i \colon i \in I\rangle = (X \cup X^{-1})^*/\tau$. From the results of Stephen [15] we know that, for each $u \in (X \cup X^{-1})^*$, the transformation monoid of the automaton $B\Gamma(u)$ is the image of M under the Schützenberger representation $P_{R_{u\tau}}$ relative to the $\mathscr{R}$-class $R_{u\tau}$ of $u\tau$ in M: we denote this by $\mathrm{Schütz}_{R_{u\tau}}(M)$ in the following theorem. Recall that the Vagner congruence on $(X \cup X^{-1})^*$ is denoted by ρ and that $\mathrm{FIM}(X) = (X \cup X^{-1})^*/\rho$. We define a language $L \subseteq (X \cup X^{-1})^*$ to be *closed* if, for all $u \in L$ and $v \in (X \cup X^{-1})^*$, $v\rho \geq u\rho$ in $\mathrm{FIM}(X)$ implies that $v \in L$. Clearly, if L is closed, then $L\rho = \{u\rho \colon u \in L\}$ is a closed subset of $\mathrm{FIM}(X)$, so L is closed if and only if L is the inverse image under the Vagner homomorphism from $(X \cup X^{-1})^*$ onto $\mathrm{FIM}(X)$ of a closed subset of $\mathrm{FIM}(X)$.

Theorem 4.4. *Let $M = \mathrm{Inv}\langle X \colon e_i = f_i, i \in I\rangle = (X \cup X^{-1})^*/\tau$ where I is a finite set and e_i, f_i are idempotents of* $\mathrm{FIM}(X)$ *for each $i \in I$. Then for* *each* $u \in (X \cup X^{-1})^*$, *the set* $u\uparrow = \{w \in (X \cup X^{-1})^* : w\tau \geq u\tau\}$ *is a closed deterministic context-free language and the graph of the minimal automaton of* $u\uparrow$ *is a (labelled) tree whose vertices embed as a rational Schreier subset of* $\mathrm{FG}(X)$. *Furthermore the syntactic monoid of the language* $u\uparrow$ *is isomorphic to* $\mathrm{Schütz}_{R_{u\tau}}(M)$, *the image of* M *under the Schützenberger representation relative to* $R_{u\tau}$.

Proof. Let $u \in (X \cup X^{-1})^*$. We have already seen that $B\Gamma(u)$ is the minimal automaton of $u\uparrow$. From Theorem 2.2 we know that $B\Gamma(u)$ is a labelled tree that embeds in the natural way in $\Gamma(X)$ so the set $L = V(B\Gamma(u))$ of vertices of $B\Gamma(u)$ forms a Schreier subset of $\mathrm{FG}(X)$. By Theorem 3.2, L forms a rational subset of $\mathrm{FG}(X)$. Again by the results of Stephen [15] we know that $u\uparrow = \{w \in (X \cup X^{-1})^* : w$ labels a path from 1 to $r(u)$ in $B\Gamma(u)\}$ and so $u\uparrow$ is a deterministic context free language by Corollary 4.3. It is obvious from the definition that $u\uparrow$ is a closed language in $(X \cup X^{-1})^*$. From the results of Stephen [15] cited above we know that the transformation monoid of $B\Gamma(u)$ is $\mathrm{Schütz}_{R_{u\tau}}(M)$, so the final statement in the theorem follows since $B\Gamma(u)$ is the *minimal* automaton of $u\uparrow$, and hence the syntatic monoid of $u\uparrow$ is isomorphic to the transformation monoid of $B\Gamma(u)$ (see, for example, Lallement [5, Proposition 1.8]).

Remark 4.5. The condition that I be finite is not essential to obtain the main conclusions of the theorem; this condition is sufficient but not necessary to guarantee that $V(B\Gamma(u))$ is rational for each $u \in (X \cup X^{-1})^*$; it is clear that if the e_i, $f_i (i \in I)$ are chosen in any way that guarantees that each set $V(B\Gamma(u))$ is rational for $u \in (X \cup X^{-1})^*$, then the remaining conclusions of the theorem hold.

Corollary 4.6. *Let* $M = \mathrm{Inv}\langle X : e_i = f_i, i \in I\rangle = (X \cup X^{-1})^*/\tau$ *where* I *is finite and* e_i, f_i *are idempotents of* $\mathrm{FIM}(X)$ *for each* $i \in I$. *Then* $1\tau = \{w \in (X \cup X^{-1})^* : w\tau 1$ *in* $M\}$ *is a (closed) deterministic context-free language.*

Proof. This is evident from Theorem 4.4 since $w\tau \geq 1\tau$ in M implies $w\tau = 1\tau$.

Note, in particular, that the well-known fact that the (restricted) Dyck languages are context free is a special case of this result (see Berstel [2]).

Concluding remarks

The results of Theorem 3.2 were announced by the authors at an international conference on semigroups in Szeged, Hungary in August 1987. The authors wish to thank J. B. Stephen and J.-C. Birget for many fruitful conversations concerning parts of this paper. In particular, further results along these lines will appear in the forthcoming paper of Birget and the authors [3].

Bibliography

1. J. Barwise, *Handbook of mathematical logic*, North-Holland, 1977.
2. J. Berstel, *Transductions and context-free languages*, Teubner, Studienbücher, 1979.
3. J. C. Birget, S. W. Margolis, and J. C. Meakin, *The word problem for inverse monoids presented by one idempotent relator*, Theoret. Comp. Sci. (to appear).
4. J. E. Hopcroft and J. D. Ullman, *Formal languages and their relation to automata*, Addison-Wesley, 1969.

5. G. Lallement, *Semigroups and combinatorial applications*, Wiley, 1979.

6. S. W. Margolis and J. Meakin, *E-unitary inverse monoids and the Cayley graph of a group presentation*, J. Pure Appl. Algebra **58** (1989), 45–76.

7. D. E. Muller and P. E. Schupp, *The theory of ends, pushdown automata, and second-order logic*, Theoret. Comput. Sci. **37** (1985), 51–75.

8. W. D. Munn, *Free inverse semigroups*, Proc. London Math. Soc. **30** (1974), 385–404.

9. M. Petrich, *Inverse semigroups*, Wiley, 1984.

10. M. O. Rabin, *Decidability of second order theories and automata on infinite trees*, Trans. Amer. Math. Soc. **141** (1969), 1–35.

11. ——, *Automata on infinite objects and Church's problem*, CBMS Regional Conf. Ser. in Math., no. 14, Amer. Math. Soc., Providence, R.I., 1971.

12. H. E. Scheiblich, *Free inverse semigroups*, Proc. Amer. Math. Soc. **38** (1973), 1–7.

13. J.-P. Serre, *Trees*, Springer-Verlag, Berlin and New York, 1980.

14. J. Stallings, *Topology of finite graphs*, Invent. Math. **71** (1983), 551–565.

15. J. B. Stephen, *Presentations of inverse monoids*, J. Pure Appl. Algebra **63** (1990), 81–112.

Department of Computer Science, Department of Mathematics and Statistics, University of Nebraska-Lincoln, Lincoln, Nebraska 68588