EQUATIONS IN FREE INVERSE MONOIDS

Timothy Deis, John Meakin[*] and Géraud Sénizergues

To John Rhodes, on his sixty-fifth birthday

Abstract

It is known that the problem of determining consistency of a finite system of equations in a free group or a free monoid is decidable, but the corresponding problem for systems of equations in a free inverse monoid of rank at least two is undecidable. Any solution to a system of equations in a free inverse monoid induces a solution to the corresponding system of equations in the associated free group in an obvious way, but solutions to systems of equations in free groups do not necessarily lift to solutions in free inverse monoids. In this paper we show that the problem of determining whether a solution to a finite system of equations in a free group can be extended to a solution of the corresponding system in the associated free inverse monoid is decidable. We are able to use this to solve the consistency problem for certain classes of single variable equations in free inverse monoids.

1 Introduction

An *inverse* monoid is a monoid M with the property that for each $a \in M$ there exists a unique element $a^{-1} \in M$ such that $a = aa^{-1}a$ and $a^{-1} = a^{-1}aa^{-1}$. Equivalently, M is a von-Neumann regular monoid whose idempotents commute. The idempotents of such a monoid form a (lower) semilattice with respect to multiplication as the meet operation, and we denote the semilattice of idempotents of an inverse monoid M by $E(M)$. Inverse monoids arise naturally as monoids of partial symmetries (partial one-one structure-preserving maps) throughout mathematics. We refer the reader to the books by Petrich [15], Lawson [7], and Patterson [14] for much information about the structure of inverse monoids and their connections with other branches of mathematics.

Inverse monoids form a variety of algebras (in the sense of universal algebra) with respect to the operations of multiplication, inversion, and choosing the identity. As such, free inverse monoids exist. We denote the free inverse monoid on a set A by $FIM(A)$. The free monoid on A will be denoted by A^* and the free group on A will be denoted by $FG(A)$. It is convenient to denote the alphabet $A \cup A^{-1}$ by $\tilde{A}$ (and the free monoid on this alphabet by $\tilde{A}^*$). It is easy to see that $FG(A)$ is the maximal group homomorphic image of $FIM(A)$. The structure of $FIM(A)$ is determined by considering finite subtrees of the Cayley tree of the free group (with respect to the usual presentation of $FG(A)$).

Denote the Cayley tree of $FG(A)$ by $\Gamma(A)$. The vertices of $\Gamma(A)$ may be identified with reduced words (elements of $FG(A)$), and there is an edge in $\Gamma(A)$ labeled by an element $a \in \tilde{A}$ from g to ga for each $g \in FG(A)$. Note that if a labels an edge from g to ga, then a^{-1} labels an edge from ga to a. For each word $w \in \tilde{A}^*$, let $MT(w)$ be the Munn tree of w. Here $MT(w)$ is the finite subtree of $\Gamma(A)$ obtained when the word w is read as a path in $\Gamma(A)$ starting at 1 and

[*] Research supported by NSF grant No. DMS-9970471

ending at the reduced form $r(w)$ of w. A theorem of Munn [13] (see also [15, 7]) states that two words u and v in $\tilde{A}^*$ are equal in $FIM(A)$ if and only if $MT(u) = MT(v)$ and $r(u) = r(v)$. This provides a solution to the word problem for $FIM(A)$. If Γ is any finite subtree of $\Gamma(A)$ containing the vertex 1 and if g is any vertex of Γ, then there is at least one word $u \in \tilde{A}^*$ (in fact infinitely many words) such that $(MT(u), r(u)) = (\Gamma, g)$. The monoid $FIM(A)$ may be identified with the set $\{(MT(w), r(w)) : w \in \tilde{A}^*\}$ with multiplication

$$(MT(u), r(u)) \times (MT(v), r(v)) = ((MT(u) \cup r(u)MT(v), r(uv)). \tag{1}$$

The idempotents of $FIM(A)$ consist of Dyck words in $\tilde{A}^*$, i.e. words whose reduced form is 1. Two such Dyck words represent the same idempotent in $FIM(A)$ if and only if they have the same Munn tree. There is a natural partial order on any inverse monoid M defined by $a \leq b$ if and only if $a = eb$ for some idempotent $e \in E(M)$. The congruence on M induced by this relation is denoted by σ_M (or just σ if M is understood) and is the minimum group congruence on M (i.e. M/σ_M is the maximum group homomorphic image of M). For $FIM(A)$, each σ-class contains a maximum element (the reduced form of a word in the σ-class) and of course $FIM(A)/\sigma \cong FG(A)$.

Let X be an alphabet that is disjoint from A. We will view letters of $\tilde{X}$ as *variables* and elements of $\tilde{A}^*$ as *constants*. The sets A and X will be assumed to be *finite and non-empty* throughout this paper. An *equation* in $FG(A)$ or in $FIM(A)$ with coefficients in $FG(A)$ (or in $FIM(A)$) is a pair (u, v), where $u, v \in (\tilde{A} \cup \tilde{X})^*$. Usually we will denote such an equation by $u = v$: if necessary for emphasis we will denote u and v by $u(X, A)$ and $v(X, A)$ if there is any possibility of confusion about the sets of variables and constants in the equation. Similarly an equation in A^* is a pair (u, v) with $u, v \in (A \cup X)^*$, and again we will denote this by $u = v$. If needed to distinguish where equations are being viewed, we will denote an equation $u = v$ in A^*, [resp. $FG(A)$, $FIM(A)$] by $u =_M v$ [resp. $u =_G v, u =_I v$].

Any map $\phi : X \to \tilde{A}^*$ extends to a homomorphism (again denoted by ϕ) from $(\tilde{A} \cup \tilde{X})^*$ in such a way that ϕ fixes the letters of A. We say that ϕ is a *solution* to the equation $u =_G v$ in $FG(A)$ [resp. $u =_I v$ in $FIM(A)$ or $u =_M v$ in A^*] if $\phi(u) = \phi(v)$ in the appropriate setting. A solution to a set of equations $u_i = v_i$ for $i = 1, \ldots n$ is a map ϕ that is a solution to each equation in the set. If a set of equations has at least one solution it is called *consistent*: otherwise it is called *inconsistent*. It is easy to give examples of equations that are inconsistent in any of the three possible settings where we are considering such equations, and it is easy to give examples of equations that are consistent in $FG(A)$ but not in $FIM(A)$ or in A^*. For example, if $A = \{a, b\}$, then the equation $ax = xb$ is inconsistent in all three settings, while the equation $ax = b$ is consistent in $FG(A)$ but inconsistent in A^* and in $FIM(A)$. On the other hand it is obvious that any set of equations that is consistent in $FIM(A)$ must be consistent in $FG(A)$: if ψ is any solution to a set of equations in $FIM(A)$ and $\psi(x) = w_x \in \tilde{A}^*$ for each $x \in X$, then $\phi : X \to \tilde{A}^*$ defined by $\phi(x) = r(w_x)$ is a solution to the same set of equations, viewed as equations in $FG(A)$.

The *consistency problem* for systems of equations in A^* [resp. $FG(A), FIM(A)$] is the problem of determining whether there is an algorithm that, on input a finite set $\{u_i = v_i : i = 1 \ldots n\}$ of equations in A^* [resp. $FG(A), FIM(A)$], produces an output of "Yes" if the system is consistent and "No" if it is inconsistent. Theorems of Makanin [11, 12] imply that the consistency problems for systems of equations in A^* and in $FG(A)$ are decidable. Much work has been done on solutions to systems of equations in free monoids and free groups: we refer the reader to [9, 6, 16, 18, 4] for just some of the extensive literature on this subject. On the other hand, a theorem of Rozenblat [19] shows that while the consistency problem for systems of equations in $FIM(A)$ is decidable if $|A| = 1$, this problem is undecidable if $|A| > 1$. The consistency problem for equations of some restricted type (for example, single variable equations, or quadratic equations) is open as far as we are aware. Some work on special cases of this problem has been done by Deis [5]. For example, Deis [5] has shown that while the consistency problem for single *multilinear* equations in $FIM(A)$ is decidable, the consistency problem for finite *systems* of multilinear equations is undecidable. We will show later in this paper that the consistency problem for single-variable equations of a particular type is decidable.

Now consider an equation $u =_I v$ in $FIM(A)$, let ψ be a solution to this in $FIM(A)$, and let ϕ be a solution to the corresponding equation in $FG(A)$, where $\phi(x)$ is a reduced word for each $x \in X$. We say that ψ is an *extension* of ϕ (or that ϕ *extends to* ψ) if for each $x \in X$ there is some Dyck word e_x such that $\psi(x) = e_x\phi(x)$. If ψ is a solution to an equation $u = v$ in $FIM(A)$ and if $\phi(x) = r(\psi(x))$ for each $x \in X$, then of course ϕ is a solution to $u = v$ in $FG(A)$ and ψ is an extension of ϕ.

A given solution ϕ to an equation $u = v$ in $FG(A)$ may admit finitely many extended solutions, infinitely many extended solutions, or no extended solutions, to the same equation in $FIM(A)$. For example, the equation $bb^{-1}x = aa^{-1}bb^{-1}$ has trivial solution in $FG(a, b)$, and this has exactly two extensions $\psi_1(x) = aa^{-1}bb^{-1}$ and $\psi_2(x) = aa^{-1}$ in $FIM(a, b)$. The equation $bb^{-1}x = aa^{-1}x$ has trivial solution in $FG(a, b)$ that extends to infinitely many solutions $\psi_e(x) = e$ for any idempotent $e \leq aa^{-1}bb^{-1}$ in the natural order on $FIM(a, b)$. The equation $a^{-1}ax = aa^{-1}$ has trivial solution in $FG(a, b)$ but no solution in $FIM(a, b)$. These facts are easy to check via the multiplication of Munn trees in the free inverse monoid, as described in equation (1).

A natural question arises here: when does a solution to an equation $u = v$ in $FG(A)$ extend to a solution to the same equation in $FIM(A)$? We refer to the corresponding algorithmic problem as the *extendibility problem* for equations in $FIM(A)$. More precisely, the extendibility problem for equations in $FIM(A)$ asks whether there is an algorithm that, on input a finite set $\{u_i = v_i : i = 1, \ldots n\}$ of equations in $FIM(A)$ that is consistent in $FG(A)$ and a solution ϕ to this system in $FG(A)$, produces the output "Yes" if ϕ can be extended to a solution to the system of equations in $FIM(A)$ and "No" if ϕ cannot be extended to a solution to this system in $FIM(A)$. Some special cases of the extendibility problem were considered by Deis [5]. The main result of this paper shows that the extendibility problem is decidable.

2 The Extendibility Problem

In order to study the extendibility problem, we first reformulate it somewhat in terms of Munn trees. Let $u = v$ be an equation in $FIM(A)$ and $\phi : X \to \tilde{A}^*$ a solution to this equation in $FG(A)$. Thus $\phi(u) = \phi(v)$ in $FG(A)$ (but not necessarily in $FIM(A)$ of course). The edges of $MT(u)$ [and $MT(v)$] are labeled over the alphabet $\tilde{A} \cup \tilde{X}$. For each variable x that occurs in the word u, there is at least one edge in $MT(u)$ labeled by x. The Munn tree $MT(\phi(u))$ has edges labeled over the alphabet $A \cup A^{-1}$. It is obtained from $MT(u)$ by replacing each (directed) edge e labeled by a variable $x \in X$ by the tree $MT(\phi(x))$: in this replacement, the initial root (i.e. 1) of this copy of $MT(\phi(x))$ is identified with the initial vertex of the edge e and the terminal root of $MT(\phi(x))$ is identified with the terminal vertex of e. This process is well defined since if u' is another word with $MT(u') = MT(u)$ and $r(u') = r(u)$ then $u' = u$ in $FIM(A \cup X)$, so $\phi(u') = \phi(u)$ in $FIM(A)$, and so $MT(\phi(u')) = MT(\phi(u))$. The relationship between $MT(v)$ and $MT(\phi(v))$ is described in a similar fashion.

The extension of ϕ to a homomorphism (again denoted by ϕ) from $(\tilde{A} \cup \tilde{X})^*$ to $\tilde{A}^*$ naturally induces a homomorphism $\bar{\phi}$ from $FG(A \cup X)$ to $FG(A)$. Thus each vertex of $MT(u)$ [resp. $MT(v)$] that is the initial vertex of an edge of $MT(u)$ [resp. $MT(v)$] labeled by a letter $x \in X$ has a unique image in $MT(\phi(u))$ [resp. $MT(\phi(v))$] under this homomorphism. We refer to the vertices obtained this way as images of initial vertices of edges of $MT(u)$ [resp. $MT(v)$] labeled by the letter $x \in X$ as *designated x-vertices* of $MT(\phi(u))$ [resp. $MT(\phi(v))$]. For example, if $w = abx_1b^{-1}bbx_2^{-1}a$ and $\phi(x_1) = b^{-1}$ and $\phi(x_2) = a$, then $MT(\phi(w))$ has two designated vertices: namely ab is a designated x_1-vertex and aba^{-1} is a designated x_2-vertex. Similarly, if $w' = abx_1b^{-1}bbx_2a$ and we take the same map ϕ as above, then ab is both a designated x_1-vertex and a designated x_2-vertex.

Now suppose that $\psi(x) = e_x\phi(x)$ for all $x \in X$, where each e_x is a Dyck word. Since the terminal root of $MT(e_x)$ is the same as its initial root (1), it follows that the designated x-vertices of $MT(\phi(w))$ and of $MT(\psi(w))$ coincide, for each word w and each $x \in X$. Furthermore, $MT(\psi(w))$ is obtained from $MT(\phi(w))$ by adjoining to $MT(\phi(w))$ a copy of $MT(e_x)$ rooted at each designated x-vertex of $MT(\phi(w))$. Recall that this map ψ defines an extension of ϕ if ψ is a solution to $u = v$ in $FIM(A)$, i.e. if $MT(\psi(u)) = MT(\psi(v))$.

Now let $\{u_i = v_i : i = 1, \ldots, n\}$ be a system of equations in $FIM(A)$, and let ϕ be a solution to this system in $FG(A)$. For each variable $x \in X$ denote the set of designated x-vertices of $MT(\phi(u_i))$ [resp. $MT(\phi(v_i))$] by $\alpha_{i,x}$ [resp. $\alpha'_{i,x}$] and denote the set of vertices of $MT(\phi(u_i))$ [resp. $MT(\phi(v_i))$] by β_i [resp. β'_i]. It is convenient to denote multiplication in $FG(A)$ by $\cdot$ and to denote the union $S \cup T$ of two subsets of $FG(A)$ or $(A \cup A^{-1})^*$ by $S + T$. Finally, let us denote the set of vertices of the Munn tree $MT(e_x)$ of some (unknown) Dyck word e_x by T_x (for each $x \in X$).

The requirement that ϕ should be extendible to some solution $\psi(x) = e_x\phi(x)$ to the system in $FIM(A)$ translates as follows. Consider the system of equations

$$\sum_x \alpha_{i,x} \cdot T_x + \beta_i = \sum_x \alpha'_{i,x} \cdot T_x + \beta'_i : i = 1, \ldots, n \tag{2}$$

Here the $\alpha_{i,x}, \alpha'_{i,x}, \beta_i, \beta'_i$ are finite subsets of $FG(A)$ and the T_x are unknowns. A solution of (2) is any collection of subsets $T_x (x \in X)$ of $FG(A)$ that satisfies this system of equations. We would like to decide whether the system of equations (2) has at least one solution such that each T_x is both finite and prefix closed. (A subset T of $FG(A)$ is *prefix closed* if the corresponding set of reduced words is prefix closed.) We will show that this problem is decidable by appealing to Rabin's tree theorem [17]. From the discussion above, this will show that the extendibility problem is decidable.

We assume some familiarity with basic definitions and ideas of (first order) logic. See, for example, Barwise [2]. In second order monadic logic, quantifiers refer to sets (i.e. unary or monadic predicates) as well as to individual members of a structure. The syntax and semantics of terms and well formed formulae are defined inductively in the usual way. Atomic formulae include those of the form $t \in Y$ where t is a term and Y is a set variable. A sentence of the form $\forall Y \nu(Y)$ where Y is a set variable, in particular, is true in a structure M iff $\nu(Y)$ is (inductively) true in M for all subsets Y of the universe of M. If a sentence θ is true in a structure M we write $M \models \theta$ and we define $Th_2(M) = \{\theta : M \models \theta\}$. The (second order monadic) theory of M is *decidable* if there is an algorithm that tests whether a given sentence θ of the language of M is in $Th_2(M)$ or not.

Let A be a countable set and consider the structure $T_A = (A^*, \{r_a : a \in A\}, \leq)$. Here $r_a : A^* \to A^*$ is right multiplication by a, $xr_a = xa, \forall x \in A^*$ and $\leq$ is the prefix order $x \leq y$ iff $\exists u \in A^*(xu = y)$. The theory $Th_2(T_A)$ is called the theory of A-successor functions. For $|A| = 2$ this is often denoted by $S2S$, and sentences in $Th_2(T_A)$ can be reformulated as sentences in $S2S$. Rabin's tree theorem stated below is one of the most powerful decidability results known in model theory: the decidability of many other results can be reduced to $Th_2(T_A)$ (see, for example, [2]).

Theorem 1 *(Rabin [17]) $Th_2(T_A)$ is decidable.*

The main theorem of this paper is the following.

Theorem 2 *There is an algorithm that will decide, on input a system of equations of the form (2), whether this system of equations has at least one solution $\{T_x : x \in X\}$ such that each T_x is a finite prefix-closed subset of $FG(A)$.*

In order to use Rabin's theorem to prove this, we need to show that the existence of a solution of the desired type to (2) is expressible in $S2S$.

Step 1: View each element of each set $\alpha_{i,x}, \alpha'_{i,x}, \beta_i$, β'_i and T_x as a reduced word in $\tilde{A}^*$. In order to translate the equations (2) over subsets of FG(A) into similar equations, but over subsets of $\tilde{A}^*$, we decompose the coefficients $\alpha_{i,x}, \alpha'_{i,x}$ and as well, the sets T_x into a finite number of components.

Let us consider the set

$$S = \{(a, u) \in (\tilde{A} \cup \{\epsilon\}) \times \tilde{A}^* \mid \exists v \in \tilde{A}^*, v \cdot u \in \sum_{i \in X, 1 \leq i \leq n} \alpha_{i,x} + \alpha'_{i,x} \text{ and } a = v^{(1)}\}$$

where $v^{(1)}$ denotes the last letter of v , if $|v| \geq 1$, and the empty word, ϵ, otherwise.
Let us denote the elements of S by $\{(a_j, u_j) \mid 1 \leq j \leq k\}$. For every $j \in [1, k]$ we write:

$$\begin{array}{lll} T_{j,x} = & \{u \in \tilde{A}^* \mid u_j^{-1} \cdot u \in T_x \text{ and } u \text{ does not begin with letter } a_j^{-1}\} & (\text{ if } a_j \in \tilde{A}) \\ T_{j,x} = & \{u \in \tilde{A}^* \mid u_j^{-1} \cdot u \in T_x\} & (\text{ if } a_j = \epsilon) \end{array}$$

Accordingly, for every $1 \leq i \leq n, 1 \leq j \leq k, x \in X$, we define the sets

$$\alpha_{i,j,x} = \{u \in A^* \mid u \cdot u_j \in \alpha_{i,x} \text{ and } u^{(1)} = a_j\}$$

and $\alpha'_{i,j,x}$ is defined similarly.

The equations (2) reduce to the system of equations

$$\sum_{x \in X} \sum_{j=1}^{k} \alpha_{i,j,x} \cdot T_{j,x} + \beta_i = \sum_{x \in X} \sum_{j=1}^{k} \alpha'_{i,j,x} \cdot T_{j,x} + \beta'_i, i = 1, \ldots, n \tag{3}$$

Note that the effect of our chosen decompositions of the sets $\alpha_{i,x}, \alpha'_{i,x}, T_x$, is that all products in the system (3) are reduced as written - so (3) may be viewed as a system of equations in the free monoid $\tilde{A}^*$, where $\alpha_{i,j,x}$, $\alpha'_{i,j,x}$, β_i and β'_i are prescribed finite subsets of this free monoid, and the $T_{j,x}$ are the unknowns. A *solution* to (3) is a vector of subsets $\{T_{j,x} : x \in X, j = 1, \ldots, k\}$ of (reduced) words in $\tilde{A}^*$ that satisfies (3). We seek to decide whether (3) has a solution so that each $T_{j,x}$ is a *finite, prefix-closed* subset of *reduced* words in $\tilde{A}^*$.

Step 2: For each set U of words in $\tilde{A}^*$, let $Pref(U)$ denote the set of prefixes of words in U. Though the existence of a finite solution $\{E_{j,x} : x \in X, j = 1, \dots, k\}$ to (3) does not necessarily imply the existence of a finite prefix-closed solution to these equations, we can note that this is in a sense "almost" the case, and we will see how to impose additional conditions to obtain a finite prefix closed solution to this system of equations. Let N be the maximum length of a word in any of the sets $\alpha_{i,j,x}, \alpha'_{i,j,x}, \beta_i$, and β'_i.

Suppose that $\{E_{j,x} : x \in X, j = 1, \dots, k\}$ is a finite solution to (3). We first prove the following Lemma.

Lemma 1 *Suppose that $u \in Pref(E_{j,x})$ for some j and some x, and that $|u| > N$. Then for all $i = 1, \dots, n$, if $v \in \alpha_{i,j,x}$, then $v \cdot u \in \sum_x \sum_j \alpha'_{i,j,x} \cdot Pref(E_{j,x}) + \beta'_i$.*

Proof. There exists a reduced word $s \in \tilde{A}^*$ such that $u \cdot s \in E_{j,x}$ and $u \cdot s$ is reduced as written. It follows that $v \cdot u \cdot s \in \sum_x \sum_j \alpha'_{i,j,x} \cdot E_{j,x} + \beta'_i$. Since $|u| > N$, it follows that $v \cdot u \cdot s \notin \beta'_i$, so $v \cdot u \cdot s \in \sum_x \sum_j \alpha'_{i,j,x} \cdot E_{j,x}$. Hence there exist $y, \bar{\jmath}$ and $v' \in \alpha'_{i,\bar{\jmath},y}$, $e' \in E_{\bar{\jmath},y}$ such that $v \cdot u \cdot s = v' \cdot e'$. But again, since $|u| > N$, s must be a suffix of e', so $e' = u' \cdot s$ for some u'. So we have $v \cdot u \cdot s = v' \cdot u' \cdot s$ in $\tilde{A}^*$. It follows that $v \cdot u = v' \cdot u'$ where $u' \in Pref(E_{\bar{\jmath},y})$. ∎

Step 3: Lemma 1 shows that if $\{E_{j,x} : x \in X,\ j = 1, \dots, k\}$ is a finite solution to (3), then $\{Pref(E_{j,x}) : x \in X,\ j = 1, \dots, k\}$ is "almost" a solution to (3). In order to arrange for a prefix-closed solution to (3) we need only assume some additional conditions on the "short" prefixes of elements of each set $E_{x,j}$. Since these prefixes must be included in a finite set that we know in advance, we are able to formulate appropriate additional conditions as follows.

Denote by $\tilde{A}^N$ [resp. $\tilde{A}^{\leq N}$] the set of words in $\tilde{A}^*$ of length N [resp. $\leq N$]. Let us introduce another vector of unknowns, $\{P_{j,x} : x \in X,\ j = 1, \dots, k\}$ and consider the additional conditions:

$$P_{j,x} \subseteq E_{j,x},\ x \in X,\ j = 1, \dots, k \tag{4}$$

$$E_{j,x} \subseteq P_{j,x} + (P_{j,x} \cap \tilde{A}^N) \cdot \tilde{A}^*,\ x \in X,\ \ j = 1, \dots, k \tag{5}$$

$$P_{j,x} \subseteq \tilde{A}^{\leq N},\ x \in X,\ j = 1, \dots, k \tag{6}$$

We have the following two lemmas.

Lemma 2 *Let $\{(P_{j,x}, E_{j,x}) : x \in X,\ j = 1, \dots, k\}$ be a finite solution to (3,4,5,6) such that each $P_{j,x}$ is prefix-closed. Then $\{Pref(E_{j,x}) : x \in X,\ j = 1, \dots, k\}$ is a finite prefix-closed solution to (3).*

Proof. Since $\{E_{j,x} : x \in X,\ j = 1, \dots k\}$ is a solution to (3), it is clear that $\beta_i \subseteq \sum_x \sum_j \alpha'_{i,j,x} \cdot E_{j,x} + \beta'_i$ for each $i = 1, \dots, n$. Let $u \in Pref(E_{j,x})$ and $v \in \alpha_{i,j,x}$ for some i, j, x. If $|u| > N$ then we already know by Lemma 1 that $v \cdot u \in \sum_x \sum_j \alpha'_{i,j,x} \cdot E_{j,x} + \beta'$. So assume that $|u| \leq N$. There exists some (reduced) word s such that $u \cdot s \in E_{j,x}$ and $u \cdot s$ is reduced as written. If $u \cdot s \in P_{j,x}$, then $u \in P_{j,x}$ since we are assuming that each $P_{j,x}$ is prefix-closed. Otherwise we must have $u \cdot s \in (P_{j,x} \cap A^N) \cdot A^*$ by (5). But then since u is a prefix of $u \cdot s$ of length $\leq N$, we must have that u is a prefix of a word in $P_{j,x}$, and so (again since $P_{j,x}$ is prefix-closed) we must have $u \in P_{j,x}$. Hence $v \cdot u \in \alpha_{i,j,x} \cdot E_{j,x}$ by (4). Since $\{E_{j,x} : x \in X,\ j = 1, \dots, k\}$ is a solution to (3), this implies that $v \cdot u \in \sum_x \sum_j \alpha'_{i,j,x} \cdot E_{j,x} + \beta'_i$. It follows that $\sum_x \sum_j \alpha_{i,j,x} \cdot Pref(E_{j,x}) + \beta_i \subseteq \sum_x \sum_j \alpha'_{i,j,x} \cdot Pref(E_{j,x}) + \beta'_i$ for each $i = 1, \dots, n$. The reverse inclusion follows dually and so $\{Pref(E_{j,x}) : x \in X,\ j = 1, \dots, k\}$ is a solution to (3), as required. ∎

Lemma 3 *Let $\{T_{j,x} : x \in X, j = 1, \ldots, k\}$ be a finite prefix-closed solution to (3) and set $E_{j,x} = T_{j,x}$ and $P_{j,x} = T_{j,x} \cap A^{\leq N}$ for each $x \in X$ and $j = 1, \ldots, k$. Then $\{(P_{j,x}, E_{j,x}) : x \in X, j = 1, \ldots, k\}$ is a solution to (3,4,5,6) and each $P_{j,x}$ is prefix-closed.*

Proof. It is trivial to verify that conditions (3), (4), and (6) are satisfied by our choice of the $P_{j,x}$ and $E_{j,x}$. To verify (5), simply note first that any word in $T_{j,x}$ of length $\leq N$ is in $P_{j,x}$ by definition of $P_{j,x}$. Also, if u is a word in $T_{j,x}$ of length $\geq N$, then we may write $u = u' \cdot s$ where u' is a prefix of u of length N and $s \in \tilde{A}^*$. But then since $T_{j,x}$ is prefix-closed, $u' \in T_{j,x}$ and so $u \in (P_{j,x} \cap \tilde{A}^N) \cdot \tilde{A}^*$. This completes the verification that (5) is satisfied. ■

Step 4 - The Decision Algorithm: By Lemmas 2 and 3, we are reduced to deciding whether, among all the prefix-closed $P_{j,x}$ satisfying (6), there is a collection such that (3) (where the unknowns are renamed $E_{j,x}$), (4), and (5) are also satisfied by some finite sets of reduced words.

Enumerate effectively all of the prefix-closed $P_{j,x}$ satisfying (6). We now translate each of the conditions (3), (4), and (5) into their "mirror" conditions in the dual semigroup to $\tilde{A}$. For each word $w = s_1 s_2 \ldots s_k$ (with each $s_j \in \tilde{A}$), we define $\hat{w}$ to be the mirror word $\hat{w} = s_k \ldots s_2 s_1$. For each subset $F \subseteq \tilde{A}^*$ we define $\hat{F} = \{\hat{w} : w \in F\}$. For a given collection of prefix-closed sets $P_{j,x}$, $x \in X, j = 1, \ldots, k$, one can consider the mirror versions of (3), (4) and (5).

The mirror version of (3) is

$$\sum_x \sum_j F_{j,x} \cdot \hat{\alpha}_{i,j,x} + \hat{\beta}_i = \sum_x \sum_j F_{j,x} \cdot \hat{\alpha}'_{i,j,x} + \hat{\beta}'_i, \; i = 1, \ldots, n. \tag{7}$$

Notice that in these equations, the variables $F_{j,x}$ are on the left and the constants are on the right. Also, the equations (7) have a solution $\{F_{j,x} : x \in X, j = 1, \ldots, k\}$ if and only if the equations (3) have a solution $\{E_{j,x} : x \in X, j = 1, \ldots, k\}$, where $F_{j,x} = \hat{E}_{j,x}$ for each $x \in X, j = 1, \ldots k$. Also notice that the existence of a solution to (7) is expressible in S2S, because right product by given words is a finite composition of successor functions. But this implies that the existence of a *finite* solution to (7) (i.e. a solution where all sets $F_{j,x}$ are finite) is also expressible in S2S, simply because finiteness is expressible in S2S. [Let us recall this standard trick: by König's Lemma, a set $F \subseteq \tilde{A}^*$ is infinite iff it admits a set of prefixes F' such that every element of F' has some successor inside F'; this characterisation is expressible in S2S].

The mirror version of (4) is

$$\hat{P}_{j,x} \subseteq F_{j,x}, \; x \in X, j = 1, \ldots, k. \tag{8}$$

Here each $\hat{P}_{j,x}$ is a fixed finite subset of $\tilde{A}^*$ (corresponding to the fixed choice of the $P_{j,x}$ that we are working with), and each $F_{j,x}$ is a variable. Clearly the existence of a solution to these conditions is expressible in S2S.

In order to express the mirror version of (5) in S2S, notice that the mirror image $R_{j,x}$ of $(P_{j,x} \cap \tilde{A}^N) \cdot \tilde{A}^*$ is the smallest subset X of $\tilde{A}^*$ such that $w \cdot s \in X$ for all $w \in \tilde{A}^*$ and all s in the fixed finite set consisting of mirror images of words in $(P_{j,x} \cap \tilde{A}^N)$. Since there are again just finitely many choices for these words s, since all variables w occur on the left, and since it is possible to express in S2S the fact that a set X is the smallest subset satisfying some other property that is expressible in S2S, membership in the sets $R_{j,x}$ is expressible in S2S. The mirror version of (5) then becomes

$$F_{j,x} \subseteq \hat{P}_{j,x} + R_{j,x}, \; x \in X, j = 1, \ldots, k. \tag{9}$$

where the $F_{j,x}$ are variables and the $R_{j,x}$ are described above. Hence it is possible to express in S2S the fact that the $F_{j,x}$ satisfy these conditions.

In addition, it is clear that it is possible to express in S2S the fact that the $F_{j,x}$ consist of *reduced* words only (this property is invariant by the mirror operation).

Finally, notice now that for fixed finite prefix-closed sets $P_{j,x}$ $(x \in X, j = 1, \ldots, k)$ satisfying (6), the existence of sets $E_{j,x}(x \in X, j = 1, \ldots, k)$ that satisfy (3,4,5) is translated in the mirror conditions to the existence of sets $F_{j,x}$ that satisfy (7),(8) and (9), and that $F_{j,x} = \hat{E}_{j,x}$ for each x and j. We can decide, using Rabin's tree theorem, whether (7,8,9) has at least one finite solution $\{F_{j,x} : x \in X, j = 1, \ldots, k\}$ in $\tilde{A}^*$, and the answer to this decides whether (3,4,5) has at least one finite solution $\{E_{j,x} : x \in X, j = 1, \ldots, k\}$ in $\tilde{A}^*$ (for the $P_{j,x}$ under scrutiny). If, for some finite prefix-closed sets $P_{j,x}$ satisfying (6), the answer is "Yes", then (3) has some finite prefix-closed solution: otherwise, (3) has no finite prefix closed solution. This completes the proof of Theorem 2.

As an immediate corollary we obtain the following result.

Theorem 3 *Let A be a finite set. Then the extendibility problem for $FIM(A)$ is decidable.*

3 The Consistency Problem for Single-variable Equations

Recall that the theorem of Rozenblat [19] shows that the consistency problem for finite systems of equations in $FIM(A)$ is undecidable. Deis [5] has shown that the consistency problem for a system consisting of one multilinear equation in $FIM(A)$ (i.e. an equation $u = v$ in which each variable labels exactly one edge in $MT(u) \cup MT(v)$) is decidable, but that the consistency problem for finite systems of multilinear equations in $FIM(A)$ is undecidable. In this section we show how the results of the previous section may be applied to study the consistency problem for systems consisting of one *single-variable* equation in $FIM(A)$. A single-variable equation in $FIM(A)$ is an equation involving just one variable x (that may occur many times in the equation, with exponent ± 1). We are able to solve the consistency problem for a large class of single variable equations in $FIM(A)$.

It is clear from Theorem 3 that the consistency problem for a class of equations in $FIM(A)$ is decidable if the corresponding equations in $FG(A)$ have only finitely many solutions. A class of single-variable equations for which this is the case was identified in a paper of Silva [20].

In the following, we consider a single-variable equation $w(x) = 1$ in $FG(A)$, where $w(x)$ is the reduced word

$$w(x) = c_1 x^{\epsilon_1} c_2 x^{\epsilon_2} \ldots c_t x^{\epsilon_t} c_{t+1}, \tag{10}$$

with each $c_i \in \tilde{A}^*$ and $\epsilon_i \in \{-1, 1\}$.

The proof of the following result in [20] is attributed to James Howie.

Theorem 4 *Let $w(x) = 1$ be a single-variable equation in $FG(A)$ and suppose that the exponent sum of the single variable x in $w(x)$ is not zero. Then the equation $w(x) = 1$ can have at most one solution in $FG(A)$.*

As an immediate corollary of this and Theorem 3, we obtain the following fact.

Corollary 1 *Consider the class $\mathcal{C}$ consisting of single-variable equations $u = v$ in $FIM(A)$ in which the sum of the exponents of the variable in u is not equal to the sum of the exponents of the variable in v. Then the consistency problem for this class is decidable. That is, there is an algorithm such that on input one equation $u = v$ in $\mathcal{C}$, will produce the output "Yes" if the equation is consistent in $FIM(A)$, and "No" if it is inconsistent.*

In order to extend this result to other classes of single variable equations in $FIM(A)$, we recall some of the established literature on single variable equations in free groups. A parametric description of the set of all solutions to a single-variable equation $w = 1$ in $FG(A)$ was obtained by Lyndon [10]. Lyndon's result was refined somewhat by Appel [1] and subsequently by Lorents [8].

Let M be twice the maximum of the lengths of the c_i in equation (10). In the following result, a *parametric word* is a word of the form $u = w_1 w_2^{\alpha} w_3$ in which α is a parameter, $w_1 w_2^{\epsilon} w_3$ is reduced for $\epsilon \in \{-1, 1\}$, and w_2 is cyclically reduced and not a proper power. A value of u is the element of $FG(A)$ obtained by substituting an integer value for α. The refinement of Lyndon's and Appel's result that we shall use (due to Lorents [8]) is the following.

Theorem 5 *The set of solutions to any equation of the form $w(x) = 1$ in $FG(A)$, where $w(x)$ is the word (10), is the union of:*
(A) a finite set of solutions whose lengths are $\leq 4M$; and
(B) the set of **all** *values of some finite set of parametric words.*

We remark that the proofs of the theorems in the papers by Appel and Lorents are effective, so the set of parametric words that can yield solutions to $w(x) = 1$ in $FG(A)$ is effectively constructible (in fact $|w_1 w_2 w_3| \leq 5M$, in the notation above). This, together with the following definition, will enable us to extend Theorem 4 to a larger class of single-variable equations for which the consistency problem is decidable.

Define $\mathcal{V} : (\tilde{A} \cup \{x, x^{-1}\})^* \to FIM(x)$ by $\mathcal{V}(a) = 1$ if $a \in \tilde{A}$ and $\mathcal{V}(x) = x$. Thus if $u = w_1 x^{\epsilon_1} w_2 \cdots x^{\epsilon_n} w_n$ where $w_j \in \tilde{A}^*$ for $j = 1, \ldots, n$ and $\epsilon_i = \pm 1$, then $\mathcal{V}(u) = x^{\epsilon_1} x^{\epsilon_2} \cdots x^{\epsilon_n}$ in $FIM(x)$.

Theorem 6 *Let $\mathcal{C}$ be the class of single-variable equations of the form $u = v$ in $FIM(A)$ for which $\mathcal{V}(u) \neq \mathcal{V}(v)$ as elements of $FIM(x)$. Then the consistency problem for $\mathcal{C}$ is decidable. That is there is an algorithm that on input an equation $u = v$ in $\mathcal{C}$, produces the output "Yes" if this equation is consistent and "No" if it is inconsistent.*

Proof: Since $u, v \in (\tilde{A} \cup \{x, x^{-1}\})^*$ then we have

$$u = u_1 x^{\epsilon_1} u_2 x^{\epsilon_2} \cdots x^{\epsilon_{n-1}} u_n \text{ and } v = v_1 x^{\delta_1} v_2 x^{\delta_2} \cdots x^{\delta_{t-1}} v_t \tag{11}$$

where $u_i, v_j \in \tilde{A}^*$ for $1 \leq i \leq n$, $1 \leq j \leq t$ and $\epsilon_i, \delta_j \in \{1, -1\}$ for $1 \leq i \leq n$, $1 \leq j \leq t$.

If there are only finitely many solutions to $u = v$ in $FG(A)$ then there is an effective bound on the length of all such solutions, by Theorem 5. Thus in this case we can decide whether the equation $u = v$ is consistent in $FIM(A)$ since the extendibility problem is decidable. So suppose that there are infinitely many solutions to $u = v$ in $FG(A)$. Then again by Theorem 5, we may effectively find finitely many parametric words of the form $w_1 w_2^{\alpha} w_3$ such that $\phi_m(x) = w_1 w_2^m w_3$ is a solution to $u = v$ in $FG(A)$ for any integer m. We will show that there are only finitely many values of the integer m (for each such parametric word) such that ϕ_m can possibly extend to a solution to $u = v$ in $FIM(A)$. Again, since the extendibility problem is decidable, this will enable us to decide whether the equation $u = v$ is consistent in $FIM(A)$.

Recall that the free group on x $FG(x)$ is isomorphic to the additive group $\mathbf{Z}$ of integers. Thus every Munn tree in the Cayley graph of $FG(x)$ can be viewed as an integer interval containing 0

$$\{i \in \mathbf{Z} \mid p \leq i \leq s\} = [p, s] \text{ where } p \leq 0 \leq s,$$

and if $w \in FIM(x)$ the rooted tree $(MT(w), r(w))$ can be identified with the triple (p, q, s) with $-p, s \in \mathbf{N}$ and $p \leq q \leq s$. The initial root of the corresponding birooted tree is 0 and $r(w) = x^q$.

Let $\mathcal{V}(u)$ be identified with the triple (l_u, n_u, r_u) and let $\mathcal{V}(v)$ be identified with the triple (l_v, n_v, r_v). Since $\mathcal{V}(u) \neq \mathcal{V}(v)$ in $FIM(x)$ it follows that $(l_u, n_u, r_u) \neq (l_v, n_v, r_v)$. If $n_u \neq n_v$ then the sum of the exponents of the variable x in u is not equal to the sum of the exponents of the variable x in v. But then from [20] there exists at most one solution to $u = v$ in $FG(A)$. Since we are assuming that $u = v$ has infinitely many solutions, this does not occur. Thus $n_u = n_v$ and since $(l_u, n_u, r_u) \neq (l_v, n_v, r_v)$ then either $l_u \neq l_v$ or $r_u \neq r_v$. Without loss of generality assume that $r_u \neq r_v$ and $r_u > r_v$. (A dual argument will apply to the case when $l_u \neq l_v$.) Note that there exists i such that $r_u = \sum_{k=1}^{i} \epsilon_k > 0$.

We will associate an integer to each reduced prefix

$$r(\phi_m(u_1 x^{\epsilon_1} u_2)), r(\phi_m(u_1 x^{\epsilon_1} u_2 x^{\epsilon_2} u_3)), \cdots, r(\phi_m(u_1 x^{\epsilon_1} u_2 x^{\epsilon_2} \cdots x^{\epsilon_{n-1}} u_n))$$

of $\phi_m(u)$. Similarly, we will do the same for the prefixes

$$r(\phi_m(v_1 x^{\delta_1} v_2)), r(\phi_m(v_1 x^{\delta_1} v_2 x^{\delta_2} v_3)), \cdots, r(\phi_m(v_1 x^{\delta_1} v_2 x^{\delta_2} \cdots x^{\delta_{t-1}} v_t))$$

of $\phi_m(v)$.

Let d be the maximum diameter of the Munn trees of any of the words
$u_1 w_1, u_1 w_3^{-1}, v_1 w_1, v_1 w_3^{-1}, w_3 u_{n+1}, w_1^{-1} u_{n+1}, w_3 v_{t+1}, w_1^{-1} v_{t+1}$, or
$w_3 u_i w_1, w_3 u_i w_3^{-1}, w_1^{-1} u_i w_1, w_1^{-1} u_i w_3^{-1}, w_3 v_j w_1, w_3 v_j w_3^{-1}, w_1^{-1} v_j w_1, w_1^{-1} u v_j w_3^{-1}$
for $i = 1, \ldots, n$ and $j = 1, \ldots, t$, and let $D = d + 1$. Choose m so that $|m| > 4MD$, where M is the maximum of n and t.

We define a function Ψ from the set of reduced words in $\tilde{A}^*$ to $\mathbf{Z}$ in the following manner. Let $z \in \tilde{A}^*$ be a reduced word. Write $z = z_1 w_2^{k_1} z_2 w_2^{k_2} z_3 \ldots z_{p-1} w_2^{k_p} z_p$ where no word z_i contains $w_2^{\pm 1}$ as a subword, the k_i are non-zero integers, and z_i is non-empty for $2 \leq i \leq p-1$. For each i let $|k_i| = s_i |m| + r_i$ where $0 \leq r_i < |m|$. If $r_i \leq |m| - 2MD$, define $\alpha_i = s_i$: if $r_i > |m| - 2MD$ define $\alpha_i = s_i + 1$. Then define $\Psi(z) = \sum_i \alpha_i \frac{k_i}{|k_i|}$. Note that $\Psi(z) = 0$ if z has no subword w_2^k with $|k| > |m| - 2MD$.

We will now show that for each reduced prefix,

$$\Psi(r(\phi_m(u_1 x^{\epsilon_1} u_2 x^{\epsilon_2} \cdots x^{\epsilon_i} u_{i+1}))) = \sum_{j=1}^{i} \epsilon_j \tag{12}$$

and

$$\Psi(r(\phi_m(v_1 x^{\delta_1} v_2 x^{\delta_2} \cdots x^{\delta_i} v_{i+1}))) = \sum_{j=1}^{i} \delta_j \tag{13}$$

Let $p_i = \phi_m(u_1 x^{\epsilon_1} u_2 x^{\epsilon_2} \cdots x^{\epsilon_i} u_{i+1})$. Then write

$$p_i = u_1' w_2^{m\epsilon_1} u_2' w_2^{m\epsilon_2} \dots u_i' w_2^{m\epsilon_i} u_{i+1}' \tag{14}$$

where u_1' is $u_1 w_1$ if $\epsilon_1 = 1$ or $u_1 w_3^{-1}$ if $\epsilon_1 = -1$, and the u_i' are defined similarly for $i > 1$. We have $r(u_1') = u_1'' w_2^{c_1}$ where c_1 is an integer (possibly zero), $|c_1| < D$ and the word $u_1'' w_2^{c_1}$ is reduced as written. Similarly, we have reduced words $r(u_j') = w_2^{d_{j-1}} u_j'' w_2^{c_j}$ for $j = 1, \dots i$ and $r(u_{i+1}') = w_2^{d_i} u_{i+1}''$, where each c_j, d_j is an integer (possibly zero) with $|c_j|, |d_j| < D$. The (possibly empty) words u_j'' are reduced and contain no subword of the form w_2^k where $|k| \geq D$, so no word u_j'' contains a subword of the form w_2^k where $|k| > |m| - 2MD$, by choice of m. It follows from (14) that

$$r(p_i) = u_1'' w_2^{m\epsilon_1 + c_1 + d_1} u_2'' w_2^{m\epsilon_2 + c_2 + d_2} \dots u_i'' w_2^{m\epsilon_i + c_i + d_i} u_{i+1}''. \tag{15}$$

Suppose that $u_j'' = 1$ for $j = s, s+1, \dots, s+p$ (some s, p with $2 \leq s \leq s+p \leq i$) Note that in this case $r(p_i)$ contains the corresponding subword w_2^k where $k = \sum_{j=s-1}^{s+p} (m\epsilon_j + c_j + d_j)$. Denote $|\sum_{j=s-1}^{s+p} \epsilon_j|$ by E. Now $E|m| - 2(p+1)D < |k| < E|m| + 2(p+1)D$, so $E|m| - 2MD < |k| < E|m| + 2MD$. If $|k| \geq E|m|$ then we have $|k| = E|m| + r$ where $r < 2MD < |m| - 2MD$ since $|m| > 4MD$. So in this case, $\Psi(w_2^{|k|}) = E$, and so $\Psi(w_2^k) = \sum_{j=s-1}^{s+p} \epsilon_j$. If $|k| < E|m|$ then we have $|k| = (E-1)|m| + r$ where $r = |k| - (E-1)|m| > |m| - 2MD$ since $E|m| - 2MD < |k|$. So again we have $\Psi(w_2^{|k|}) = E$ and $\Psi(w_2^k) = \sum_{j=s-1}^{s+p} \epsilon_j$.

In general, there is some (possibly empty) sequence of integers $i_1, i_2, \dots i_f$ with $1 \leq i_1 < i_2 < \dots < i_f \leq i+1$ such that $u_{i_j}'' \neq 1$ for each j, but all other words u_j'' (for $1 \leq j \leq i+1$) are empty. We must have $r(p_i) = u_1'' w_2^{k_1} u_{i_1}'' w_2^{k_2} u_{i_2}'' \dots w_2^{k_f} u_{i+1}''$, where this word is reduced as written and where each k_j is the sum over the appropriate range on the subscript e of terms of the form $m\epsilon_e + c_e + d_e$. By the argument above, $\Psi(w_2^{k_j}) = (\sum_e \epsilon_e)$, where the sum is taken over the same range of values of e. Since none of the words u_{i_j} is empty or contains a subword of the form w_2^k with $|k| > |m| - 2MD$, we clearly have $\Psi(r(p_i)) = \sum_{j=1}^{f} \Psi(w_2^{k_j})$. This, combined with the result in the previous sentence, shows that $\Psi(r(p_i)) = \sum_{j=1}^{i} \epsilon_j$, as required. The proof of (13) is similar.

We now claim that if $|m| > 4MD$, then the solution ϕ_m of $u = v$ in $FG(A)$ does not extend to a solution in $FIM(A)$. This will show that the consistency problem reduces to checking extendibility of finitely many solutions to $u = v$ in $FG(A)$, which is decidable by Theorem 3.

Define a *branch* of a Munn tree $MT(w)$ of some word w to be a path $\mathbf{b}$ in $MT(w)$ starting at the initial vertex of $MT(w)$ and labeled by a reduced word $l(\mathbf{b})$. Note that according to this definition, a branch of $MT(w)$ does not necessarily end at an extremal vertex of the tree $MT(w)$. It is clear that if $MT(w_1) = MT(w_2)$ for two words w_1 and w_2, and if k is any integer such that $\Psi(l(\mathbf{b_1})) = k$ for some branch $\mathbf{b_1}$ of $MT(w_1)$, then there must exist a branch $\mathbf{b_2}$ of $MT(w_2)$ such that $\Psi(l(\mathbf{b_2})) = k$ (and vice-versa).

We claim that for any word w, if w' labels a branch of $MT(w)$, then there must be some prefix p of the word w such that $r(p) = w'$. To see this, note that if w' is the reduced word $a_1 a_2 \dots a_s$, then the edges corresponding to $a_1, a_2, \dots a_s$ must eventually be traced out (in the order listed) in $MT(w)$ as we read the word w in the Cayley graph of $FG(A)$. Thus there must exist words $w_1, w_2, \dots, w_{s+1}$ such that $w = w_1 a_1 w_2 a_2 \dots a_s w_{s+1}$ in $\tilde{A}^*$ where each w_i $(i = 1, \dots, s)$ is a Dyck word. It follows that $w' = r(p)$ where $p = w_1 a_1 w_2 a_2 \dots a_s$ is a prefix of w.

Since $\mathcal{V}(u) \neq \mathcal{V}(v)$ as elements of $FIM(x)$ and since we assumed that $r_u > r_v \geq 0$, it follows that there exists i such that $r_u = \sum_{k=1}^{i} \epsilon_k > \sum_{k=1}^{j} \delta_k$ for all $j \leq t-1$. Choose i to be maximal with this property. Clearly we must have $\epsilon_i = 1$. Let $F = \Psi(r(\phi_m(u_1 x^{\epsilon_1} u_2 x^{\epsilon_2} \cdots x^{\epsilon_i} u_{i+1})))$. By (12) and (13), we have

$$F > \Psi(r(\phi_m(v_1 x^{\delta_1} v_2 x^{\delta_2} \cdots x^{\delta_k} v_{k+1}))),\ k = 1, \ldots, t-1 \tag{16}$$

Equation (16) shows that there is a branch $\mathbf{b_1}$ in $MT(\phi_m(u))$ such that $\Psi(l(\mathbf{b_1})) = F$, but we will see that it also implies that there is no branch $\mathbf{b_2}$ in $MT(\phi_m(v))$ for which $\Psi(l(\mathbf{b_2})) = F$. To see this, note first that (13) shows that if w is any prefix of v, then the value of $\Psi(r(\phi_m(w)))$ depends only on the largest integer k for which $u_1 x^{\epsilon_1} \ldots x^{\epsilon_k}$ is a prefix of w. Thus there is no prefix w of $\phi_m(v)$ for which $\Psi(r(w)) = F$. By the claim above, this means that there is no branch $\mathbf{b_2}$ of $MT(\phi_m(v))$ such that $\Psi(l(\mathbf{b_2})) = F$. This in turn implies that $\phi_m(u) \neq \phi_m(v)$ in $FIM(A)$.

Now assume that ϕ_m extends to a solution ψ of $u = v$ in $FIM(A)$ where $\psi(x) = e\phi_m(x)$ for some idempotent e. Choose a branch $\mathbf{b}$ of $MT(e)$ for which $\Psi(l(\mathbf{b})))$ is maximal. Certainly $\Psi(l(\mathbf{b}))) \geq 0$.

If $i < n-1$, then $\epsilon_{i+1} = -1$. In this case let α be the designated vertex of $MT(\phi(u))$ corresponding to the initial vertex of the edge labeled by $x^{\epsilon_{i+1}}$ in $MT(u)$. When we adjoin a copy of $MT(e)$ to $MT(\phi(u))$ at the vertex α we see that we can construct a branch $\mathbf{b_3}$ of $MT(\psi_m(u))$ such that $\Psi(l(\mathbf{b_3})) = \sum_{k=1}^{i} \epsilon_k + \Psi(l(\mathbf{b}))$, so there must be some branch $\mathbf{b_4}$ of $MT(\psi(v))$ with $\Psi(l(\mathbf{b_4})) = \sum_{k=1}^{i} \epsilon_k + \Psi(l(\mathbf{b}))$. But this is impossible since $\Psi(r(w)) < \sum_{k=1}^{i} \epsilon_k$ for all prefixes w of v by (16).

If $i = n-1$, then $\epsilon_{n-1} = 1$. In this case let α be the designated vertex of $MT(\phi(u))$ corresponding to the initial vertex of the edge labeled by $x^{\epsilon_{n-1}}$ in $MT(u)$. When we adjoin a copy of $MT(e)$ to $MT(\phi(u))$ at the vertex α, we see that we can construct a branch $\mathbf{b_5}$ of $MT(\psi_m(u))$ such that $\Psi(l(\mathbf{b_5})) = \sum_{k=1}^{n-2} \epsilon_k + L$, where L is the maximum of 1 and $\Psi(l(\mathbf{b}))$. But again there is no such branch in $MT(\psi(v))$ by (16), so we have a contradiction. Hence ϕ_m does not extend to any solution to $u = v$ in $FIM(A)$ if $|m| > 4MD$. Decidability of the consistency problem follows from Theorem 3. ■

In order to study the consistency problem for equations $u = v$ for which $\mathcal{V}(u) = \mathcal{V}(v)$, it is convenient to note the following lemma.

Lemma 4 *Let $u = v$ be an arbitrary equation in $FIM(A)$ and let $\phi : X \to \tilde{A}^*$ be a solution to $u = v$ in $FG(A)$. If the set of designated vertices in $MT(\phi(u))$ is equal to the set of designated vertices in $MT(\phi(v))$, then ϕ extends to a solution in $FIM(A)$.*

Proof: Let $\{w_1, w_2, \cdots, w_k\}$ be the set of designated vertices in $MT(\phi(u))$ (and in $MT(\phi(v))$). View each w_j as a reduced word in $\tilde{A}^*$. Let

$$g = \phi(u)\phi(u)^{-1}\phi(v)\phi(v)^{-1}$$

and note that $MT(g) = MT(\phi(u)) \cup MT(\phi(v))$. Let

$$E = (w_1^{-1} g w_1)(w_2^{-1} g w_2) \cdots (w_k^{-1} g w_k),$$

and let $T = MT(E)$.

Extend the map ϕ by defining

$$\psi : X \to \tilde{A}^* \text{ by } \psi(x_i) = E\phi(x_i)$$

for each variable x_i in the content of u and in the content of v.

From the definition of a designated vertex, it follows that at each vertex labeled by w_j ($j = 1, \ldots, k$), the tree

$$w_jT = \cup_{i=1}^k w_jMT(w_i^{-1}gw_i)$$

is a subtree of $MT(\psi(u))$ and that in fact

$$MT(\psi(u)) = w_1T \cup w_2T \ldots w_kT \cup MT(\phi(u)).$$

If z is an element of $FG(A)$ that is a vertex in $MT(\phi(u)$, then $w_jw_j^{-1}zz^{-1}w_j$ labels a path in $w_j(w_j^{-1}MT(g)w_j) \subseteq w_jT$, starting at 1 and ending at w_j. But then $w_j^{-1}zz^{-1}w_j$ labels a path in w_jT, starting and ending at w_j, so z is a vertex in w_jT. It follows that $MT(\phi(u)) \subseteq w_jT$ for all j, and so $MT(\psi(u)) = w_1T \cup w_2T \ldots w_kT$. Similarly, $MT(\psi(v)) = w_1T \cup w_2T \ldots w_kT$, and so $MT(\psi(u)) = MT(\psi(v))$, whence ψ is a solution to $u = v$ in $FIM(A)$ that extends ϕ. ■

We will introduce the concept of a standard factorization or a Choffrut factorization of an element in $FIM(A)$. Let $u \in FIM(A)$. A *reduced factorization* of $r(u)$ is $r(u) = u_1u_2\cdots u_n$ where $u_1, u_2, \cdots, u_n$ are nonempty words and $u_1u_2\cdots u_n$ is reduced as written. For $i = 1, 2, \cdots, n$ set $u_i = u_i'c_i$ where c_i is the last letter of u_i. Note that while $r(u)$ has many different possible reduced factorizations, there are only finitely many such factorizations.

Theorem 7 (Choffrut [3]) *Let $u \in FIM(A)$ and let $u_1u_2\cdots u_n$ be a reduced factorization of $r(u)$. Then there exists a unique factorization (which we call a Choffrut factorization) of u in $FIM(A)$ such that*

$$u = e_0u_1e_1u_2\cdots e_{n-1}u_ne_n$$

where $e_0, e_1, e_2, \cdots, e_n$ are idempotents satisfying the conditions:

$$\begin{array}{ll} (i) & \text{for all } i = 1, 2, \cdots, n,\ u_iu_i^{-1} \not\geq e_{i-1} \text{ and } u_i'u_i'^{-1} \geq e_{i-1} \\ (ii) & \text{for all } i = 1, 2, \cdots, n,\ c_i^{-1}c_i \not\geq e_i. \end{array}$$

For more information on this decomposition, see [3].

The following theorem will use the well known parameterization of the solution set to the equation $x_1x_2 = x_2x_3$ in the free monoid (see, for example, [9]). All solutions $\phi : X \to \tilde{A}^*$ to this equation can be parameterized as:

$$\begin{array}{rcl} \phi(x_1) & = & rs \\ \phi(x_2) & = & r(sr)^m \\ \phi(x_3) & = & sr. \end{array}$$

where $r, s \in \tilde{A}^*$ and $m \in \mathbf{N}$.

Theorem 8 *The consistency problem for equations of the form $u_1x^{\zeta_1}u_2 = v_1x^{\zeta_2}v_2$ where $u_i, v_i \in (A \cup A^{-1})^*$ and $\zeta_i = \pm 1$ for $i = 1, 2$ in $FIM(A)$ is decidable.*

Proof: The extendibility problem for $u_1x^{\zeta_1}u_2 = v_1x^{\zeta_2}v_2$ is decidable by Theorem 3. If $\zeta_1 \neq \zeta_2$ then there is at most one solution to $u_1x^{\zeta_1}u_2 = v_1x^{\zeta_2}v_2$ in the setting of $FG(A)$, which implies that the consistency problem is decidable.

We will assume that $\zeta_1 = \zeta_2$. Without loss of generality assume that $\zeta_1 = \zeta_2 = 1$. The case when $\zeta_1 = \zeta_2 = -1$ will follow similarly. If there are finitely many solutions to the equation $u_1xu_2 = v_1xv_2$ in $FG(A)$ then the consistency problem is decidable for $u_1xu_2 = v_1xv_2$ in $FIM(A)$ since the extendibility problem is decidable.

So we suppose that $u_1xu_2 = v_1xv_2$ has infinitely many solutions in $FG(A)$. From Theorem 5 we know that there exists a finite set of parametric words defining the solution set to this equation in $FG(A)$. Choose one such parametric word: there are corresponding reduced words w_1, w_2 and w_3 so that $w_1w_2^{\pm 1}w_3$ is reduced as written, w_2 is cyclically reduced and primitive, and $w_1w_2^nw_3$ is a solution in $FG(A)$ for all $n \in \mathbf{Z}$.

Let $E = u_1u_1^{-1}u_2u_2^{-1}v_1v_1^{-1}v_2v_2^{-1}$. Assume that for some value of n, the solution $w_1w_2^nw_3$ extends to a solution to the equation in $FIM(A)$. Without loss of generality we may assume that $n \geq 0$ (Replace w_2 by w_2^{-1} if necessary.) Choose $N \in \mathbf{N}$ minimal such that $\phi(x) = w_1w_2^Nw_3$ extends to a solution $\psi(x)$ in $FIM(A)$. We will show that

$$N < 3Diam(MT(E)) + 2.$$

It follows that we need to check extendability of only finitely many solutions to $u = v$ in $FG(A)$. By Theorem 3, this implies that the consistency problem is decidable.

Suppose on the contrary that

$$N \geq 3Diam(MT(E)) + 2.$$

Factor $\psi(x)$ using the Chouffrut factorization based on the reduced word $w_1w_2^Nw_3$ to get

$$\psi(x) = e_{-1}w_1e_0w_2e_1 \cdots e_{N-1}w_2e_Nw_3e_{N+1}. \tag{17}$$

Let α_u [resp. α_v] be the largest integer k such that $u_1w_1w_2^k$ [resp. $v_1w_1w_2^k$] labels a path starting at 1 in $MT(u_1w_1)$ [resp. $MT(v_1w_1)$]. Dually, let β_u [resp. β_v] denote the largest integer k such that $u_2^{-1}w_3^{-1}w_2^{-k}$ [resp. $v_2^{-1}w_3^{-1}w_2^{-k}$] labels a path starting at 1 in $MT(u_2^{-1}w_3^{-1})$ [resp. $MT(v_2^{-1}w_3^{-1})$].

Let $\alpha = max\{\alpha_u, \alpha_v\} + 1$ and $\beta = max\{\beta_u, \beta_v\} + 1$.

Thus by choice of N, the length of the geodesic $[w_2^{\alpha+1}, w_2^{N-\beta-1}]$ in $MT(\phi(u)) \cap MT(\phi(v))$ is greater than the diameter of $MT(E)$. Consider a Choffrut factorization of $\psi(u) = \psi(v)$ relative to the reduced factorization

$$r(u_1w_1w_2^{\alpha})\underbrace{w_2 \cdots w_2}_{N-(\alpha+\beta)}r(w_2^{\beta}w_3u_2) \tag{18}$$

of $r(\phi(u)) = r(\phi(v)$. We will get

$$\psi(u) = \psi(v) = f_0r(u_1w_1w_2^{\alpha})f_1w_2f_2 \cdots f_{N-(\alpha+\beta)}w_2f_{N-(\alpha+\beta+1)}r(w_2^{\beta}w_3u_2)f_{N-(\alpha+\beta+2)} \tag{19}$$

where f_i are idempotents for $0 \leq i \leq N - (\alpha + \beta + 2)$.

Either $r(u_1) = r(v_1)$ or $r(u_1) \neq r(v_1)$ in $FG(A)$. When $r(u_1) = r(v_1)$, then the designated u-vertex and designated v-vertex are the same and so by Lemma 4 the equation $u = v$ is consistent, so we may suppose that $r(u_1) \neq r(v_1)$ in $FG(A)$. But this implies that $r(u_1 w_1 w_2^{\alpha}) \neq r(v_1 w_1 w_2^{\alpha})$ in $\tilde{A}^*$. (In general if $r(u_1 z) = r(v_1 z)$ in $\tilde{A}^*$, then $u_1 z = v_1 z$ in $FG(A)$, so we obtain $u_1 = v_1$ in $FG(A)$, a contradiction.) Thus the reduced factorizations (18) and

$$r(v_1 w_1 w_2^{\alpha}) \underbrace{w_2 \cdots w_2}_{N-(\alpha+\beta)} r(w_2^{\beta} w_3 v_2) \tag{20}$$

of $r(\phi(u)) = r(\phi(v))$ are not identical.

Without loss of generality assume that $r(v_1 w_1 w_2^{\alpha})$ is a proper initial segment of $r(u_1 w_1 w_2^{\alpha})$. Then there exists $w' \in \tilde{A}^*$ such that $r(u_1 w_1 w_2^{\alpha}) \equiv r(v_1 w_1 w_2^{\alpha}) w'$ in $\tilde{A}^*$. By choice of N, w' is a proper initial segment of $w_2^{N-(\alpha+\beta})$ and we may write $w' w_2^{N-(\alpha+\beta)} \equiv w_2^{N-(\alpha+\beta)} w''$ for some $w'' \in \tilde{A}^*$. By examining overlaps between the two factorizations of the word above, we see that if w' is not a power of w_2, then there must exist non-trivial words $q, t, q' \in \tilde{A}^*$ such that $w_2 \equiv qt \equiv tq'$, where q' is also a prefix of w_2. But $|q| + |t| = |w_2| = |t| + |q'|$, so $|q| = |q'|$, and since both are prefixes of w_2 this forces $q \equiv q'$. So $w_2 \equiv qt \equiv tq$. But it is well-known (see [9]) that this implies that $q \equiv p^i$ and $t \equiv p^j$ for some word p, which contradicts the fact that w_2 is primitive. Hence w' must be a power of w_2, say $w' \equiv w_2^d$ for some $d \geq 1$. A similar argument shows that $w'' \equiv w_2^e$ for some $e > 1$. Note that $d, e < \alpha$.

We now compare the Choffrut factorizations of $\psi(u) = \psi(v)$ relative to the factorizations (18) and (20), and the Choffrut Factorization (17) of $\psi(x)$. Note that the observation above about w' and w'' implies that the segment of $N-(\alpha+\beta+e)$ occurrences of w_2 occurring after $r(u_1 w_1 w_2^{\alpha})$ in (18) coincides with the segment of $N-(\alpha+\beta+e)$ occurrences of w_2 occurring after $r(v_1 w_1 w_2^{\alpha}) w_2^d$ in the factorization (20). Let $k = N - (\alpha + \beta + 1)$ and consider $f_1 w_2 f_2 \cdots f_k w_2$ as being a word in the free monoid generated by $\{f_1, f_2, \cdots, f_k, w_2\}$. We will denote this free monoid by F.

By definition of α we must have

$$e_{\alpha+1} w_2 e_{\alpha+2} w_2 e_{\alpha+3} w_2 \cdots e_{\alpha+k} w_2 = f_1 w_2 f_2 \cdots f_k w_2$$

as words in F. But we also must have

$$e_{\alpha+d+1} w_2 e_{\alpha+d+2} w_2 e_{\alpha+d+3} w_2 \cdots e_{\alpha+k-d} w_2 = f_1 w_2 f_2 \cdots f_k w_2$$

as words in F.

This implies that

$$e_{\alpha+1} e_{\alpha+2} e_{\alpha+3} \cdots e_{\alpha+k-d} = e_{\alpha+d+1} e_{\alpha+d+2} e_{\alpha+d+3} \cdots e_{\alpha+k}.$$

Since $d < \alpha$ then $\alpha + k - d > 0$. Denote this sequence by X. It follows that

$$(e_{\alpha+1} e_{\alpha+2} \cdots e_{\alpha+d}) X = X (e_{\alpha+k+1-d} e_{\alpha+k+2-d} \cdots e_{\alpha+k}).$$

We will denote $(e_{\alpha+1} \cdots e_{\alpha+d})$ by Y and the sequence $(e_{\alpha+k+1-d} e_{\alpha+k+2-d} \cdots e_{\alpha+k})$ by Z.

We now consider the equation

$$YX = XZ$$

in the setting of the free monoid on the alphabet $\{e_{\alpha+1}, e_{\alpha+2}, \ldots, e_{\alpha+k}\}$. Recall the equation $YX = XZ$ has a parameterized solution of the form

$$Y = UV, Z = VU \text{ and } X = U(VU)^m$$

where $m \in \mathbf{N}$. It follows that there exists $l \in \mathbf{N}$ where $\alpha + 1 < l < \alpha + k - d$ and $U = e_{\alpha+1}e_{\alpha+2}\cdots e_l$.Then the equation $YU = UZ$ holds in the free monoid on the alphabet $\{e_{\alpha+1}, e_{\alpha+2}, \ldots, e_{\alpha+k}\}$.

If $\phi(x) = w_1 w_2^N w_3$ is a solution to the equation in $FG(A)$ then $\phi'(x) = w_1 w_2^{N-l'} w_3$ where $l' = m|UV|$ is also a solution to the equation in $FG(A)$ by Theorem 5. Now $\phi(x) = w_1 w_2^N w_3$ extends to a solution

$$\psi(x) = e_{-1}w_1e_0w_2e_1\cdots w_2e_{\alpha+1}\cdots w_2e_{\alpha+k}w_2e_{\alpha+k+1}\cdots e_{N-1}w_2e_Nw_3e_{N+1}.$$

in $FIM(A)$.

We then consider an extension of $\phi'(x)$ defined by

$$\psi'(x) = e_{-1}w_1e_0w_2e_1\cdots e_{\alpha+1}w_2e_{\alpha+2}\cdots w_2e_lw_2e_{\alpha+k+1}\cdots w_2e_{N-1}w_2e_Nw_3e_{N+1}.$$

Since

$$e_{\alpha+1}w_2e_{\alpha+2}w_2\cdots w_2e_lw_2 = f_1w_2f_2w_2\cdots w_2f_lw_2$$

and

$$e_{\alpha+d+1}w_2e_{\alpha+d+2}w_2\cdots w_2e_{l+d}w_2 = f_1w_2f_2w_2\cdots w_2f_lw_2$$

then

$$f_0r(u_1w_1w_2^{\alpha})f_1w_2f_2\cdots f_lw_2f_{N-(\alpha+\beta+1)}r(w_2^{N-\beta}w_3u_2)f_{N-(\alpha+\beta+2)}$$

corresponds to the elements $\psi'(u)$ and $\psi'(v)$ in $FIM(A)$. Thus this substitution is a solution in FIM(A). Since $N - l' < N$, this contradicts the minimality of N. Hence if there is any integer such that $w_1w_2^nw_3$ extends to a solution in $FIM(A)$, then there must be such an integer n with $n < 3Diam(MT(E)) + 2$. By Theorem 3, this implies that the consistency problem is decidable. ■

It seems plausible that the argument involved in the proof of this theorem may be extended to show decidability of the consistency problem for *all* single-variable equations in $FIM(A)$, but we have not carried out this technical argument.

References

[1] K. I. Appel, One-variable equations in free groups, *Proc. Amer. Math. Soc.* 19 (1968), 912–918.

[2] J. Barwise (Ed.) *Handbook of Mathematical Logic*, North Holland (1978).

[3] C. Choffrut, Conjugacy in free inverse monoids, *Int. J. Alg. Comp.* 3.2 (1993), 169–188.

[4] L. P. Comerford and C. C. Edmunds, Products of Commutators and Products of Squares in a Free Group, *Int. J. Alg. Comp.* 4.3 (1994) 469-480.

[5] T. Deis, *Equations in Free Inverse Monoids*, Ph.D. Thesis, Univ. of Nebraska (1999).

[6] C. Gutiérrez, Satisfiability of Equations in Free Groups is in PSPACE, *32nd Ann. ACM Symp. Theory Comput. (STOC'2000)*, ACM Press 2000.

[7] Mark. V. Lawson, *Inverse Semigroups; the theory of Partial Symmetries*, (World Scientific 1998).

[8] A. A. Lorents, Representations of sets of solutions of systems of equations with one unknown in a free group, *Dokl. Akad. Nauk. SSSR* 178 (1968), 290–292 (Russian).

[9] M. Lothaire, *Algebraic Combinatorics on Words*, (Cambridge University Press 2001).

[10] R. C. Lyndon, Equations in free groups, *Trans. Amer. Math. Soc.* 96 (1960), 445–457.

[11] G. S. Makanin, Equations in a Free Group, *Izv. Akad. Nauk. SSR, Ser. Math* 46 (1983) 1199-1273. English transl. in *Math. USSR Izv.* 21 (1983).

[12] G. S. Makanin, Problem of Solvability of Equations in Free Semigroup, *Math. Sbornik* 103 (1977) 147-236. English transl. in *Math. USSR Sbornik* 32 (1977).

[13] W. D. Munn, Free inverse semigroups, *Proc. London Math .Soc.* (3) 29 (1974), 385–404.

[14] A.L.T. Patterson, *Groupoids, Inverse Semigroups, and their* $\mathbf{C}^*$*-algebras*, (Birkhäuser 1998).

[15] M. Petrich, *Inverse semigroups*, (Wiley 1984).

[16] W. Plandowski, Satisfiability of Word Equations with Constants is in PSPACE, *Proc. 40th Ann. Symp. Found. Comput. Sci. (FOCS'99), IEEE Computer Society Press*, (1999) 495-500.

[17] M. O. Rabin, Decidability of Second Order Theories and Automata on Infinite Trees, *Trans. Amer. Math. Soc.* 141 (1969) 1-35.

[18] A. A. Razborov, On Systems of Equations in a Free Group, *Math. USSR-Izv.* 25 (1985) 115-162.

[19] B. V. Rozenblat, Diophantine theories of free inverse semigroups, *Sibirskii Mat. Zhurnal* (6) 26 (1986), 101–107 (Russian); english transl. in pp. 860–865.

[20] P. V. Silva, Word Equations and Inverse Monoid Presentations, in *Semigroups and Applications, Including Semigroup Rings*, ed. S. Kublanovsky, A. Mikhalev, P. Higgins, J. Ponizovskii, "Severny Ochag", St. Petersburg (1999).

Timothy Deis
Department of Mathematics
University of Wisconsin
Platteville, WI 53818, USA
deist@uwplatt.edu

John Meakin
Department of Mathematics and Statistics
University of Nebraska
Lincoln, NE 68588, USA
jmeakin@math.unl.edu

Géraud Sénizergues
LaBRI-CNRS
Université Bordeaux I Nouvelle
351, cours de la Libération
F-33405 Talence Cedex, France
ges@labri.fr