Satisfiability of Systems of Equations over Finite Monoids

Cristopher Moore[1], Pascal Tesson[2], and Denis Thérien[2*]

[1] University of New Mexico, Albuquerque and The Santa Fe Institute
moore@santafe.edu
[2] School of Computer Science, McGill University
{ptesso, denis}@cs.mcgill.ca

Abstract. We study the computational complexity of determining whether a systems of equations over a fixed finite monoid has a solution. In [6], it was shown that in the restricted case of groups the problem is tractable if the group is Abelian and NP-complete otherwise. We prove that in the general case, the problem is in P if the monoid is the direct product of an Abelian group and a commutative idempotent monoid, and is NP-complete if it does not divide such a monoid. In the restricted case where only constants appear on the right-hand side, we show that the problem is in P if the monoid is in the class $\mathbf{R}_1 \vee \mathbf{L}_1$, and is NP-complete otherwise.
Furthermore interesting connections to the well known Constraint Satisfiability Problem are uncovered and exploited.

1 Introduction

Using ideas and tools from algebraic automata theory, a number of algebraic characterizations of complexity classes have been uncovered ([1, 7] among many others). This has increased the importance of the study of problems whose computational complexity is parametrized by the properties of an underlying algebraic structure [2, 6, 12].

In [6], Goldmann and Russell studied the computational complexity of solving equations and systems of equations over a fixed finite group. These investigations were partly completed in [2] where we considered the more general problem of checking satisfiability of a single equation over a fixed finite monoid. Formally, an equation over the monoid M is given as: $c_0X_1c_1 \ldots X_kc_k = d_0Y_1d_1 \ldots Y_ld_l$ where the $c_i, d_i \in M$ are constants and the X_i's and Y_i's are variables.

The System of Equations Satisfiability problem for the finite monoid M (denoted EQN^*_M) is to determine whether a given system of equations over M has a solution. We also consider the restriction of EQN^*_M where the right-hand side of each equation is a constant, i.e. variables occur only on the left, and denote it $\mathrm{T-EQN}^*_M$ (where "T" stands for "target"). For groups these problems are equivalent, and it is shown in [6] that they are in P for Abelian groups but NP-complete otherwise.

* Research supported by NSERC, FCAR and the Von Humboldt Foundation.

Since this problem is understood for groups, we examine the complementary case where M contains no non-trivial groups, i.e. when it is aperiodic. We prove that a dichotomy exists for aperiodic M similar to that for groups, namely that EQN^*_M is in P if M is idempotent and commutative, and is NP-complete otherwise. For the case with only constants on the right, we show that $\mathrm{T-EQN}^*_M$ is tractable when M belongs to the class $\mathbf{R}_1 \vee \mathbf{L}_1$ (defined below), and is NP-complete when M is aperiodic but outside $\mathbf{R}_1 \vee \mathbf{L}_1$. For the general case, we show that EQN^*_M is NP-complete outside the class $\mathbf{J}_1 \vee \mathbf{G}_{\mathrm{com}}$ where $\mathbf{G}_{\mathrm{com}}$ is the variety of Abelian groups. Furthermore, we explore (and exploit) the relationship between this work and the extensively studied CONSTRAINT SATISFACTION PROBLEM [5, 10, 3].

2 Background and Notation

2.1 Finite Monoids

We review here some basic notions of finite semigroup theory. A detailed introduction on the material sketched here can be found in [4, 11].

A monoid M is a set with a binary associative operation and an identity. This operation defines a canonical surjective morphism $eval_M : M^* \to M$ by

$$eval_M(m_1 m_2 \ldots m_k) = m_1 \cdot m_2 \cdots m_k$$

which just maps a sequence of monoid elements to their product in M. Throughout this paper, M denotes a finite monoid.

We say that a monoid N divides M and write $N \prec M$ if N is the homomorphic image of a submonoid of M. A class of finite monoids $\mathcal{M}$ is said to be a (pseudo)-variety if it is closed under direct product and division. Varieties are the natural unit of classification of finite monoids.

A monoid M is *aperiodic* or group-free if no subset of it forms a non-trivial group or, equivalently, if it satisfies $m^{t+1} = m^t$ for some integer t and all $m \in M$. Any finite monoid divides a wreath product[1] of aperiodics and simple groups. Thus aperiodics and groups act as "building blocks" for all monoids, so they are the two special cases to consider first.

An element $m \in M$ is said to be *idempotent* if $m^2 = m$ and if this holds for all $m \in M$, we say that M is idempotent.

The *left (right) ideal* generated by an element x is the set Mx (resp. xM), and its *two-sided ideal* is the set MxM. The following equivalence relations on M, called Green's relations, ask whether two elements generate the same ideals.
- $x\mathcal{J}y$ if $MxM = MyM$
- $x\mathcal{L}y$ if $Mx = My$
- $x\mathcal{R}y$ if $xM = yM$
- $x\mathcal{H}y$ if both $x\mathcal{R}y$ and $x\mathcal{L}y$

It is known that $\mathcal{L}$ and $\mathcal{R}$ commute and that for finite monoids $\mathcal{L} \circ \mathcal{R} = \mathcal{J}$.

[1] See e.g. [11] for a formal definition

If we instead ask whether element's ideals are contained in one another, we can define natural pre-orders $\leq_{\mathcal{J}}, \leq_{\mathcal{R}}, \leq_{\mathcal{L}}$ on M with e.g. $x \leq_{\mathcal{J}} y$ iff $MxM \subseteq MyM$. We will say that "x is (strictly) $\mathcal{J}$-above y" if $x \geq_{\mathcal{J}} y$ (resp. $x >_{\mathcal{J}} y$), and so on. Note that $x \leq_{\mathcal{J}} y$ iff there exists u, v such that $x = uyv$ and similarly $x \leq_{\mathcal{R}} y$ iff there is u with $x = yu$. We will make extensive use of the following fact:

Lemma 1. *If $x\mathcal{J}y$ and x, y are $\mathcal{R}$-comparable (resp. $\mathcal{L}$-comparable), then $x\mathcal{R}y$ (resp. $x\mathcal{L}y$). Thus if $x\mathcal{J}y$ and x, y are both $\mathcal{R}$- and $\mathcal{L}$-comparable, then $x\mathcal{H}y$.*

Monoid M is said to be $\mathcal{J}$-trivial (resp. $\mathcal{R}$-, $\mathcal{L}$-, or $\mathcal{H}$-trivial) if $x\mathcal{J}y$ (resp. $x\mathcal{R}y$, $x\mathcal{L}y$ or $x\mathcal{H}y$) implies $x = y$; that is, if each of the relevant equivalence classes contain exactly one element. A $\mathcal{J}$-class (resp. $\mathcal{R}, \mathcal{L}, \mathcal{H}$-class) is *regular* if it contains an idempotent.

An $\mathcal{H}$ class contains at most one idempotent and is regular iff it forms a group whose identity element is the idempotent of the class. Conversely, any maximal subgroup of M is precisely a regular $\mathcal{H}$ class. Therefore, M is aperiodic iff it is $\mathcal{H}$-trivial. More generally, if all its $\mathcal{H}$-classes are regular then M is the union of its maximal subgroups, in which case it is called a *union of groups*. We denote by $\mathcal{R}_s$ (and similarly for $\mathcal{J}, \mathcal{H}, \mathcal{L}$) the $\mathcal{R}$-class of s. One can prove the following:

Lemma 2. *Let $s\mathcal{J}t$, then $st \in \mathcal{R}_s \cap \mathcal{L}_t$ if and only if $\mathcal{L}_s \cap \mathcal{R}_t$ contains an idempotent. Otherwise, we have $st <_{\mathcal{J}} s$.*

2.2 CSP

Let D be a finite domain and Γ be a finite set of relations on D. To each pair D, Γ corresponds a Constraint Satisfaction Problem (CSP). An instance of CSP is a list of constraints, i.e. of pairs $R_i(S_i)$ where $R_i \in \Gamma$ is a k-ary relation and S_i, the scope of R_i, is an ordered list of of k-variables (with possible repetitions) and we want to determine whether the variables can be assigned values in D such that each constraint is satisfied.

This class of combinatorial decision problems has received a lot of attention because of the wide variety of problems which it emcompasses. CSP is NP-complete in general, and one important question concerning CSP is to identify restrictions on Γ which make the problem tractable for a given domain D. Recently, tools from universal algebra [3], relational database theory [5], and group theory [5] have been used to identify tractable classes of CSP when the relations in Γ are closed under certain families of operations. In our context, the domain is the monoid and Γ is the set of constraints definable as equations over M.

The CSP problem on boolean domains is also known as Generalized Satisfiability and was studied by Schaefer in [13]. He proved a complete dichotomy and showed the problem to be NP-complete unless it was one of six tractable special cases: 2-SAT, 0-valid SAT, 1-valid SAT, affine-SAT, Horn-SAT and anti-Horn SAT. Affine-SAT is the case where each relation is the solution set of a system of equations over the cyclic group $\mathbb{Z}_2$. There exists a second 2-element monoid, namely U_1, the two element aperiodic having elements 1 and 0 and multiplication given by $0 \cdot 1 = 1 \cdot 0 = 0 \cdot 0 = 0$ and $1 \cdot 1 = 1$. The following provides a connection to our concern in this paper.

Lemma 3. *A boolean relation is Horn or anti-Horn, i.e. expressible as tuples satisfying a conjunction of disjuncts containing each at most one un-negated (resp. negated) variable, iff it is the set of solutions of a system of equations over U_1.*

Proof. Identify the element 1 of U_1 with TRUE and 0 with FALSE. Then the Horn clause $X_1 \wedge X_2 \wedge \ldots X_n \rightarrow Y$ is satisfied when one of the X_i's is FALSE or when all X_i's and Y are TRUE. These are exactly the tuples which satisfy the equation

$$X_1 X_2 \ldots X_n = X_1 \ldots X_n Y$$

over U_1.

Conversely, the equation $X_1 \ldots X_n = Y_1 \ldots Y_m$ corresponds to the Horn formula:

$$\bigwedge_{1 \leq i \leq m} (X_1 \wedge \ldots \wedge X_n \rightarrow Y_i) \wedge \bigwedge_{1 \leq i \leq n} (Y_1 \wedge \ldots \wedge Y_n \rightarrow X_i)$$

If on the other hand we choose to identify 1 with FALSE and 0 with TRUE, a similar argument shows the relationship of U_1 systems to anti-Horn formulas.

3 Hardness Results

In this section we show several hardness results for EQN^*_M. Note that the problem trivially lies in NP.

Theorem 1. *If M is aperiodic but not idempotent, then EQN^*_M is NP-complete.*

Proof. Let $m \neq m^2$ be a $\mathcal{J}$-maximal non-idempotent element of M.

We use the following reduction from 1-3SAT, where each clause consists of three literals, exactly one of which must be true. For each Boolean variable X_i, we use two variables x_i, $\overline{x}_i$ in M. For each i we have equations

$$(1)\ x_i \overline{x}_i = m \text{ and } \qquad (2)\ \overline{x}_i x_i = m$$

and for each clause of the formula, e.g. $(X_1 \vee \overline{X_2} \vee X_3)$ we add the equation

$$(3)\ x_1 \overline{x}_2 x_3 = m$$

Suppose that the original 1-in-3 SAT formula is satisfiable. Then for each i we set $x_i = m$ and $\overline{x}_i = 1$ whenever X_i is TRUE, and $x_i = 1$ and $\overline{x}_i = m$ whenever X_i is FALSE. It is easy to see that this satisfies the sets of Equations (1), (2) and (3), so these equations are satisfiable if the 1-in-3 SAT formula is.

Conversely, suppose that this system of equations is satisfable. Let us first look at the Equations (1) and (2). Note that x_i (resp. $\overline{x}_i$) is both $\mathcal{R}$-above and $\mathcal{L}$-above m. So if x_i (say) lies in m's $\mathcal{J}$-class, it must be $\mathcal{R}$ and $\mathcal{L}$ equivalent to m, hence, by Lemma 1, $\mathcal{H}$-equivalent and thus equal to m by aperiodicity. It follows that at least one of x_i or $\overline{x}_i$ lies strictly $\mathcal{J}$-above m; otherwise we would have $x_i = \overline{x}_i = m$, and since m is not idempotent this would violate Equation (1). Moreover, since m is $\mathcal{J}$-maximal among the non-idempotent elements of M, whichever one of x_i, $\overline{x}_i$ is strictly $\mathcal{J}$-above m must be some idempotent e.

Therefore, suppose $x_i = e$ where $e >_{\mathcal{J}} m$ is idempotent. Then Equation (1) gives us $m = e\overline{x}_i = ee\overline{x}_i = em$, and similarly (2) gives $m = me$. We cannot also have $\overline{x}_i >_{\mathcal{J}} m$, since then $\overline{x}_i$ would also be idempotent, and this leads to the contradiction $m^2 = e\overline{x}_i\overline{x}_ie = e\overline{x}_ie = em = m$. Thus $\overline{x}_i\mathcal{J}m$; by Lemma 1 this gives $\overline{x}_i\mathcal{H}m$, and so $\overline{x}_i = m$ from aperiodicity.

Similarly, if $\overline{x}_i >_{\mathcal{J}} m$ then $x_i = m$. So if we set X_i to TRUE when $x_i = m$ and FALSE when $\overline{x}_i = m$, Eqs (1) and (2) insure that our mapping between Boolean variables and variables in M is consistent, in the sense that for all i, exactly one of x_i, $\overline{x}_i$ is m and the other is an idempotent in a higher $\mathcal{J}$-class.

Finally, suppose that all 3 variables in Equation (3) have idempotent values. By a previous argument, these values fix m and so $m^2 = x_1\overline{x}_2x_3m = x_1\overline{x}_2m = x_1m = m$, a contradiction. We get a similar contradiction if two or more of the variables are set to m. Therefore (3) insures that exactly one literal in each clause is true, and the 1-in-3 SAT formula is satisfiable.

Theorem 2. *If M contains some irregular $\mathcal{H}$-class, i.e. if M is not a union of groups, then* EQN^*_M *is NP-complete.*

Proof. Let $m \neq m^2$ be an element of a $\mathcal{J}$-maximal irregular $\mathcal{H}$-class, i.e. such that any monoid element u with $u >_{\mathcal{J}} m$ lies in a regular $\mathcal{H}$-class. In particular, since regular $\mathcal{H}$-classes form groups, there must exist $q \geq 1$ such that for all such u we have u^q is idempotent and $u^{q+1} = u$.

We use the same reduction from 1-3SAT, namely: for each literal x_i and its complement $\overline{x}_i$, we add equations $x_i\overline{x}_i = m$ and $\overline{x}_ix_i = m$ and for each clause we add the equation e.g.

$$x_1\overline{x}_2x_3 = m$$

Once again, a satisfiable instance of 1-3SAT trivially gives rise to a satisfiable system of equations.

For the converse, Eqs. (1) and (2) show that if $x_i\mathcal{J}m$ then in fact $x_i\mathcal{H}m$, just as in Theorem 1. Since the $\mathcal{H}$-class of m contains no idempotent, Lemma 2 shows that the product of any two elements of $\mathcal{H}_m$ lies strictly $\mathcal{J}$-below m. Eqs. (1) and (2) thus force at least one of x_i, $\overline{x}_i$ to be strictly $\mathcal{J}$-above m.

Suppose both x_i and $\overline{x}_i$ are strictly $\mathcal{J}$-above m. Then x_i^q and $\overline{x}_i^q$ are idempotents, and $m = x_i\overline{x}_i = x_i\overline{x}_i^{q+1} = m\overline{x}_i^q$. Similarly $m = mx_i^q$. Moreover, x_i and $\overline{x}_i$ commute by Eqs. (1) and (2), so we get $m = mx_i^q\overline{x}_i^q = m(x_i\overline{x}_i)^q = m^{q+1} <_{\mathcal{J}} m$, a contradiction. Therefore, at least one of $x_i, \overline{x}_i$ must be $\mathcal{H}$-equivalent to m and the other fixes the $\mathcal{H}$-class of m, so if we identify true literals with variables taking a value $\mathcal{H}$-equivalent to m, we obtain a consistent truth assignment, and the same argument as before shows that exactly one literal in each clause corresponding to Equation (3) must be true.

Note that the reductions used to prove Theorems 1 and 2 actually show that $\mathrm{T-EQN}^*$ is NP-complete in both these cases.

Theorem 3. *If M is aperiodic and idempotent but is not commutative, then EQN^*_M is NP-complete.*

Proof. Let a and b be two non-commuting elements, $ab \neq ba$, such that a is a $\mathcal{J}$-maximal element which is not *central* in M (i.e. which does not commute with every element) and b is a $\mathcal{J}$-maximal element which does not commute with a. We reduce from 3-SAT. For each Boolean variable X_i, we create variables $x_i, \overline{x}_i, y_i, \overline{y}_i$ and equations

$$\begin{array}{llll} (1) & x_i\overline{x}_i = a & (2) & \overline{x}_i x_i = a \\ (3) & y_i\overline{y}_i = b & (4) & \overline{y}_i y_i = b \\ (5) & x_i\overline{y}_i = \overline{y}_i x_i & (6) & \overline{x}_i y_i = y_i \overline{x}_i \end{array}$$

Also, for each 3-SAT clause, e.g. $X_1 \vee \overline{X_2} \vee X_3$, we add an equation

$$x_1\overline{x}_2 x_3 = a$$

It is easy to verify that if the 3-SAT instance is satisfiable, the above system is satisfiable by setting $x = a$, $\overline{x} = 1$, $y = b$, and $\overline{y} = 1$ whenever X is TRUE, and $x = 1$, $\overline{x} = a$, $y = 1$, $\overline{y} = b$ whenever X is FALSE.

Conversely, suppose the system of equations is satisfiable. Eq. (1) shows that $x_i, \overline{x}_i \geq_{\mathcal{J}} a$. Since a and b don't commute, a cannot be the product of two elements commuting with b. However, any element strictly $\mathcal{J}$-above a is central and commutes with everything, so at least one of x_i, $\overline{x}_i$ must be $\mathcal{J}$-equivalent to a. Moreover, Eqs. (1) and (2) insure that x_i, $\overline{x}_i$ are both $\mathcal{L}$-above and $\mathcal{R}$-above a, so if $x_i \mathcal{J} a$ (say) we must also have $x \mathcal{H} a$ by Lemma 1 and so $x = a$ by aperiodicity. Thus at least one of x_i, $\overline{x}_i$ must be a, Similarly at least one of y_i, $\overline{y}_i$ must be b, since any elements strictly $\mathcal{J}$-above b commute with a.

If $x_i = a$, then $\overline{y}_i$ commutes with a by Eq. (5). Thus $\overline{y}_i \neq b$ must be strictly $\mathcal{J}$-above b. If $y_i = b$, then $\overline{x}_i$ commutes with b by Eq. (6), so $\overline{x}_i \neq a$ is strictly $\mathcal{J}$-above a. If we identify true literals with a and b and false ones with elements strictly $\mathcal{J}$-above a and b, this shows that our truth assignment is consistent.

Since every element strictly $\mathcal{J}$-above a is central but a is not, a cannot be a product of such elements. Therefore, at least one of the variables in $x_1\overline{x}_2x_3 = a$ must be a, so the 3-SAT clause is satisfiable.

Goldmann and Russell already showed that $\mathrm{T-EQN}^*$ is NP-complete for a non-Abelian group. We can extend this result to any M having an $\mathcal{H}$-class that forms a non-Abelian group G. Indeed, let e be the idempotent of this $\mathcal{H}$-class and consider a system S of equations over G with no variables on the right-hand side. By replacing each occurence of a variable x in S by exe, we obtain a system S' which is satisfiable over M if and only if S is satisfiable over G since for any $x \in M$, either exe is in G or exe is strictly $\mathcal{J}$-below the equation's target.

We can combine the results of this section into one algebraic condition on M insuring NP-completeness of EQN^*_M. Let $\mathbf{J}_1 \vee \mathbf{G}_{\mathrm{com}}$ be the variety generated by idempotent commutative monoids and Abelian groups. It can be shown that M belongs to $\mathbf{J}_1 \vee \mathbf{G}_{\mathrm{com}}$ if and only if it is a union of groups and is commutative.

Theorem 4. *If M is not in $\mathbf{J}_1 \vee \mathbf{G}_{\mathrm{com}}$, then EQN^*_M is NP-complete.*

Proof. By Theorem 2 and remarks following its proof, this is certainly true unless M is a union of maximal Abelian subgroups. Such a monoid can easily be shown to be commutative iff all four of Green's relations coincide in M.

If this is not the case, however, there must exist a $\mathcal{J}$-maximal $\mathcal{J}$-class containing two idempotents a, b which are not $\mathcal{H}$-equivalent and thus, by Lemma 2, do not commute. Moreover, because each $\mathcal{H}$-class is a subgroup, there exists some q such that for all $m \in M$, m^q is idempotent. We can therefore show NP-completeness using the reduction in the proof of Theorem 3 where we replace all occurences of a variable x by x^q.

4 Tractable cases

It is shown in [6] that EQN^* is in P for Abelian groups. This could also be proved using Theorem 33 of [5] which insures tractability of CSP over a finite group G when each relation of Γ is a coset[2] of G^k. For aperiodics, we have:

Theorem 5. *If M is idempotent and commutative, then EQN^*_M is in P.*

Proof. It is known that idempotent and commutative finite monoids are $\mathcal{J}$-trivial [4]. In particular, they always have a unique $\mathcal{J}$-minimal element 0 such that $0 \cdot x = x \cdot 0 = 0$ for all $x \in M$. Our algorithm looks for a "$\mathcal{J}$-minimal" solution to the system.

For each equation of our system in which the variable x_i appears only on one side, $ux_iv = E$ where x_i does not appear in E, we write an inequality $x_i \geq_{\mathcal{J}} E$. Note that any satisfying assigment must satisfy all these $\mathcal{J}$-inequalities.

Initially, we set all our variables to 0. We update these values in the following way: for each variable x_i, we set its new value to the least value satisfying all inequalities $x_i \geq_{\mathcal{J}} E$ where E is evaluated using the current values for $x_1, \ldots, x_n$. We iterate this until we reach a fixed point. It should be clear that no satisfying assignment of the system assigns to x_i a value which is smaller than this one in the $\mathcal{J}$-ordering. Moreover, this process terminates in at most $n \cdot |M|$ steps where n is the number of variables, since the new values are increasing in the $\mathcal{J}$-order.

If the assignment resulting from this process does not satisfy the system, then some equation is violated. For this to happen, there must be a constant appearing on only one side of the equation such that all the variables appearing on the same side are $\mathcal{J}$-above it, and all variables on the other side are strictly $\mathcal{J}$-above it. Thus, if this assignment does not satisfy our system, the system is unsatisfiable because any assignment $\mathcal{J}$-above it will also violate this inequality.

[2] Note that, in general, the solution set of an equation over G can be assumed to have this property only when G is Abelian.

Note once more that when the arity of the equations is bounded, this can be obtained by appealing to [8] which shows that CSP is tractable when the relations are preserved by an associative, commutative and idempotent operation. In our case, this operation is simply multiplication in M.

As a corollary, EQN^*_M is tractable if M is a direct product of an Abelian group and a commutative idempotent. On the other hand, when M is not a divisor of such a direct product, i.e. when M is not in $\mathbf{J}_1 \vee \mathbf{G}_{\mathrm{com}}$, then by Theorem 2, EQN^*_M is NP-complete.

5 $\mathrm{T}-\mathrm{EQN}^*$: Equations with only Constants on the Right-Hand Side

If we restrict our equations to have only constants on the right-hand side, we can do better. Note that we can assume that the left-hand side contains no constants by adding to our system equations $X_m = m$ for all $m \in M$ and replacing constants on the left side by the appropriate variable.

We denote as $\mathbf{R}_1 \vee \mathbf{L}_1$ the variety generated by idempotent $\mathcal{R}$-trivial and idempotent $\mathcal{L}$-trivial monoids. A monoid M is in $\mathbf{R}_1 \vee \mathbf{L}_1$ iff it satisfies $m^2 = m$ for all $m \in M$ and $abaca = abca$ for all $a, b, c \in M$ (See e.g. [4]).

Theorem 6. *If M belongs to $\mathbf{R}_1 \vee \mathbf{L}_1$, then $\mathrm{T}-\mathrm{EQN}^*_M$ can be solved in polynomial time.*

Proof. Our algorithm works by shrinking a list of possible values for each variable and implicitly uses the fact that the relations defined by equations over M are closed under a set function [5]. We use the following lemma:

Lemma 4. *Let $M \in \mathbf{R}_1 \vee \mathbf{L}_1$ and suppose $x_1 \dots x_k = m$ and $y_1 \dots y_l = m$. For all shuffles K of $x_1 \dots x_k$ with $y_1 \dots y_l$, we have $K = m$.*

To see this, recall [4] that in idempotent monoids, the product of two elements $\mathcal{J}$-above some $u \in M$ is also $\mathcal{J}$-above u. Hence we have $K \geq_{\mathcal{J}} m$ since K is a product of elements $\mathcal{J}$-above m. On the other hand, since all x_i's appear in K, we can use the relation $abaca = abca$ to get $KmK = Kx_1 \cdots x_k K = K^2 = K$. Thus, $m \geq_{\mathcal{J}} K$ and so $m \mathcal{J} K$. Note also that every prefix of $x_1 \dots x_k$ is $\mathcal{R}$-above m and so $x_1 \dots x_i m = m$ for all $i \leq k$.

We claim that $K \geq_{\mathcal{R}} m$. Indeed, we have $Km = Kmx_1 \dots x_k y_1 \dots y_l$. Using again the relation $abaca = abca$, we can replace the occurrence of x_i in K on the right-hand side of this equation with the prefix $x_1 \dots x_i$ since all the x_j with $j \leq i$ appear both before and after x_i. Hence Km can be written as a product of prefixes of $x_1 \cdots x_k$ or $y_1 \dots y_l$ times m. Thus $Km = m$ and $K \geq_{\mathcal{R}} m$.

By a symmetric argument, $K \geq_{\mathcal{L}} m$. Since $m \mathcal{J} K$, Lemma 1 implies that $m \mathcal{H} K$ and $m = K$ by aperiodicity and we obtain the lemma.

Now consider the following algorithm. For each variable X_i, $1 \leq i \leq n$, initialize a set $A_i = M$ of "possible values." Then repeat the following until either the A_i are fixed or some $A_i = \emptyset$: for all i from 1 to n, for each equation E

involving X_i, and each $a_i \in A_i$, if there exists no n-tuple $(a_1, \ldots, a_i, \ldots a_n)$ with $a_j \in A_j$ that satisfies E, then set $A_i \Rightarrow A_i - \{a_i\}$.

If some A_i is empty, the system clearly has no solution. Conversely, we are left with sets A_i such that for all $a_i \in A_i$ and all equations E in the system, there are $a_j \in A_j$ for all $i \neq j$ such that the n-tuple $(a_1, \ldots, a_n)$ satisfies E. We claim that this guarantees the existence of a solution to the system.

Indeed, let t_i be the product in M of all elements of $A_i = \{a_i^{(1)}, \ldots, a_i^{(s_i)}\}$ in some arbitrary order. Then $(t_1, \ldots, t_n)$ satisfies all equations in the system. To see this, consider some equation $E = X_1 X_2 \ldots X_k = m$. The product $t_1 t_2 \ldots t_k$ is a shuffle of solutions to this equation by definition of the A_i's, so by the above lemma $(t_1, \ldots, t_n)$ is also a solution.

It remains to show that the algorithm has a polynomial running time. It is sufficient to show that we can test in polynomial time for the existence of a tuple $\overline{a} = (a_1, \ldots, a_i, \ldots, a_n)$, $a_j \in A_j$ satisfying a given equation, say $X_1 \ldots X_k = m$.

For each $m \in M$, we will denote M_m the subset of M^* given by $M_m = \{m_1 m_2 \ldots m_k : eval_M(m_1, m_2 \ldots m_k) = m\}$. Because $\mathbf{R}_1 \vee \mathbf{L}_1$ is contained in the aperiodic variety $\mathbf{DA}$ (see [14]), each M_m can be expressed as the disjoint union of languages of the form[3] $B_0^* b_1 B_1^* \ldots b_r B_r^*$ with $b_j \in M$ and $B_j \subseteq M$. So to test for the existence of an $\overline{a}$ with the right properties, it is sufficient to consider, for each of the concatenations $B_0^* b_1 B_1^* \ldots b_r B_r^*$ that compose M_m, the $\binom{k}{r}$ possibilities of placing the b_i's among the k variables. To validate this choice, it is enough to check that $b_i \in A_j$ if we set $X_j = b_i$, and that for all other variables the corresponding A_i contains at least one element which belongs to the right B_j's.

Here again, we seek a dichotomy result for aperiodic monoids. By our proof of Theorem 1, T$-$EQN* is NP-complete for non-idempotent aperiodics. So we only need to consider idempotent aperiodics outside $\mathbf{R}_1 \vee \mathbf{L}_1$ that is idempotent monoids that fail to satisfy the relation $abaca = abca$.

Theorem 7. *If M is idempotent aperiodic outside $\mathbf{R}_1 \vee \mathbf{L}_1$ then* T$-$EQN* *is NP-complete.*

Proof. We start by stating a technical lemma about idempotent monoids outside $\mathbf{R}_1 \vee \mathbf{L}_1$ whose proof is due to Klima [9].

Lemma 5. *Suppose M is an idempotent aperiodic monoid outside of $\mathbf{R}_1 \vee \mathbf{L}_1$. There exists $a, b, c \in M$ such that $abaca \neq abca$ and moreover for all e, f lying strictly $\mathcal{J}$-above b (resp. lying strictly $\mathcal{J}$-above c) the product ef is also strictly $\mathcal{J}$-above b (resp. strictly $\mathcal{J}$-above c).*

Choose now $a, b, c \in M$ be as in Lemma 5. We now reduce from Monotone NAE-3SAT, whose NP-hardness is insured by Schaefer's theorem.

For each variable X_i, create variables s_i, t_i and equations:

[3] We do not need this in our proof, but it can also be assumed that these concatenations are unambiguous, i.e. that if $w \in B_0^* b_1 B_1^* \ldots b_r B_r^*$, its factorization is unique.

(1) $bs_i = b$ (2) $s_i b = b$
(3) $t_i c = c$ (4) $ct_i = c$
(5) $abas_i t_i aca = abaca$

For each clause e.g. $X_1 \vee X_2 \vee X_3$ we further add two equations:

(6) $s_1 s_2 s_3 = b$ (7) $t_1 t_2 t_3 = c$

If we have a satisfying NAE-3SAT assignment, we can satisfy this system by setting $s_i = b$ and $t_i = 1$ if X_i is TRUE and $s_i = 1$ and $t_i = c$ if X_i is FALSE.

Conversely, Equations (1-4) show that s_i and t_i are $\mathcal{R}$- and $\mathcal{L}$-above b and c respectively. By Lemma 1 and aperiodicity, $s_i \mathcal{J} b$ forces $s_i = b$ and $t_i \mathcal{J} c$ forces $t_i = c$. Thus either $s_i = b$ or $s_i >_{\mathcal{J}} b$, and similarly either $t_i = c$ or $t_i >_{\mathcal{J}} c$.

Moreover, if we have both $s_i = b$ and $t_i = c$ then Eq. (5) is violated since $(ab)(ab)(ca)(ca) = abca \neq abaca$. Thus if we interpret $s_i = b$ as X_i is TRUE and $t_i = c$ as X_i is FALSE, we are guaranteed that no variable has both truth values.

If neither $s_i = b$ nor $t_i = c$, then we view X_i as indeterminate, and unable to contribute to the satisfaction of a NAE-3SAT clause: that is, among the remaining two variables there must be one true and one false. This corresponds to the fact that if (say) $s_i >_{\mathcal{J}} b$, then it cannot help satisfy Eq. (6), and one of the remaining two variables must be b.

Indeed, by Lemma 5 neither b nor c is the product of elements $\mathcal{J}$-above them, so if Eqs. (6) and (7) are satisfied, at least one of the s's must be set to b and one of the t's to c. Since we are considering *monotone* formulas, if a satisfying assignment exists with some variables indeterminate, then one exists with these variables set to true. So the Monotone NAE-3SAT formula is also satisfiable.

6 Conclusion

Some questions regarding the case where M is neither a group nor aperiodic are left open. Very recently however, these authors and, independently, Klima [9] proved that EQN^*_M is tractable for any $M \in \mathbf{J}_1 \vee \mathbf{G}_{\mathrm{com}}$. This, together with Theorem 4, completes the dichotomy for EQN^*. Similar arguments show that $\mathrm{T}-\mathrm{EQN}^*_M$ is tractable whenever M belongs to the variety $\mathbf{R}_1 \vee \mathbf{L}_1 \vee \mathbf{G}_{\mathrm{com}}$. We conjecture the matching hardness result for $\mathrm{T}-\mathrm{EQN}^*$, i.e. that $\mathrm{T}-\mathrm{EQN}^*_M$ is NP-complete whenever $M \notin \mathbf{R}_1 \vee \mathbf{L}_1 \vee \mathbf{G}_{\mathrm{com}}$.

Unfortunately, time and space constraints prevented these results (and a proof of Lemma 5) from being included in our final submission to MFCS. They should however be available from one of the authors' web pages by the time these proceedings are published.

The connection to CSP can also be further explored. For instance, it would be interesting to relate, as in Lemma 3, tractable sets of constraints with solution sets to equations over the domain viewed as a monoid.

We wish to thank Ondrej Klima for sending us a proof of Lemma 5 and many helpful comments and suggestions.

References

1. D. A. M. Barrington and D. Thérien. Finite monoids and the fine structure of NC^1. *Journal of the ACM*, 35(4):941–952, Oct. 1988.
2. D. M. Barrington, P. McKenzie, C. Moore, P. Tesson, and D. Thérien. Equation satisfiability and program satisfiability for finite monoids. In *MFCS'00*, pages 172–181, 2000.
3. V. Dalmau. *Computational Complexity of Problems over Generalized Formula*. PhD thesis, Universita Politécnica de Catalunya, 2000.
4. S. Eilenberg. *Automata, Languages and Machines*, volume B. Academic Press, 1976.
5. T. Feder and M. Y. Vardi. The computational structure of monotone monadic SNP and constraint satisfaction: A study through datalog and group theory. *SIAM J. on Computing*, 28(1):57–104, 1999.
6. M. Goldmann and A. Russell. The complexity of solving equations over finite groups. In *Proc. 14th Conf. on Computational Complexity*, pages 80–86, 1999.
7. U. Hertrampf, C. Lautemann, T. Schwentick, H. Vollmer, and K. Wagner. On the power of polynomial time bit-reductions. In *Conf. on Structure in Complexity Theory*, 1993.
8. P. Jeavons, D. Cohen, and M. Gyssens. Closure properties of constraints. *J. ACM*, 44(4):527–548, 1997.
9. O. Klima. Private communication, 2001.
10. J. Pearson and P. Jeavons. A survey of tractable constraint satisfaction problems. Technical Report CSD-TR-97-15, University of London, 1997. Available at http://www.dcs.rhbnc.ac.uk/research/.
11. J.-E. Pin. *Varieties of formal languages*. North Oxford Academic Publishers Ltd, London, 1986.
12. J.-F. Raymond, P. Tesson, and D. Thérien. An algebraic approach to communication complexity. *Lecture Notes in Computer Science (ICALP'98)*, 1443:29–40, 1998.
13. T. J. Schaefer. The complexity of satisfiability problems. In *Proc. 10th ACM STOC*, pages 216–226, 1978.
14. M. P. Schützenberger. Sur le produit de concaténation non ambigu. *Semigroup Forum*, 13:47–75, 1976.