Dichotomies in the Complexity of Solving Systems of Equations over Finite Semigroups

Ondřej Klíma[1⋆], Pascal Tesson[2,3⋆⋆], and Denis Thérien[3⋆⋆⋆]

[1] Department of Mathematics, Masaryk University,
Brno, Czech Republic,
klima@math.muni.cz

[2] Wilhelm Schickard Institut für Informatik, University of Tübingen
Tübingen, Germany
tesson@informatik.uni-tuebingen.de

[3] School of Computer Science, McGill University
Montréal (Qc.), Canada
{ptesso,denis}@cs.mcgill.ca

Abstract. We consider the problem of testing whether a given system of equations over a fixed finite semigroup S has a solution. For the case where S is a monoid, we prove that the problem is computable in polynomial time when S is commutative and is the union of its subgroups but is NP-complete otherwise. When S is a monoid or a regular semigroup, we obtain similar dichotomies for the restricted version of the problem where no variable occurs on the right-hand side of each equation.
We stress connections between these problems and constraint satisfaction problems. In particular, for any finite domain D and any finite set of relations Γ over D, we construct a finite semigroup S_Γ such that $CSP(\Gamma)$ is polynomial-time equivalent to the equation satifiability problem over S.

1 Introduction

M. Goldmann and A. Russell studied in [10] the relationship between the algebraic properties of a finite group and the complexity of determining the solvability of an equation or a system of equations over that fixed group. In particular, they showed that determining whether a system of equations over G has a solution is NP-complete for any non-Abelian G and polynomial time computable for any Abelian G. Partial results concerning the complexity of the more general problem of solving equations over finite semigroups were obtained in [2, 22, 23] for the case of a single equation.

This paper, on the other hand, is concerned with the complexity of solving systems of equations over a fixed finite semigroup. Formally, an equation over

⋆ Research supported by the Ministry of Education of the Czech Republic under the project MSM 143100009.

⋆⋆ Research supported by the Alexander von Humboldt Foundation

⋆⋆⋆ Research supported by NSERC and FQRNT grants.

a finite semigroup S is given as: $s_1 s_2 \dots s_k = t_1 t_2 \dots t_n$ where each s_i or t_i is either a constant $c \in S$ or a variable x_j. We will further say that an equation is a *target-equation* if its right-hand side contains no variable.

The SYSTEM OF EQUATIONS SATISFIABILITY problem for the finite semigroup S (denoted EQN^*_S) is to determine whether a given system of equations over S has a solution. We will also consider the restriction of EQN^*_S where each equation is a target-equation and denote this problem $\mathrm{T\text{-}EQN}^*_S$.

Our motivation to study this question is twofold. On one hand, the deep links between finite semigroup theory and automata theory have lead to a number of algebraic characterizations of complexity classes ([1, 12] among many others) and this has given increased importance to the study of problems whose computational complexity is parametrized by the properties of an underlying algebraic structure [2, 10, 23, 24].

Our second motivation relates to constraint satisfaction problems (CSP's) which provide a unified framework for the study of many combinatorial problems. A conjecture of Feder and Vardi [9] states that every CSP is either in P or NP-complete and such dichotomies have been established in a number of special cases (e.g. [3, 10, 11, 21]). Both EQN^*_S and $\mathrm{T\text{-}EQN}^*_S$ can be viewed as special cases of CSP's and they present an interesting case study since most of the recent progress towards this conjecture has relied heavily on universal algebra methods (e.g. [3–6, 8]).

Using these techniques, B. Larose and L. Zádori have recently studied the complexity of solving equations over arbitrary finite algebras rather than just semigroups and obtained very broad results [18] which overlap some of our work about finite monoids. Furthermore, the complexity of the counting problem associated with EQN^*_S has been investigated by G. Nordh and P. Jonsson [15], also using a universal algebra point of view.

We are able to classify completely the complexity of EQN^*_S and $\mathrm{T\text{-}EQN}^*_S$ for very wide classes of semigroups. Specifically, we obtain the three following dichotomy results:

Main Theorems

- *If M is a finite monoid then EQN^*_M is computable in polynomial time if M is commutative and is the union of its subgroups and is NP-complete otherwise.*
- *If M is a finite monoid then $\mathrm{T\text{-}EQN}^*_M$ is computable in polynomial time if M divides the direct product of an Abelian group with a monoid satisfying the identities $x^2 = x$ and $xyxzx = xyzx$, and is NP-complete otherwise.*
- *If S is a finite regular semigroup then $\mathrm{T\text{-}EQN}^*_S$ is computable in polynomial time if S divides the direct product of an Abelian group with a semigroup satisfying the identities $x^2 = x$ and $xyxzx = xyzx$, and is NP-complete otherwise.*

We cannot as of yet provide a similar dichotomy for $\mathrm{T\text{-}EQN}^*_S$ when S is not a monoid or a regular semigroup or a dichotomy for EQN^*_S when S is not a monoid. We obtain partial results but also show that the questions cannot be resolved unless we settle the long standing conjecture about the complexity of CSP's. Indeed, we show that for every constraint satisfaction problem Γ, there exist semigroups S_Γ and T_Γ such that Γ is polynomial-time equivalent to $\mathrm{EQN}^*_{S_\Gamma}$ and to T-$\mathrm{EQN}^*_{T_\Gamma}$. We further show that if $P \neq NP$ (which we will assume throughout this paper) the class of semigroups for which T-EQN^*_S or EQN^*_S lies in P is not closed under subsemigroups or homomorphic images, a fact which is bound to hamper the progress of further investigations.

In Section 2, we give a short introduction to CSP's and present the fundamental results from semigroup theory which are necessary for the development of our discussion. In Section 3, we prove a number of sufficient algebraic conditions for the NP-hardness of EQN^*_S and T-EQN^*_S while Section 4 presents polynomial-time algorithms for solving systems of equations over "easy" classes of semigroups. In Section 5, our three dichotomy theorems are proved as corollaries to the results of the two preceding sections. Finally, we prove in Section 6 that the complete classification of the complexity of either T-EQN^*_T or EQN^*_S would result in a classification of the complexity of CSP's. Some of our results require rather technical semigroup theory and we have postponed some the more tedious proofs to an appendix to improve the paper's readability.

Some of the results appeared in the proceedings of MFCS'00 [19] and in two of the authors' Ph.D. theses [17, 23].

2 Background

2.1 Constraint Satisfaction Problems

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(Γ) 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 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. For example Graph k-colorability is the CSP where the domain consists of the k colors, Γ contains the single binary relation encoding inequality and constraints correspond to edges in the graph.

This class of combinatorial decision problems has received a lot of attention because of the wide variety of problems which it encompasses and because constraint satisfaction problems arise so naturally in artificial intelligence. Deep connections with database theory and finite model theory have also been uncovered [9].

For a finite semigroup S, the problem EQN^*_S can be seen as a CSP problem over the domain S with Γ being the class of relations which are the solution set of an equation over S. There is a technical caveat: since we place no bound on the arity of these equations, Γ is a priori *not* finite but we will argue in Section 4 that the arity can in fact be bounded to 3 without loss of generality. Such a construction is impossible for T-EQN^*_S and this problem can only be considered a CSP in a looser sense, unless we forcibly restrict the arity of equations. Still, this technicality has no bearing on our discussion.

For any D, Γ, CSP(Γ) lies in NP and is easily seen to be NP-complete in general so one seeks to identify tractable restrictions of the problem. One can choose, for instance, to impose certain conditions on the structure of constraints appearing in a given instance. A lot of research has also dealt with identifying necessary and sufficient conditions on Γ such that CSP(Γ) is tractable. This approach was pioneered by T. Schaefer [21] who studied the CSP problem on Boolean domains. In this case, the problem is usually known as GENERALIZED SATISFIABILITY and Schaefer proved that this problem was 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 C_2. The only other 2-element monoid is U_1 the semilattice with two elements $\{1, 0\}$ whose multiplication is given by $1x = x1 = x$ and $0x = x0 = 0$. Interestingly, we can relate the last two of Schaefer's tractable cases to systems of equations over U_1.

Lemma 1. *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, if and only if 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. □

Our hardness results in Section 3 will use reductions from 3SAT, 1-3SAT (where we require that every clause contains exactly one literal set TRUE) and the restriction of the latter where each clause contains at least one negated and one unnegated literal. The NP-completeness of the three problems is given by Schaefer's Theorem [21].

Recently, tools from universal algebra [5, 6], group theory and relational database theory [9] have been used to identify "islands of tractability", i.e. classes of relations for which CSP is tractable. As noted in our introduction, it is conjectured that for any domain D and any set of relations Γ the problem CSP(Γ) either lies in P or is NP-complete. Let us define a k-ary operation to be any function $f : D^k \to D$ and say that a relation $R \in D^t$ is *closed under* f if for any k t-tuples lying in R:

$$(d_1^1, d_2^1, \ldots, d_t^1), \ldots, (d_1^k, d_2^k, \ldots, d_t^k)$$

we also have the t-tuple

$$(f(d_1^1, \ldots, d_1^k), \ldots, f(d_t^1, \ldots, d_t^k))$$

in R. The algebraic properties of the operations that preserve every relation in Γ can be studied to determine the complexity of CSP(Γ) [5]. Using this approach, A. Bulatov obtained a dichotomy theorem similar to the one of Schaefer for domains of size three [3]. Furthermore, two known islands of tractability are defined with the help of semigroups.

Theorem 1 ([4]). *Let S be a finite semigroup and Γ_S be the set of relations which are closed under the multiplication in S then CSP(Γ_S) is tractable if S is a so-called "block group" and is NP-complete otherwise.*

Theorem 2 ([9]). *If G is a group and Γ is a set of relations such that for each $R \in \Gamma$ of arity t we have R a coset of G^t, then over the domain G the problem CSP(Γ) can be solved in polynomial time.*

Recall that a subset of G^t forms a coset if and only if it is closed under the ternary operation $x \cdot y^{-1} \cdot z$. Although our results are incomparable to these two theorems, the mechanics of some of our upper bounds, as we will point out, are quite similar.

For a digraph G, we define the DIGRAPH RETRACT PROBLEM (or DRP_G) as follows: given an input digraph H containing G as a subgraph, is there a surjective graph homomorphism from H to G which is the identity on G. Equivalently, DRP_G can be viewed as a CSP whose domain consists of G's vertices and a set of relations consisting of one binary relation (corresponding to edges in G) and for each element d in the domain, the unary relation consisting of the singleton $\{d\}$.

Theorem 3 ([9]). *For every set of relations Γ there exists a digraph G_Γ such that CSP(Γ) is polynomial-time equivalent to DRP_{G_Γ}.*

2.2 Semigroups

We give here a brief introduction to the theory of finite semigroups and present the necessary definitions and results. Although the paper is mostly self-contained, we refer the reader to e.g. [13, 20] for a more thorough overview. In particular, the latter reference stresses connections to automata theory and computer science.

Recall that a semigroup S is a set with a binary associative operation, which we will write multiplicatively (save one noted exception in the proof of Lemma 23). A monoid M is a semigroup with a distinguished identity element. We will be solely concerned with *finite* semigroups and monoids and in the rest of this paper S and M will always denote respectively a finite semigroup and a finite monoid.

We denote by S^1 the monoid obtained from S by adding an identity element if there is none in S. For any semigroup S, we introduce five equivalence relations known as *Green's relations* which describe whether two elements generate the same ideals in S. Formally:

- $x \mathcal{J} y$ iff $S^1xS^1 = S^1yS^1$;
- $x \mathcal{L} y$ iff $S^1x = S^1y$;
- $x \mathcal{R} y$ iff $xS^1 = yS^1$;
- $x \mathcal{H} y$ iff both $x \mathcal{R} y$ and $x \mathcal{L} y$;
- $x \mathcal{D} y$ iff $x \mathcal{R} \circ \mathcal{L} y$, that is there exists z such that $x \mathcal{R} z$ and $z \mathcal{L} y$.

It can be shown that $\mathcal{R}$ is a left-congruence (i.e. $x \mathcal{R} y$ implies $cx \mathcal{R} cy$ for all c) and that $\mathcal{L}$ is a right-congruence. Moreover $\mathcal{R}$ and $\mathcal{L}$ commute (i.e. $\mathcal{D} = \mathcal{R} \circ \mathcal{L} = \mathcal{L} \circ \mathcal{R}$) and so all five of these relations are indeed equivalence relations. Moreover, the relations $\mathcal{J}$ and $\mathcal{D}$ coincide for any *finite* S. Since we are only interested in the structure of finite semigroups, we will consequently always refer to the $\mathcal{J}$-relation.

For an element x of S, we denote by $\mathcal{J}_x$ (resp. $\mathcal{R}_x$, $\mathcal{L}_x$, $\mathcal{H}_x$) the $\mathcal{J}$-class (resp. $\mathcal{R}$-, $\mathcal{L}$-, $\mathcal{H}$-class) of x. We also define natural pre-orders $\leq_{\mathcal{J}}$, $\leq_{\mathcal{R}}$, $\leq_{\mathcal{L}}$ on S with e.g. $x \leq_{\mathcal{J}} y$ if and only if $S^1xS^1 \subseteq S^1yS^1$. We will say that "$x$ is (strictly) $\mathcal{J}$-above y" if $x \geq_{\mathcal{J}} y$ (resp. $x >_{\mathcal{J}} y$), and similarly for $\leq_{\mathcal{R}}$ and $\leq_{\mathcal{L}}$. Note that $x \leq_{\mathcal{J}} y$ if and only if there exists $u, v \in S^1$ such that $x = uyv$. Similarly, $x \leq_{\mathcal{R}} y$ if and only if there is u with $x = yu$ and $x \leq_{\mathcal{L}} y$ if and only if there is u with $x = uy$. One can easily prove:

Lemma 2. *For any $a, b \in S$ such that $a \mathcal{J} b$, if $a \leq_{\mathcal{R}} b$ (resp. $a \leq_{\mathcal{L}} b$) then in fact $a \mathcal{R} b$ (resp. $a \mathcal{L} b$).*

The following lemma is the fundamental result about Green's relations:

Lemma 3 (Green's Lemma). *Suppose a and b are two elements of the same $\mathcal{R}$-class, i.e. there exist u, v s.t. $au = b$ and $bv = a$. Denote by $\rho_u : S \to S$ the function defined by $\rho_u(s) = su$. Then ρ_u and ρ_v are inverse bijections from $\mathcal{L}_a$ to $\mathcal{L}_b$ and from $\mathcal{L}_b$ to $\mathcal{L}_a$ respectively and they preserve $\mathcal{H}$-classes.*

The basic properties of Green's relations lead to the so-called "egg-box" representation of (finite) semigroups. Each $\mathcal{J}$-class of the semigroup is represented as a table where rows correspond to $\mathcal{R}$-classes, columns to $\mathcal{L}$-classes and cells to $\mathcal{H}$-classes. From Green's Lemma, we also know that all the cells of a given $\mathcal{J}$-class contain the same number of elements. When writing out the egg-box representation, the $\mathcal{J}$-classes are often laid out with respect to the $\leq_{\mathcal{J}}$ preorder (see later examples).

We say that $e \in S$ is *idempotent* if $e^2 = e$. Idempotents play an important role in the structure of semigroups: in particular, the identity element 1_M is an idempotent of M. We say that S has a *zero* is there is an element $0 \in S$ such that $0s = s0 = 0$ for all $s \in S$. Note that 0 is also idempotent.

Lemma 4. *Let $e = e^2$ be an idempotent of S. Then $a \leq_{\mathcal{R}} e$ if and only if $ea = a$. Similarly $a \leq_{\mathcal{L}} e$ if and only if $ae = a$.*

Lemma 5. *Let $a, b \in S$ with $a \mathcal{J} b$. Then $ab \in \mathcal{R}_a \cap \mathcal{L}_b$ if and only if $\mathcal{L}_a \cap \mathcal{R}_b$ contains an idempotent $e = e^2$. Otherwise, $ab <_{\mathcal{J}} a$.*

The subsemigroup generated by an element s of S is finite of course, so there must exist t, p such that $s^{t+p} = s^t$ and the subsemigroup can be shown to have a unique idempotent. We will denote by ω the *exponent* of S, that is the smallest integer such that s^ω is idempotent for all $s \in S$. For any idempotent $e \in S$, the set eSe forms a monoid of S with identity e which we call the *local submonoid* of S associated with e.

Groups are a well-known special case of monoids. Recall that a monoid G is a *group* if every element $g \in G$ has an inverse g^{-1} such that $gg^{-1} = g^{-1}g = 1_G$. Every idempotent in S forms a trivial subgroup of S. Note also that by Lemma 5 an $\mathcal{H}$-class containing an idempotent is closed under multiplication and, more generally, one can show:

Lemma 6. *Let H be any $\mathcal{H}$-class of S, then H contains an idempotent if and only if H is a maximal subgroup of S.*

Consequently every $\mathcal{H}$-class contains at most one idempotent. Using Green's Lemma, one can further show that any two maximal subgroups of a common $\mathcal{J}$-class are isomorphic. If every maximal subgroup of S is trivial then S is said to be *aperiodic* or *group-free*. An important consequence of Lemma 6 is:

Lemma 7. *A semigroup S is aperiodic if and only if all its $\mathcal{H}$-classes contain a single element.*

A $\mathcal{J}$-class is said to be *regular* if it contains an idempotent. It can be shown in fact that a regular $\mathcal{J}$-class contains at least one idempotent in each of its $\mathcal{R}$ and $\mathcal{L}$ classes. A semigroup is *regular* if all its $\mathcal{J}$-classes are regular. We further say that S is a *union of groups* if each $\mathcal{H}$-class contains an idempotent and thus forms a maximal subgroup of S. This is equivalent to the requirement that $s^{\omega+1} = s$ for each $s \in S$.

A semigroup is *completely simple* if it consists of a single $\mathcal{J}$-class. Note that by Lemma 5, a $\mathcal{J}$-class of S forms a completely simple subsemigroup if and only if it all its $\mathcal{H}$-classes are subgroups.

Let us denote as $E(S)$ the subsemigroup generated by the idempotents of S: if $E(S)$ contains *only* idempotent elements then we say that S is *orthodox*. It can be shown that if S is a union of groups then S is orthodox if and only if all its $\mathcal{J}$-classes are completely simple orthodox subsemigroups [13].

We say that the semigroup T *divides* S if T is the morphic image of a subsemigroup of S. A class of finite semigroups is a *(pseudo)-variety*[4] if it is closed under finite direct product and division. For two varieties $\mathbf{V}, \mathbf{W}$, we denote by $\mathbf{V} \vee \mathbf{W}$

[4] In this paper, we will use the term *variety* as a shorthand for the more technically correct *pseudo-variety*.

the smallest variety containing both $\mathbf{V}$ and $\mathbf{W}$: it consists of the semigroups which divide a direct product $S \times T$ with $S \in \mathbf{V}$ and $T \in \mathbf{W}$.

Some varieties will bear particular importance for this work, mainly subvarieties of the variety of *bands*, i.e. semigroups in which every element is idempotent. In particular, we will consider the varieties of *regular bands* $\mathbf{RB}$ satisfying $xyxzx = xyzx$, *normal bands* $\mathbf{NB}$ satisfying $xyzx = xzyx$ and *semilattices* $\mathbf{SL}$ satisfying $xy = yx$. Clearly, $\mathbf{SL} \subseteq \mathbf{NB} \subseteq \mathbf{RB}$. In a semilattice, the $\leq_{\mathcal{J}}$ forms a partial order and multiplication in the semigroup corresponds to the semilattice *meet* ($\wedge$). Note that two elements might not have a *join* ($\vee$), i.e. a least upper bound but if they do then it is unique. Note also that every band is aperiodic and is a union of (trivial) groups.

We will further denote $\mathbf{Ab}$ the variety of Abelian groups, $\mathbf{UG}$ the variety of unions of groups and $\mathbf{DS}$ the variety of semigroups whose regular $\mathcal{J}$-classes form completely simple subsemigroups (note that $\mathbf{UG} \subseteq \mathbf{DS}$). Mostly, we will look at semigroups S in $\mathbf{UG}$ and it is worth mentioning that over such S, the $\mathcal{J}$-relation is a congruence and the quotient $S/\mathcal{J}$ is a semilattice.

Definition 1. *If $\mathbf{B}$ is a variety of bands and $\mathbf{H}$ is a variety of groups, we say that S is a* strong $\mathbf{B}$ band of $\mathbf{H}$-groups *if there exists a band $E \in \mathbf{B}$, a family of disjoint groups $\{G_e | e \in E\}$, all of which lie in $\mathbf{H}$, and for every $e, f \in E$ such that $e \geq_{\mathcal{J}} f$ (in the $\mathcal{J}$-order of E) a group homomorphism $\phi_{e,f} : G_e \to G_f$ such that:*

1. *S is the union of the G_e;*
2. *$\phi_{e,e} = id_{G_e}$ for all $e \in E$;*
3. *for any $e \geq_{\mathcal{J}} f \geq_{\mathcal{J}} d$ we have $\phi_{f,d} \circ \phi_{e,f} = \phi_{e,d}$;*
4. *for $x \in G_e$ and $y \in G_f$ the multiplication in S is given by the formula*
$$x \cdot y = \phi_{e,ef}(x) \cdot \phi_{f,ef}(y).$$

One can verify that the multiplication defined above is associative so that S is indeed a semigroup.

The proof of the next lemma is included in the appendix.

Lemma 8. *For a semigroup S, a variety of bands $\mathbf{B}$ and a variety of groups $\mathbf{H}$, the following are equivalent:*

1. *S is a strong $\mathbf{B}$-band of $\mathbf{H}$-groups;*
2. *S belongs to $\mathbf{B} \vee \mathbf{H}$.*
3. *S is an orthodox union of groups all of which lie in $\mathbf{H}$, such that $E(S)$ is a band in $\mathbf{B}$ and $\mathcal{H}$ is a congruence. In particular, the idempotents form a subsemigroup and $S/\mathcal{H} \equiv E(S)$.*

3 Hardness Results

One would intuitively expect that solving a system of equations over some semigroup S is no easier than solving a system of equations over a subsemigroup of

S or a morphic image of S. As we will see in Section 6, this intuition is unfortunately incorrect but the following definition allows us to salvage it partly. We will say that a subset T of S is *inducible* if there exists some expression E over S (i.e. a product of variables and constants) whose image is exactly T.

Lemma 9. *If T is an inducible subsemigroup of S, then* $\mathrm{EQN}^*_T \leq_P \mathrm{EQN}^*_S$ *and* $\text{T-EQN}^*_T \leq_P \text{T-EQN}^*_S$.

This simple fact was established in [10]. We will make extensive use of it and note that in particular, the following subsets of S are always inducible: every local monoid eSe, the set of idempotents of S and the semigroup $E(S)$ which they generate and the subsemigroup I of elements lying in or $\mathcal{J}$-below some regular $\mathcal{J}$-class J of S (and similarly for $\mathcal{R}$ and $\mathcal{L}$ classes). For the latter, we use the expression xey where x, y are variables and e is some idempotent in J. Note that I is the two-sided ideal generated by e. Often, we will simply write that we "force a variable x to be idempotent" to mean that each of its occurrences is replaced by x^ω.

When establishing our lower bounds, it is often convenient to think of a certain variable, say x, as being restricted to a set of particular values $T \subseteq S$. This can clearly be done without loss of generality as long as there exists a system of (target-)equations $\mathcal{E}$ with variables $x, y_1, \ldots, y_k$ such that s is in T if and only if $\mathcal{E}$ has a solution when x is set to s. We will say that such T are *(target)-definable* in S. Of course, we have:

Lemma 10. *If T is a definable (resp. target-definable) subsemigroup of S then* $\mathrm{EQN}^*_T \leq_P \mathrm{EQN}^*_S$ *(resp.* $\text{T-EQN}^*_T \leq_P \text{T-EQN}^*_S$*).*

In particular, if J is a regular $\mathcal{J}$-class containing the idempotent e, the target-equation $uxv = e$ defines the set $\{x | x \geq_{\mathcal{J}} e\}$. The proof of the next lemma serves as a good example to illustrate the usefulness of the above observations.

Lemma 11. *If S contains a non-Abelian subgroup, then* T-EQN^*_S *is NP-complete.*

Proof. As we mentioned in our introduction, it has been shown that T-EQN^*_G is NP-complete for any non-Abelian group G [10]. Let e be the idempotent of a non-Abelian subgroup H of S. The local semigroup eSe is inducible and its subgroup H can be defined (as a subset of eSe) by the target-equation $x^\omega = e$. We thus have

$$\text{T-EQN}^*_H \leq_P \text{T-EQN}^*_{eSe} \leq_P \mathrm{EQN}^*_S,$$

yielding our result. □

Recall that a band is said to be *normal* if it satisfies $xyzx = xzyx$. In the appendix, we will prove the following technical result:

Lemma 12. *A band S is normal if and only if all its local monoids are semilattices.*

Lemma 13. *If S is a band but is not a normal band, then* EQN^*_S *is NP-complete.*

Proof. Since every local monoid of S is inducible, it suffices, by Lemma 12, to prove the NP-completeness of EQN^*_M for a non-commutative, idempotent *monoid* M. Let a, b in M be such that $ab \neq ba$. We can choose a, b 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 now obtain a reduction from 3SAT. For each Boolean variable X_i in the formula, we create variables $x_i, \bar{x}_i, y_i, \bar{y}_i$ and equations

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

Also, for each 3SAT clause, e.g. $X_1 \vee \overline{X}_2 \vee X_3$, we add an equation

$$(7) \quad x_1\bar{x}_2x_3 = a.$$

Given a satisfying assignment to the formula, we can construct a solution to the above system by setting $x_i = a,\ \bar{x}_i = 1,\ y_i = b$, and $\bar{y}_i = 1$ whenever X_i is TRUE, and $x_i = 1,\ \bar{x}_i = a,\ y_i = 1,\ \bar{y}_i = b$ whenever X_i is FALSE.

Conversely, suppose the system of equations is satisfiable. Equation (1) shows that both x_i and $\bar{x}_i$ lie $\mathcal{J}$-above 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 so at least one of x_i, $\bar{x}_i$ must be $\mathcal{J}$-equivalent to a. Moreover, Eqs. (1) and (2) insure that x_i, $\bar{x}_i$ are both $\mathcal{L}$-above and $\mathcal{R}$-above a, so if $x_i \mathcal{J} a$ (say) we must also have $x \mathcal{H} a$ and thus $x = a$ by aperiodicity. So at least one of x_i, $\bar{x}_i$ must be a. Similarly at least one of y_i, $\bar{y}_i$ must be b, since any elements strictly $\mathcal{J}$-above b commute with a.

If $x_i = a$, then $\bar{y}_i$ commutes with a by Eq. (5). Thus $\bar{y}_i$ must be strictly $\mathcal{J}$-above b. If $y_i = b$, then $\bar{x}_i$ commutes with b by Eq. (6), so $\bar{x}_i$ is strictly $\mathcal{J}$-above a. We can thus obtain a consistent truth assignment to the literals by setting X_i to TRUE if and only if $x_i = a$ and $y_i = b$ and $\overline{X}_i$ to TRUE if and only $\bar{x}_i = a$ and $\bar{y}_i = b$.

Since every element strictly $\mathcal{J}$-above a is central but a is not, a cannot be a product of elements $\mathcal{J}$-above it. Therefore, if $x_1\bar{x}_2x_3 = a$ then at least one of $x_1, \bar{x}_2, x_3$ must be a and the corresponding 3SAT clause is satisfied. □

Lemma 14. *If S does not lie in* **DS** *then* T-EQN^*_S *is NP-complete.*

Proof. If S is not in **DS**, then it contains a regular $\mathcal{J}$-class K which is not a subsemigroup. Equivalently, some $\mathcal{H}$-class of K does not contain an idempotent. We will work over the inducible subsemigroup of elements that lie in or $\mathcal{J}$-below K. In this subsemigroup K is a maximal $\mathcal{J}$-class and since we can force each variable to be idempotent and define with a target equation the set of elements lying in or above K, we can insure that each variable is one of the idempotents of K. Let I be the set of idempotents of K: by introducing additional target-equations, we will define a small subset of I whose properties will allow us to build our reduction.

Let $G=(V,E)$ be the undirected graph such that $V=I$ and $E=\{(e_i,e_j)|e_ie_j\notin K \text{ or } e_je_i\notin K\}$. Note that by Lemma 5 the edge (e_i,e_j) is in E if and only if $e_ie_je_i$ is strictly $\mathcal{J}$-below K. Since we assume that K is not a subsemigroup, the graph contains at least one edge.

Suppose that there are distinct idempotents $e_i,e_j,e_k\in V$ such that $(e_i,e_j)\in E$ but neither (e_i,e_k) nor (e_j,e_k) are in E. By definition, we have $e_ie_k\in K$ and thus $e_ie_k\mathcal{L}e_k$. Similarly $e_ke_j\,\mathcal{R}\,e_k$. So by Lemma 5, we know that $e_ie_ke_j=(e_ie_k)(e_ke_j)$ also lies in K because the intersection of the $\mathcal{L}$-class of e_ie_k and the $\mathcal{R}$-class of e_ke_j contains the idempotent e_k. Symmetrically we also have $e_je_ke_i\in K$.

Now the pair of target-equations

$$e_ixe_j=e_ie_j \qquad\qquad e_jxe_i=e_je_i$$

is satisfied when x is e_i or e_j. However one of e_ie_j or e_je_i lies outside K and so $x=e_k$ is not a solution. In other words, this pair of equations defines a set of elements that contains e_i and e_j but not e_k.

Next, suppose that the graph G contains a triangle with vertices e_i,e_j,e_k. We distinguish two cases. Suppose first that $(e_ie_je_i)^\omega=(e_ie_ke_i)^\omega$. Then the target-equation

$$(e_ixe_i)^\omega=(e_ie_je_i)^\omega$$

is satisfied for $x=e_j$ and $x=e_k$ but not $x=e_i$. If on the other hand we have $(e_ie_je_i)^\omega\neq(e_ie_ke_i)^\omega$ then we can assume w.l.o.g. that in fact $(e_ie_je_i)^\omega$ does not lie $\mathcal{R}$-above $(e_ie_ke_i)^\omega$. (Indeed, if the two values are both $\mathcal{R}$-related and $\mathcal{L}$-related then they must be equal for each $\mathcal{H}$-class contains at most one idempotent.) Then the equation

$$(e_ixe_i)^\omega w=(e_ie_ke_i)^\omega$$

can be satisfied by setting $x=e_i$ and $w=(e_ie_ke_i)^\omega$ or $x=e_k$ and $w=e_i$. But one cannot choose $x=e_j$ since we assumed that $(e_ie_je_i)^\omega$ does not lie $\mathcal{R}$-above our target $(e_ie_ke_i)^\omega$. In all cases, we can introduce an equation defining a set of elements containing only two of the three idempotents e_i,e_j,e_k.

By iteratively adding such constraints, we can thus define smaller and smaller subsets of I and we can continue until the corresponding graph is such that for any three points e_i,e_j,e_k, two out of the three possible edges are present. It is easy to see that this means that the graph is a complete bipartite graph. Note also that the graph still contains at least an edge. In other words, we have defined a subset H of I with $H=\{e_1,\ldots,e_s,f_1,\ldots,f_t\}$ such that any product of e_i's or any product of f_i's lies in K but both $e_if_je_i$ and $f_je_if_j$ lie outside K, for any i,j.

We can now show the NP-completeness of T-EQN$^*_S$ with a very simple reduction from 3SAT. Of course, we begin by constraining the variables, as we just described, so that each of them lies in H and for every Boolean literal X_i, we then pose the equations

$$(1)\ ex_i\bar{x}_ie=efe \qquad\qquad (2)\quad fx_i\bar{x}_if=fef$$

(where, say, $e = e_1$ and $f = f_1$) and for a clause $X_1 \vee \overline{X}_2 \vee X_3$ the equation

$$(3) \quad (fx_1\bar{x}_2x_3f)^\omega = (fef)^\omega.$$

If the formula is satisfiable then one can verify that the system is satisfied by setting $x_i = e$ and $\bar{x}_i = f$ when X_i is TRUE and $x_i = f$ and $\bar{x}_i = e$ otherwise.

Conversely, Equations (1,2) force exactly one of $x_i, \bar{x}_i$ to be some e_j and the other to be some f_k. If we set each Boolean literal to TRUE iff the corresponding variable is some e_j then Equation (3) will insure that each clause contains at least one true literal since any product $ff_rf_sf_tf$ lies in K. □

Lemma 15. *If M is a monoid which is not a union of groups, then* T-EQN$^*_M$ *is NP-complete.*

Proof. There must exist some $m \in M$ which is not part of a subgroup and thus, by Lemma 5, such that m^2 lies strictly $\mathcal{J}$-below m. Let us choose a $\mathcal{J}$-maximal such m: Any element $t >_\mathcal{J} m$, is an element of a subgroup and thus satisfies $t^{\omega+1} = t$. In particular, for any two elements s, t lying strictly above m we cannot have $st = ts = m$ for then s and t commute and so

$$m^{\omega+1} = (st)^{\omega+1} = s^{\omega+1}t^{\omega+1} = st = m,$$

a contradiction. Furthermore, if u, v are $\mathcal{J}$-related to m with $uv = m$ then we have $u\,\mathcal{R}\,m$ and $v\,\mathcal{L}\,m$ and this implies that we cannot have $uv = vu = m$: In that case, we would have $u, v \in \mathcal{H}_m$ but the product of any two elements of $\mathcal{H}_m$ must lie strictly $\mathcal{J}$-below m.

We use these observations to obtain the following reduction from 1-3SAT: for each Boolean variable X_i in the formula, we introduce variables x_i, $\bar{x}_i$ and equations

$$(1)\ x_i\bar{x}_i = m \qquad\qquad (2)\ \bar{x}_ix_i = m.$$

Moreover, for each clause of the formula, e.g. $(X_1 \vee \overline{X}_2 \vee X_3)$ we add the equation

$$(3)\ x_1\bar{x}_2x_3 = m.$$

Suppose first that the 1-3SAT formula is satisfiable. Then one can check that the resulting system of equations is satisfied by setting $x_i = m$ and $\bar{x}_i = 1$ whenever X_i is TRUE, and $x_i = 1$ and $\bar{x}_i = m$ whenever X_i is FALSE.

Conversely, suppose that this system of equations is satisfiable. Our initial observations show that Equations (1) and (2) can only be satisfied if exactly one of $x_i, \bar{x}_i$ lies strictly $\mathcal{J}$-above m while the other is $\mathcal{J}$-related to m. We thus obtain a consistent truth assignment by setting X_i (resp. $\overline{X}_i$) to TRUE if and only if x_i (resp. $\bar{x}_i$) is $\mathcal{J}$-related to m. We claim that this truth assignment satisfies the 1-3SAT formula.

In any solution to the system, we have $x_i \geq_\mathcal{R} m$ by Eq. (1) and $x_im = x_i\bar{x}_ix_i = mx_i$ from Eqs. (1,2). We claim that if $x_i >_\mathcal{J} m$ then in fact $x_im\,\mathcal{H}\,m$. Indeed, $x_i\,\mathcal{H}\,x_i^\omega$ since it lies strictly $\mathcal{J}$-above m and by Lemma 4 $x_i^\omega m = m$. So

$x_i m \,\mathcal{L}\, m$. Furthermore, $x_i m = m x_i \leq_{\mathcal{R}} m$ and thus $x_i m \,\mathcal{R}\, m$ by Lemma 2. By Green's Lemma, we can infer that $x_i \mathcal{H}_m = \mathcal{H}_m$.

This allows to conclude that if an equation of type (3), say $x_1 \bar{x}_2 x_3 = m$ is satisfied, then it cannot be that $x_1, \bar{x}_2, x_3$ are all $\mathcal{J}$-above m for then, by our previous claim, $m^2 = x_1 \bar{x}_2 x_3 m \in \mathcal{H}_m$, a contradiction. Similarly, if any two of these variables lie in $\mathcal{H}_m$ then the product must lie strictly $\mathcal{J}$-below m and so exactly one of them lies in $\mathcal{H}_m$ while the other two lie strictly $\mathcal{J}$-above m. Thus, exactly one literal of the corresponding clause is true. $\square$

The last three hardness results of this section require rather technical arguments and their complete proofs have been postponed to the appendix. Nevertheless, we illustrate each of them using a concrete semigroup and a reduction very similar to the more general construction described later on.

Lemma 16. *If S is a union of groups such that $\mathcal{H}$ is not a congruence on S, then* T-EQN$^*_S$ *is NP-complete.*

Example 1 Let S be the four-element semigroup $\{a, a^2, e, ae\}$ with multiplication specified by $a^2 e = e, \quad ex = e$ for all x and $a^3 = a$. This semigroup (in fact a monoid since a^2 is an identity) consists of two $\mathcal{J}$-classes: the top one contains the two element group $\{a, a^2\}$ and the bottom one two idempotents e, ae which are $\mathcal{L}$-related, as pictured in Figure 1. Thus S is indeed a union of groups but $\mathcal{H}$ is not a congruence since we have $a \,\mathcal{H}\, a^2$ but $ae \,\not\mathcal{H}\, a^2 e = e$.
We claim that T-EQN$^*_S$ is NP-complete and construct the reduction from 3SAT. For each Boolean variable X_i, we introduce variables $x_i, \bar{x}_i, v_i, s_i, t_i$ such that v_i, s_i, t_i are all $\mathcal{H}$-related to a and add the equations:

$$\begin{array}{llll} (1) & x_i e = e & (2) & \bar{x}_i e = e \\ (3) & v_i x_i s_i e = e & (4) & v_i a \bar{x}_i t_i e = e \end{array}$$

Moreover, for each 3SAT clause, e.g. $X_1 \vee \overline{X}_2 \vee X_3$ we introduce the equation

$$(5) \quad x_1 \bar{x}_2 x_3 = e.$$

Given an assignment to the Boolean literals satisfying the 3SAT formula, one can verify that this system has a solution by setting $x_i = e$, $\bar{x}_i = a^2$, $t_i = a$ and $v_i = s_i = a^2$ whenever X_i is TRUE and $x_i = a^2$, $\bar{x}_i = e$, $t_i = a^2$ and $v_i = s_i = a$ whenever X_i is FALSE. Furthermore, each equation of type (5) is indeed satisfied since at least one of the three terms is e (by the satisfiability of the formula) while the others are a^2.
Conversely, suppose that there exists a solution to the constructed system. Equations (1,2) show that x_i and $\bar{x}_i$ take values in $\{a^2, e\}$ but we cannot have $x_i = \bar{x}_i = e$ for otherwise no value of v_i can simultaneously satisfy Eqns. (3,4) (while the values of s_i and t_i are irrelevant in that case). Correspondingly, if we set X_i (resp. $\overline{X}_i$) to TRUE when $x_i = e$ (resp. $\bar{x}_i = e$) then a litteral and its complement are never both true. Finally, if $x_1 \bar{x}_2 x_3 = e$, one of those three variables must be e and so each clause in the formula is indeed satisfied. $\square$

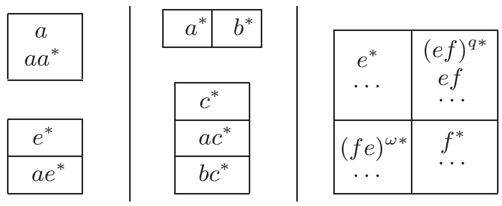

Fig. 1. Egg-box pictures for the semigroups of Examples 1, 2 and 3 respectively. Idempotents are marked with $*$.

Lemma 17. *If S is a band but is not a regular band, then* T-EQN$^*_S$ *is NP-complete.*

Example 2 Consider the band $S = \{a, b, c, ac, bc\}$ pictured in Figure 1 and where $ab = b$, $ba = a$ and $c = ca = cb$. It is not a regular band, because $abca = bc$ but $abaca = ac$.
We can now obtain the following reduction from 3SAT to T-EQN$^*_S$. For each Boolean literal X_i in the formula, we introduce the variables $x_i, \bar{x}_i, y_i$ and construct the equations

$$\begin{array}{llll} (1) & ax_i a\bar{x}_i = ac & (2) & bx_i b\bar{x}_i = bc \\ (3) & a\bar{x}_i ax_i = ac & (4) & b\bar{x}_i bx_i = bc \\ (5) & y_i x_i ac = ac & (6) & y_i \bar{x}_i bc = bc \end{array}$$

Consider some solution of this system. Suppose that both x_i and $\bar{x}_i$ are $\mathcal{J}$-related to a then $ax_i a\bar{x}_i \,\mathcal{J}\, a$, it is a contradiction. From Equation (1) we have $x_i \neq bc$ and from Equation (2) we have $x_i \neq ac$. This means that if x_i is $\mathcal{J}$-related to c then $x_i = c$. The same holds for $\bar{x}_i$ by Equations (3,4). Suppose that both x_i and $\bar{x}_i$ are $\mathcal{J}$-related to c, i.e. both are equal to c. Then $y_i x_i ac = y_i c$ and $y_i \bar{x}_i bc = y_i c$ which is contradiction with Equations (5,6) and the fact $ac \neq bc$. Altogether one of x_i and $\bar{x}_i$ is $\mathcal{J}$-related to a and the second one is equal to c.
We complete our reduction by introducing, for each of clause of the 3SAT formula, e.g. $X_1 \vee \overline{X}_2 \vee X_3$, the equation:

$$(7) \quad ax_1 a\bar{x}_2 ax_3 = ac.$$

One can now verify that if the 3SAT instance is satisfiable, then we can satisfy the system obtained through our reduction by letting $x_i = c$, $\bar{x}_i = a$, $y_i = a$ whenever X_i is TRUE, and $x_i = a$, $\bar{x}_i = c$, $y_i = b$ whenever X_i is FALSE.
Conversely, suppose the system of the equations is satisfiable. Since exactly one of $x_i, \bar{x}_i$ is equal to c, we get a consistent truth assignment to the literals by setting X_i (resp. $\overline{X_i}$) to TRUE if and only if $x_i = c$ (resp. $\bar{x}_i = c$). This assignment satisfies every clause of the original formula for if the variables occurring in Eq. (7) all are $\mathcal{J}$-related to a we have $ax_1 a\bar{x}_2 ax_3 = a$. $\square$

Lemma 18. *If S contains a $\mathcal{J}$-class K forming a completely simple but unorthodox semigroup then* T-EQN$^*_S$ *is NP-complete.*

Example 3 Consider a completely simple semigroup S with two $\mathcal{R}$-classes and two $\mathcal{L}$-classes as represented in the eggbox-picture of Figure 1. So S contains four idempotents which we can denote as $e, f, (ef)^\omega, (fe)^\omega$ because of Lemma 5. We assume however that $ef \neq (ef)^\omega$ and let q be the smallest integer such that $(ef)^q = (ef)^\omega$. Again using Lemma 5, we know that if $st = u$ in S then $s \,\mathcal{R}\, u$ and $t \,\mathcal{L}\, u$.
We show the NP-completeness of T-EQN$^*_S$ using a reduction from 1-3SAT: for each Boolean variable X_i we create two variables $x_i, \bar{x}_i$ and force them to be idempotents. We include equations

$$(1)\ x_i\bar{x}_i = (ef)^q \qquad\qquad (2)\quad x_i fe\bar{x}_i = ef$$

For each clause, e.g. $X_1 \vee \overline{X}_2 \vee X_3$ we include the equation

$$(3)\quad (x_1f)(e\bar{x}_2)(x_3f) = ef.$$

If the formula is satisfiable, then one can easily verify that a solution to the system is obtained by setting $x_i = e$ and $\bar{x}_i = (ef)^q$ when X_i is TRUE and $x_i = (ef)^q$ and $\bar{x}_i = f$ when X_i is FALSE.
Conversely, since both x_i and $\bar{x}_i$ are idempotents, Eqn. (1) shows that $x_i \in \{e, (ef)^q\}$ while $\bar{x}_i \in \{f, (ef)^q\}$. We cannot have simultaneously $x_i = e$ and $\bar{x}_i = f$ (by Eqn. 1) and we cannot have $x_i = \bar{x}_i = (ef)^q$ for then $x_i fe\bar{x}_i = (ef)^q fe(ef)^q = (ef)^q$, in violation of Eqn. (2).
We can therefore choose X_i to be TRUE when $x_i = e$ and $\overline{X}_i$ to be TRUE when $\bar{x}_i = f$. Now note that the product x_if is ef when $x_i = e$ and $(ef)^q$ when $x_i = (ef)^q$. Similarly $e\bar{x}_i$ is ef when $\bar{x}_i = f$ and $(ef)^q$ when $\bar{x}_i = (ef)^q$. Let us first assume that $q \geq 3$: if an equation of type (3) is satisfied: $(x_1f)(e\bar{x}_2)(x_3f) = ef$ then exactly one of the three terms in the product is ef and so the corresponding clause contains exactly one literal set to TRUE.
If $q = 2$ then an equation of type (3) will also be satisfied if all three corresponding literals are TRUE. But we can assume w.l.o.g. using Schaefer's Theorem that each clause contains at least one unnegated literal (say X_1) and one negated literal (say $\overline{X}_2$). We then add a fourth type of equation: $x_1\bar{x}_2 = (ef)^q$ which can be satisfied only by setting at least one of $x_1, \bar{x}_2$ to $(ef)^q$. Correspondingly, at least one of $X_1, \overline{X}_2$ is FALSE and this prevents the problematic case of all three literals in a clause being TRUE. □

4 Upper Bounds

We chose to establish our upper bounds for EQN* and T-EQN* by presenting explicit algorithms rather than by combining established general results in the study of CSP's as this clearly shows the relation between the algebraic structure of a semigroup S and the tractability of solving equations over S. An alternative presentation was chosen in [23] and we will outline its main ideas.

If we introduce for each $s \in S$ a dummy variable x_s and an equation $x_s = s$, we can assume that no other equation in a system over S uses constants. Furthermore, the equation $x_1 \ldots x_{n+m} = y$ is equivalent to the pair $x_1 \ldots x_n = z$ and $zx_{n+1} \ldots x_{n+m} = y$ and by using this trick repeatedly, we can assume that every equation in a system is either $x_s = s$ for some constant s or $x_i x_j = x_k$ (see discussion in [18]). However the dummy variable z that we introduced appears on the right-hand side of an equation so this construction does not work for systems of target-equations.

Lemma 19. *If S is a semilattice, then* EQN^*_S *is computable in polynomial time.*

Proof. Observe that if $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_n)$ are solutions to a system of equations $\mathcal{E}$ in n variables over S, then $(u_1 v_1, \ldots, u_n v_n)$ is also a solution to $\mathcal{E}$. Indeed if $x_{i_1} x_{i_2} = x_{i_3}$ is an equation of $\mathcal{E}$ then we have

$$u_{i_1} v_{i_1} u_{i_2} v_{i_2} = u_{i_1} u_{i_2} v_{i_1} v_{i_2} = u_{i_3} v_{i_3}$$

because S is commutative. Equations of the form $x_i = s$ for $s \in S$ are also satisfied because of idempotency. Note that $(u_1 v_1, \ldots, u_n v_n)$ is the meet of $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_n)$ in the semilattice S^n. It is in fact known that this closure property of relations induced by equations over S suffices to obtain a polynomial time algorithm for EQN^*_S [14]. Still, we sketch an explicit algorithm, since it will serve as the basis for an algorithm solving equations over a larger class of semigroups.

Our algorithm maintains a lower bound $\boldsymbol{y} = (y_1, \ldots, y_n)$ for the minimal solution to $\mathcal{E}$. We initialize $\boldsymbol{y}$ as $(0, \ldots, 0)$ and update it as follows. For each equation of the form $x_s = s$ for $s \in S$ we begin by setting the corresponding y_s to s. In each subsequent step, if $(y_1, \ldots, y_n)$ is a solution to $\mathcal{E}$, the algorithm halts. If some equation in $\mathcal{E}$, say $x_{i_1} x_{i_2} = x_{i_3}$ is not satisfied then since we are maintaining $\boldsymbol{y}$ as a lower bound to any assignment satisfying $\mathcal{E}$, we know that in any such assignment y_{i_1} and y_{i_2} will be bounded below by y_{i_3}. Thus, if y_{i_1} is not $\mathcal{J}$-above y_{i_3}, we can update our lower bound by setting $y_{i_1} := y_{i_1} \vee y_{i_3}$ i.e. the $\mathcal{J}$-minimal element of S lying above both of them[5]. We do similar updates for y_{i_2} and y_{i_3}.

We iterate this until we reach a fixed point for $\boldsymbol{y}$. The process terminates in at most $n \cdot |S|$ steps since the value of $\boldsymbol{y}$ always increases in the semilattice S^n and if the fixed point is not a solution to the system, then it must be that $\mathcal{E}$ contains the equation $x_s = s$ but the corresponding y_s lies above s and so $\mathcal{E}$ is unsatisfiable. $\square$

Recall that a band is a regular band if it satisfies the identity $abaca = abca$.

Lemma 20. *If S is a regular band, then* T-EQN^*_S *is computable in polynomial time.*

[5] If no such element exists, we conclude that the system is unsatisfiable.

In order to establish this upper bound, we will use the following property of solutions to a target-equation over a regular band. Recall that a shuffle of two strings $x_1 \dots x_k$ and $y_1 \dots y_l$ is a string formed by these $k+l$ elements and in which the x_i's and y_i's appear in their original order.

Lemma 21. *Let S be a regular band and suppose $x_1 \dots x_k = s$ and $y_1 \dots y_l = s$ for some $x_i, y_i, s \in S$. For all shuffles K of $x_1 \dots x_k$ with $y_1 \dots y_l$, we have $K = s$.*

Proof. In any band, the product of two elements $\mathcal{J}$-above some $u \in S$ is also $\mathcal{J}$-above u [13]. Hence we have $K \geq_{\mathcal{J}} s$ since each x_i, y_i lies $\mathcal{J}$-above s. On the other hand, since all x_i's appear in K, we can use the identity $abaca = abca$ to get $KsK = Kx_1 \cdots x_k K = K^2 = K$. Thus, $s \geq_{\mathcal{J}} K$ and so $s \mathcal{J} K$. Furthermore, for any $i \leq k$ we have $x_1 \dots x_i s = x_1 \dots x_i x_1 \dots x_i \dots x_k = s$.

We claim that $K \geq_{\mathcal{R}} s$. Indeed, we have $Ks = Ksx_1 \dots x_k y_1 \dots y_l$. Using again the identity $abaca = abca$, 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 Ks can be written as a product of prefixes of $x_1 \cdots x_k$ or $y_1 \dots y_l$ times s. Thus $Ks = s$ and $K \geq_{\mathcal{R}} s$.

By a symmetric argument, $K \geq_{\mathcal{L}} s$. Since $s \mathcal{J} K$, we have $s \mathcal{H} K$ and $s = K$ by aperiodicity. $\square$

From the universal algebra perspective, Lemma 21 can be used to show that the relations induced by target-equations over a regular band are closed under a so-called *set function* [8] and this is known to be a sufficient condition for the tractability of the corresponding CSP. The algorithm we describe next is implicitly using this fact.

Proof. of Lemma 20 For each variable x_i, $1 \leq i \leq n$, we initialize a set $A_i = S$ of "possible values" for x_i and 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 $(a_1, \dots, a_i, \dots, a_n)$ with $a_j \in A_j$ that satisfies E, then we set $A_i := A_i - \{a_i\}$.

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 $(a_1, \dots, a_n)$ satisfies E. We claim that this guarantees the existence of a solution to the system.

Indeed, let t_i be the product in S of all elements of $A_i = \{a_i^{(1)}, \dots, a_i^{(r_i)}\}$ in some arbitrary order. Then $(t_1, \dots, t_n)$ satisfies all equations in the system. To see this, consider some equation $E : x_{j_1} x_{j_2} \dots x_{j_k} = s$. One can easily show using idempotency that by definition of the A_i's and t_i's, the product $t_{j_1} t_{j_2} \dots t_{j_k}$ is a shuffle of solutions to this equation. So by Lemma 21, the tuple $(t_1, \dots, t_n)$ also satisfies the equation.

To show that our algorithm runs in polynomial time, it suffices to show that we can efficiently test whether a given equation $x_{j_1} \dots x_{j_k} = s$ has a solution $\boldsymbol{a} = (a_1, \dots, a_i, \dots, a_n)$ where a_i is given and for each $j \neq i$ we want $a_j \in A_j$. Since a regular band S satisfies the identities $xyxzx = xyzx$ and $x^2 = x$, the value of a product $s_1 s_2 \dots s_m$ with $s_i \in S$ is completely determined by the set of elements of S occurring among the s_i and the order in which they first appear from left to right and from right to left (see e.g. [16]). If we require that $x_{j_1} \dots x_{j_k} = s$ we can thus guess which elements of S do appear in this product and where they first and last appear: because $|S|$ is a constant, there are only polynomially many such possibilities. For each of these, it is easy to check whether or not we can find $\boldsymbol{a}$ such that $x_{j_1} \dots x_{j_k} = s$. $\square$

Note that for any solution $(s_1, \dots, s_n)$ to the system, we have $t_i \leq_{\mathcal{J}} s_i$ for each i because we must have $s_i \in A_i$.

For a set $Q = \{q_1, \dots, q_t\}$ of integers with $q_i \geq 2$ for all i, we define a decision problem LEQN_Q as follows: given a system $\mathcal{E}$ of linear equations modulo q_i for some $q_i \in Q$ (with different equations using possibly different moduli), determine if $\mathcal{E}$ has a solution over the integers.

Lemma 22. *The problem* LEQN_Q *lies in P for any set of moduli* Q.

This is not hard to prove using elementary arithmetic. Alternatively, one can see that LEQN_Q is a constraint satisfaction problem over the domain $D = \{0, 1, \dots, lcm(Q)-1\}$ whose relations are solution sets of linear equations (whose arity can be bounded to 3) over the different moduli. The domain D can be viewed as a cyclic group under addition and it is clear that every k-ary relation is then a coset of D^3. The problem is thus tractable by Theorem 2.

For the following two lemmas, it is useful to recall that in an orthodox union of groups S the idempotents form a subsemigroup $E(S)$. If S is also a strong band of groups then $E(S) \equiv S/\mathcal{H}$ is the image of the homomorphism $\phi : s \mapsto s^\omega$. If $\mathcal{E}$ is a system of equations over S and $\phi : S \to T$ is a homomorphism, we can naturally construct a system $\phi(\mathcal{E})$ over T by replacing every constant c appearing in $\mathcal{E}$ by $\phi(c)$. If $\mathcal{E}$ has a solution then of course so does $\phi(\mathcal{E})$.

Lemma 23. *If S is in* $\mathbf{SL} \vee \mathbf{Ab}$ *then* EQN^*_S *is computable in polynomial-time.*

Proof. Let $\mathcal{E}$ be a system of equations over S in n variables. We know that S is a strong semilattice of Abelian groups and if $\mathcal{E}$ is satisfiable, then the corresponding system over $S/\mathcal{H} \equiv E(S)$ is also satisfiable. It is useful to note that the system over $E(S)$ can be obtained from $\mathcal{E}$ by raising every variable and constant to its ω power. Using the algorithm of Lemma 19, we can find the $\mathcal{J}$-minimal solution $(e_1, \dots, e_n)$ of the system over $E(S)$. If $(u_1, \dots, u_n)$ is an arbitrary solution of $\mathcal{E}$ then we have $u_i \geq_{\mathcal{J}} e_i$ and thus $e_i u_i \mathcal{J} e_i$ (and in fact $e_i u_i \,\mathcal{H}\, e_i$ since $\mathcal{J} = \mathcal{H}$ for $S \in \mathbf{SL} \vee \mathbf{Ab}$). Furthermore, $(e_1 u_1, \dots, e_n u_n)$ is also a solution of the system $\mathcal{E}$: say $x_1 x_2 = x_3$ is some equation of $\mathcal{E}$, then because S is a commutative we have

$$(e_1 u_1)(e_2 u_2) = (e_1 e_2)(u_1 u_2) = e_3 u_3.$$

Also, $s^\omega s = s$ for every $s \in S$ since it is a union of groups and so equations $x_s = s$ are also satisfied.

So if $\mathcal{E}$ has a solution, it has a solution $(u_1, \dots, u_n)$ such that $u_i \,\mathcal{H}\, e_i$ or, in other words such that u_i belongs to the subgroup G_i whose identity element is

e_i. Note that if we know the group in which each variable lies, we consequently know the group in which a product of them lies and we can associate to each equation E_j of $\mathcal{E}$ a subgroup H_j in which both its right-hand and left-hand sides will sit.

Suppose for simplicity that each group is cyclic (the more general case can easily be handled by decomposing the groups into their cyclic factors). We write the group operations additively and for every variable x_i of $\mathcal{E}$ introduce an integer variable y_i such that $x_i = y_i g_i$ where g_i is the generator of G_i. Each equation E_j of $\mathcal{E}$ can be viewed as an equation over H_j: using the homomorphism mapping the relevant G_i to H_j, we can thus rewrite each E_j as a linear equation modulo some integer q_j over variables y_i's. Clearly, $\mathcal{E}$ has a solution iff the resulting instance of LEQN_Q has a solution and we can check this in polynomial time. $\square$

This upper bound technique combines ideas from two classes of polynomial time algorithms for CSP's. This has lead to the identification of an apparently new "island of tractability" [7] which supersedes both Theorems 1 and 2: if S is a block group (i.e. a semigroup in which the idempotents generate a $\mathcal{J}$-trivial subsemigroup) and Γ is a set of relations over S closed by the operation $t(x, y, z) = xy^{\omega-1}z$ then $\mathrm{CSP}(\Gamma)$ is tractable. The latter is a so-called *Taylor operation* and an easy exercise shows that it indeed closes the relations defined by equations over a semigroup in $\mathbf{SL} \vee \mathbf{Ab}$.

Lemma 24. *If S is in $\mathbf{RB} \vee \mathbf{Ab}$ then T-EQN^*_S is computable in polynomial-time.*

Proof. We proceed exactly as in the previous proof: if $\mathcal{E}$ is our system of target-equations over S, we begin by considering the corresponding system over $S/\mathcal{H} \equiv E(S)$ and running the algorithm of Lemma 20: if it has no solution then $\mathcal{E}$ is also unsolvable. If $(e_1, \dots, e_n)$ is a solution over $E(S)$ then let $(u_1, \dots, u_n)$ be a solution to $\mathcal{E}$: by our remark following Lemma 20 we have $e_i \leq_{\mathcal{J}} u_i$ and thus $e_i u_i e_i \,\mathcal{H}\, e_i$.

We can adapt Lemma 21 to show that if S is a strong regular band of Abelian groups and $x_1 \dots x_k = s$ and $y_1^\omega \dots y_l^\omega = s^\omega$ then for any shuffle K of $x_1 \dots x_k$ and $y_1^\omega \dots y_l^\omega$ we have $K = s^{\omega+1} = s$. Therefore $(e_1 u_1 e_1, \dots, e_n u_n e_n)$ is a solution to $\mathcal{E}$.

So if $\mathcal{E}$ has any solution then it has a solution $(u_1, \dots, u_n)$ such that $u_i^\omega = e_i$. Once again, this means that we have identified in polynomial time the subgroup to which each u_i will belong and the rest of the proof is identical to that of Lemma 23. $\square$

5 Three Dichotomy Theorems

In this section we combine the results obtained so far to characterize the complexity of EQN^*_M and T-EQN^*_M for every finite monoid M and the complexity of T-EQN^*_S for every regular semigroup S.

Theorem 4. *If M is a finite monoid then* T-EQN$^*_M$ *is computable in polynomial time if M lies in* $\mathbf{RB} \vee \mathbf{Ab}$ *and is NP-complete otherwise.*

Proof. The upper bound is provided by Lemma 24. For the lower bound, if M is not a union of Abelian groups then EQN$^*_M$ is NP-complete by Lemmas 11 and 15. If M is a union of groups but is not orthodox then it must have a completely simple unorthodox subsemigroup [13] and NP-completeness follows from Lemma 18. If M is an orthodox union of Abelian groups then the inducible subsemigroup $E(M)$ is a band and Lemma 17 insures the NP-hardness of T-EQN$^*_M$ if this band is not regular. Finally, if M is an orthodox union of Abelian groups over which $\mathcal{H}$ is not a congruence, we can use Lemma 16. By Lemma 8, our proof is complete. □

Theorem 5. *If S is a finite regular semigroup then* T-EQN$^*_S$ *is computable in polynomial time if S lies in* $\mathbf{RB} \vee \mathbf{Ab}$ *and is NP-complete otherwise.*

Proof. Once again, Lemma 24 yields the upper bound. For the lower bound, note that if a regular semigroup S is not a union of groups then it must lie outside **DS**. In that case, the NP-completeness of T-EQN$^*_S$ follows from Lemma 14. If S is a union of groups, we can argue as in Theorem 4. □

Theorem 6. *If M is a finite monoid then* EQN$^*_M$ *is computable in polynomial time if M lies in* $\mathbf{SL} \vee \mathbf{Ab}$ *and is NP-complete otherwise.*

Proof. The upper bound is simply Lemma 23. For the lower bound: if M lies outside $\mathbf{RB} \vee \mathbf{Ab}$ then EQN$^*_M$ is NP-complete by Theorem 4. Otherwise, M is a strong band of groups and if the underlying band $E(M)$ is not a semilattice, NP-completeness follows from Lemma 13. □

B. Larose and L. Zádori recently showed that the NP-completeness half of this result can be obtained alternatively using universal algebra [18]. They show that if the set of relations defined by equations over M is closed under a so-called Taylor operation then M is a commutative union of groups (i.e. lies in $\mathbf{SL} \vee \mathbf{Ab}$) and this yields the lower bound. The converse of this statement is also true, a fact we implicitly exploited to obtain the matching upper bound.

6 Obstacles for more General Dichotomies

Ideally, we would want to prove that such dichotomies hold for EQN* and T-EQN* for all finite semigroups but our results in this section indicate that this is as difficult as obtaining a dichotomy for all CSP's.

Theorem 7. *For every set of relations Γ, there exists a semigroup S_Γ satisfying the identity $xyz = uvw$ such that CSP(Γ) is polynomial-time equivalent to* T-EQN$^*_{S_\Gamma}$.

Proof. By Theorem 3, we can assume that Γ contains a single binary relation R and all constants, i.e. all unary relations consisting of a singleton. In other words, every constraint of the CSP(Γ) instance either sets a variable to some constant value or constrains a pair of variables to lie in R.

We construct the semigroup S_Γ from generators $d_1, \ldots, d_k$ corresponding to the k elements of Γ's domain. Furthermore, we add a semigroup element $\langle d_i d_j \rangle$ for every pair of domain elements such that $(d_i, d_j) \notin R$. The last two elements of S_Γ are r and 0 and the multiplication is given by
- $d_i d_j = \langle d_i d_j \rangle$ if $(d_i, d_j) \notin R$ and $d_i d_j = r$ otherwise;
- $xy = 0$ unless x and y are among the k generators.
In particular, the product of any three elements of the semigroup is 0 so S_Γ satisfies the identity $xyz = uvw$.

The reduction from CSP(Γ) to T-EQN$^*_{S_\Gamma}$ is now quite transparent: for every variable y_i of the CSP instance, we create a variable x_i. If the CSP variable y_i is bound to the domain value d_s then we correspondingly impose $x_i = d_s$ in the system and for each constraint $(y_i, y_j) \in R$ we introduce the equation $x_i x_j = r$. The correctness of the reduction is clear.

Conversely, consider a system of target-equations over S_Γ. We can assume that the left-hand side of each equation contains no constants (for we can introduce dummy variables $x_c = c$) and no more than two variables (since $xyz = 0$ for all $x, y, z \in S_\Gamma$). Furthermore, we can replace any equation of the form $xy = \langle d_i d_j \rangle$ by the pair $x = d_i$, $y = d_j$. In any solution to the system, a variable x that occurs in a target-equation of the form $xy = r$, $yx = r$, or $x = d_i$ must be one of the generators. Consider an equation of the form $xy = 0$: if both x and y are forced to be generators, then the system will be unsatisfiable. Otherwise, this equation can be removed without affecting the system's satisfiability. Hence, we can assume that our system of equations contains only equations of the form $x = d_i$ and $xy = r$ and the reduction to CSP(Γ) is obvious. $\square$

On the other hand, T-EQN$^*_{S_k}$ is computable in polynomial time for the semigroup[6] S_k generated by the k element set $D = \{d_1, \ldots, d_k\}$ and subject to $xyz = 0$ for all $x, y, z \in S$.

Indeed, we can assume that each target equation of a system $\mathcal{E}$ over S_k involves at most two variables and that no constants appear on the left. Moreover, equations of the form $x_i x_j = d$ for some $d \in D$ have no solution so we can assume they do not occur in the system. To solve $\mathcal{E}$ we consider the set F of variables that are *not* bound by an equation $x_i = c$, with $c \in S_k$. If F is empty, the satisfiability of $\mathcal{E}$ is trivial to verify and our algorithm can replace any equation of the form $x_i x_j = ab$ with $a, b \in D$ by the pair $x_i = a$ and $x_j = b$. Once all such equations have been removed, every variable x_i remaining in F occurs only in equations of the form $x_i x_j = 0$ or $x_j x_i = 0$ so we can safely set it to 0 and delete the corresponding equations: the solvability of the remaining system is trivial to check.

The existence of such a polynomial time algorithm for T-EQN$^*_{S_k}$ is surprising since we have just shown that arbitrary CSP's are equivalent to T-EQN$^*_T$ for some

[6] In semigroup jargon, S_k is the free nilpotent semigroup of threshold 3 on k generators.

T in the variety generated by the S_k's and so the class of semigroups S for which T-EQN^*_S (and as we will see EQN^*_S) lies in P does not form a variety.

Theorem 6 describes the complexity of EQN^*_M for every monoid M and in light of Theorem 5 we would expect that is also possible to describe the complexity of EQN^*_S for every *regular* S. Our results already allow us to show two basic results in this direction.

Lemma 25. *If S is regular but is not in* $\mathbf{NB} \vee \mathbf{Ab}$ *then* EQN^*_S *is NP-complete.*

Proof. We already know from Theorem 5 that EQN^*_S is NP-complete unless S lies in $\mathbf{RB} \vee \mathbf{Ab}$. If S does lie in $\mathbf{RB} \vee \mathbf{Ab}$, then $E(S)$ (an inducible subsemigroup of S) must form a band which is not normal and the NP-completeness follows from Lemma 13. □

Lemma 26. *If S lies in* $\mathbf{NB} \vee \mathbf{Ab}$*, then* EQN^*_S *is polynomial-time equivalent to* $\mathrm{EQN}^*_{E(S)}$.

Proof. Only one direction requires proof since $E(S)$ is an inducible subsemigroup of S. For the converse, note that every band K in $\mathbf{NB}$ and every group G in $\mathbf{Ab}$ satisfy the identity $uxyv = uyxv$ so this also holds in S. Consider the relation $\sim$ on S defined by $x \sim y$ when $x^\omega y x^\omega = x$ and $y^\omega x y^\omega = y$. In fact, we claim that $x \sim y$ iff $uxv = uyv$ for any $u, v \in S$. The right to left implication is clear since $x^{\omega+1} = x$ in a union of groups. For the converse, we get first

$$y^\omega x^\omega y^\omega = y^{\omega^2} x^\omega y^{\omega^2} = (y^\omega x y^\omega)^\omega = y^\omega$$

and thus

$$uxv = ux^\omega y x^\omega v = ux^\omega y^\omega y^\omega y v = uy^\omega x^\omega y^\omega y v = uy^{\omega+1} v = uyv.$$

Hence, $\sim$ is an equivalence relation and in fact a congruence. Also $xy \sim yx$ so the quotient semigroup $Q = S/\sim$ is commutative and since it is also a union of groups then $Q \in \mathbf{SL} \vee \mathbf{Ab}$ [13]. For $q \in S$, we will denote by $[q]$ the $\sim$-class of q.

Let $\mathcal{E}$ be a system of equations over S: if $\mathcal{E}$ is satisfiable, then the two corresponding systems over $E(S)$ and Q must also be satisfiable. We claim that the converse statement also holds: let $\boldsymbol{e} = (e_1, \ldots, e_n)$ and $\boldsymbol{q} = ([q_1], \ldots, [q_n])$ be the respective solutions of these systems. We show that $\boldsymbol{s} = (e_1 q_1 e_1, \ldots, e_n q_n e_n)$ is a solution to $\mathcal{E}$. Every equation of $\mathcal{E}$ is either of the form $x_i x_j = x_k$ or $x_i = c$. In the first case, we have:

$$\begin{aligned} e_i q_i e_i e_j q_j e_j &= e_i e_j q_i q_j e_i e_j && \text{Since } S \text{ satisfies } uxyv = uyxv \\ &= e_i e_j q_k e_i e_j && \text{Since } q_i q_j \sim q_k \\ &= e_k q_k e_k && \text{Since } e_i e_j = e_k \\ &= e_k q_k e_k. \end{aligned}$$

For the second case, we get $e_i = c^\omega$ and $q_i \sim c$ so $e_i q_i e_i = c$.

Since we can check in polynomial time the solvability of the system over Q, we can reduce the solvability of $\mathcal{E}$ to that of the system over the normal band $E(S)$. □

It thus remains to understand the complexity of EQN^*_S when S is a normal band. We will first consider the a priori simpler problem of systems over *right-normal bands* which satisfy $x^2 = x$ and $xyz = yxz$. An important consequence of these identities is that if S is a right normal band and s and t are $\mathcal{J}$-related then $su = tu$ for each $u \in S$. Also, all $\mathcal{L}$-classes of S are trivial, i.e. $\mathcal{J} = \mathcal{R}$.

Theorem 8. *For every domain D and every set of relations Γ over D, there exists a right-normal band S_Γ such that $CSP(\Gamma)$ is polynomial-time equivalent to* $\mathrm{EQN}^*_{S_\Gamma}$.

Proof. Again, we can assume w.l.o.g. that Γ contains a single binary relation R and the constant relations. Let $r = |R|$ and $d = |D|$ with $D = \{t_1, \ldots, t_d\}$. We construct a semigroup S_Γ with 7 $\mathcal{J}$-classes $\{\alpha, \beta, \gamma, \delta, \epsilon, \rho, 0\}$ as in the egg-box picture of Figure 2. These $\mathcal{J}$-classes form a semilattice $S/\mathcal{J}$ also represented in the figure. The $\alpha, \beta, \gamma, \delta$ and ϵ classes all have d elements and we think of the elements $\alpha_i, \beta_i, \gamma_i, \delta_i, \epsilon_i$ as "representing" the element t_i of D. The ρ-class has r elements and we similarly think of each element of ρ as representing a pair $(t_i, t_j) \in R$.

We want S_Γ to be a right-normal band so for any x, x', y with $x \mathcal{J} x'$ we have $xy = x'y$. To stress this, we will abuse notation and when x lies, say in the $\mathcal{J}$-class γ, write $xy = \gamma y$.

The multiplication is described by the following equalities. First,

$$\begin{aligned} \beta\alpha_i = \epsilon\alpha_i = \alpha\beta_i = \epsilon\beta_i = \alpha\epsilon_i = \beta\epsilon_i &= \epsilon_i \\ \alpha\gamma_i = \gamma\alpha_i = \rho\alpha_i = \rho\gamma_i &= \gamma_i \\ \beta\delta_i = \delta\beta_i = \rho\beta_i = \rho\delta_i &= \delta_i \end{aligned}$$

for every $1 \leq i \leq d$. Next, if ρ_k is associated to the pair (t_i, t_j) we have

$$\alpha\rho_k = \gamma\rho_k = \gamma_i \quad \text{and } \beta\rho_k = \delta\rho_k = \delta_j$$

so that multiplying ρ_k on the left by γ and δ respectively "extracts" the information about the first and second component of the pair (t_i, t_j) that ρ_k represents. Also, for any $x \,\mathcal{J}\, y$ we have $xy = y$ because each $\mathcal{J}$-class is a single $\mathcal{R}$-class.

The element x lies in the $\mathcal{J}$-class, say, γ if and only if $\gamma_1 x = x$ and $x\gamma_1 = \gamma_1$. We will abbreviate this pair of equations by simply writing $x \in \gamma$ and similarly for the other $\mathcal{J}$-classes.

To get a reduction from $CSP(\Gamma)$ to $\mathrm{EQN}^*_{S_\Gamma}$, we create, for each CSP variable v_i two variables x_i and y_i and include equations:

$$(1)\ x_i \in \alpha \qquad (2)\ y_i \in \beta \qquad (3)\ \epsilon x_i = \epsilon y_i.$$

Since we have $x_i \in \alpha$ and $y_i \in \beta$, Equation (3) imposes that $x_i = \alpha_j$ iff $y_i = \beta_j$ and thus "synchronizes" the two variables. If our CSP instance uses the constant relation $v_i = t_j$ then we further set $x_i = \alpha_j$ (which also forces $y_i = \beta_j$). Next, for each constraint $(v_p, v_q) \in R$ we create a variable $z_{p,q}$ and include in our system the equations

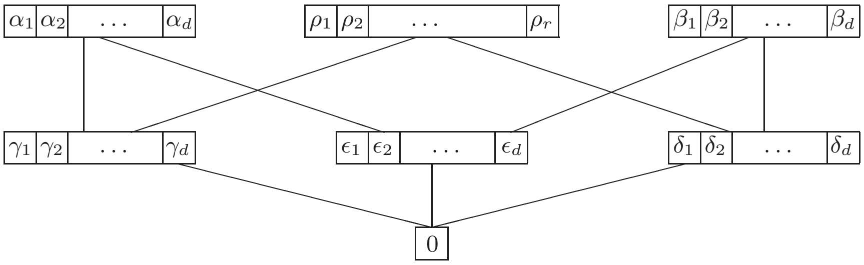

Fig. 2. Egg-box representation of S_Γ. Lines indicate the order in the semilattice of $\mathcal{J}$-classes.

$$(4)\ z_{p,q} \in \rho \qquad (5)\ \gamma z_{p,q} = \gamma x_p \qquad (6)\ \delta z_{p,q} = \delta y_q.$$

If $x_p = \alpha_i$ and $y_q = \beta_j$ then our multiplication rules show that there is a $z_{p,q}$ satisfying these equations iff $(t_i, t_j) \in R$. It is now clear that this system is satisfiable if and only if the CSP(Γ) instance is.

Conversely, suppose we start with a system $\mathcal{E}$ of equations over S_Γ and assume w.l.o.g. that all equations are of the form $x = s$ or $x_i x_j = x_k$. We will show that we can construct in polynomial time a system $\mathcal{F}$ which is exactly in the form we have just described and which has a solution if and only if $\mathcal{E}$ does. It is then easy to reconstruct a CSP(Γ) instance which is satisfiable if and only if $\mathcal{F}$ is.

As in Lemma 26, if $\mathcal{E}$ has a solution, then the corresponding system over the semilattice consisting of elements $\{\alpha, \beta, \gamma, \delta, \epsilon, \rho, 0\}$ must also have a solution. We can check this in polynomial time and if a solution exists, we can obtain the minimal one $\bar{e} = (e_1, \ldots, e_n)$ with each $e_i \in \{\alpha, \beta, \gamma, \delta, \epsilon, \rho, 0\}$. Once again, it can be shown that if $\mathcal{E}$ has any solution $(u_1, \ldots, u_n)$ then it has one, namely $(e_1 u_1, \ldots, e_n u_n)$ where x_i lies in the $\mathcal{J}$-class e_i.

Therefore, we do not affect the solvability of $\mathcal{E}$ if we add for each x_i the pair of equations equivalent to $x_i \in e_i$. Furthermore, we can replace every equation $x_i x_j = x_k$ by $e_i x_i e_j x_j = e_k x_k$. Furthermore, because $e_i x_i e_j x_j = e_i e_j x_j$ and since we must have $e_i e_j = e_k$ in the semilattice, the equation can be rewritten as $e x_j = e x_k$ with $e \in \{\alpha, \beta, \gamma, \delta, \epsilon, \rho, 0\}$.

We are thus left with solving a system, say $\mathcal{E}'$, where every variable is constrained by some condition $x_i \in e_i$ and every other equation is of the form $x = c$ for c a constant or $e x_i = e x_j$. Moreover, if the system contains an equation $e x_i = e x_j$ then $e_i \geq_{\mathcal{J}} e$. Of course every equation of the form $0x = 0y$ is trivially satisfied and can be discarded.

Suppose a variable x is constrained by $x \in \gamma$. In every equation of the form $ex = ey$ in which x occurs, we have $e \leq_{\mathcal{J}} \gamma$ and this means in fact $e = \gamma$ since we removed equations with $e = 0$. Suppose that in $\mathcal{E}'$ we replace the requirement $x \in \gamma$ by $x \in \alpha$ and the equation $x = \gamma_i$ (if such an equation exists) by $x = \alpha_i$: we claim that the solvability of the system will be unaffected. Indeed, whenever we had a solution with $x = \gamma_i$, we will have a solution with $x = \alpha_i$ since for any

$1 \leq i \leq d$ we have $\gamma\gamma_i = \gamma\alpha_i$. By the same token, we can replace requirements $x \in \delta$ or $x \in \epsilon$ by $x \in \beta$ without affecting the solvability of the system.

These observations allow us to assume that every variable in our system is constrained by $x_i \in e_i$ where e_i is one of $\{\alpha, \beta, \rho\}$.

Let r_i, r_j be variables of the system constrained by $r_i \in \rho$ and $r_j \in \rho$. We can assume that our system does not contain an equation of the form $\gamma r_i = \gamma r_j$ since it is equivalent to the equations $\gamma r_i = \gamma z, \quad \gamma r_j = \gamma z$ with $z \in \alpha$ and where z is a new variable. Symmetrically, we can replace equations of the form $\delta r_i = \delta r_j$.

In the same way, for any constant ρ_s there are unique constants α_u and β_v such that $\gamma\rho_s = \gamma\alpha_u$ and $\delta\rho_s = \delta\beta_v$ so we can replace $r_i = \rho_s$ by the equations $x = \alpha_u, \quad y = \beta_v, \quad \gamma r_i = \gamma x$ and $\delta r_i = \delta y$. Also, if we have $x_i \in \alpha$ and $x_j \in \alpha$ then the equation $\epsilon x_i = \epsilon x_j$ (and similarly $\gamma x_i = \gamma x_j$) forces $x_i = x_j$. In the same way, for $z \in \rho$ (resp. $y \in \beta$) and $x_i, x_j \in \alpha$ the pair of equations $\gamma x_i = \gamma z$ and $\gamma x_j = \gamma z$ (resp. $\epsilon x_i = \epsilon y$ and $\epsilon x_j = \epsilon y$) yields $x_i = x_j$. Symmetrical results hold for y_i and y_j in β.

In the system $\mathcal{F}$ thus obtained, we can view each variable as being one of three types X, Y, Z where each $x_i \in X$ is constrained by $x_i \in \alpha$, each $y_i \in Y$ by $y_i \in \beta$ and each $z_i \in Z$ by $z_i \in \rho$. For each $z_k \in Z$, there is at most one (exactly one if we use dummy variables) $x_i \in X$ and one $y_j \in Y$ such that the system contains $\gamma x_i = \gamma z_k$ and $\delta y_j = \delta z_k$. Also, for each $x_i \in X$, there exists exactly one $y_j \in Y$ such that the system contains $\epsilon x_i = \epsilon y_j$. All other equations are $x_i = \alpha_s$ or $y_j = \beta_s$ for some s. By construction, $\mathcal{F}$ has a solution if and only if $\mathcal{E}$ had one and $\mathcal{F}$'s solvability easily reduces to a CSP(Γ) instance. $\square$

In contrast, an easy exercise can show that solving a system of equations over the free right-normal band or even the free normal band on any finite number k of generators is doable in polynomial time. This also is part of more general results of one of the authors [16]. So the class of semigroups for which EQN^*_S lies in P is not closed under morphic images. It is not closed under taking subsemigroups either. Indeed, the right-normal band S_Γ that we constructed is a subsemigroup of the right-normal band that we would obtain when R consists of all pairs in $D \times D$ but the latter CSP clearly lies in P.

If S is not regular, then our partial results allow us to show that EQN^*_S and T-EQN^*_S are NP-complete unless S lies in the variety $\mathbf{DO} \cap \overline{\mathbf{Ab}}$ of semigroups whose regular $\mathcal{J}$-classes are orthodox unions of Abelian groups and which has already been shown of particular relevance in computational complexity contexts [23, 24].

7 Conclusion

Although the complexity of EQN^*_S or T-EQN^*_S for semigroups is a question that will not find a resolution until we can settle the CSP conjecture, we are able to give complete dichotomies in the case of monoids and the classes of monoids for which each problem is tractable form varieties. We do not have a good explanation for this phenomenon and it would be interesting to see, for instance, whether one can get a simple and direct proof (in the case of monoids) that the tractability of EQN^*_M implies the tractability of EQN^*_S for every subsemigroup $S \subseteq M$. As we noted, B. Larose and L. Zádori have reproved using universal algebra that EQN^*_M is NP-complete if M is not in $\mathbf{SL} \vee \mathbf{Ab}$ and their proof crucially depends on the presence of an identity element [18].

Reducing every CSP to the problem of solving systems of equations over a finite semigroup might be useful given that so many of the successful machinery to study CSP is of algebraic nature. In any case, it would be surprising if equations over such simple classes of semigroups defined problems whose complexity form a very wide spectrum and in that sense, these results constitute additional if weak evidence in favor of the CSP conjecture.

Finally, let us note that EQN^*_G for a finite group can be solved well within NC (and thus has a very efficient parallel algorithm) but EQN^*_S is P-hard as soon as S contains two idempotents $e \neq f$ such that $ef = fe = f$. Indeed, Lemma 1 essentially shows that in this case the P-complete problem HORN-SAT has a logspace reduction to EQN^*_S. In particular, the semigroup S_Γ which we construct in Theorem 8 is such that $\mathrm{EQN}^*_{S_\Gamma}$ will be P-complete, even if $\mathrm{CSP}(\Gamma)$ has much lower complexity.

Acknowledgments. Cris Moore was a key participant in our first investigations of this topic. We also want to thank Andrei Krokhin for suggesting that Theorem 8 might be true and Benoit Larose for many helpful discussions about CSP's and universal algebra. We are grateful to Gustav Nordh and Peter Jonsson for sending us a preprint of their paper and discussing their findings.

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 *Proc. Math. Foundations of Comp. Sci (MFCS'00)*, pages 172–181, 2000.
3. A. Bulatov. A dichotomy theorem for constraints on a three-element set. In *Proc. of 43rd Foundations of Comp. Sci. (FOCS'02)*, pages 649–658, 2002.
4. A. Bulatov, P. Jeavons, and M. Volkov. Finite semigroups imposing tractable constraints. In G. Gomez, P. Silva, and J-E.Pin, editors, *Semigroups, Algorithms, Automata and Languages*. WSP, 2002.
5. A. Bulatov, A. Krokhin, and P. Jeavons. Constraint satisfaction problems and finite algebras. In *Proceedings 27th International Colloquium on Automata, Languages and Programming—ICALP'00*, volume 1853 of *Lecture Notes in Computer Science*, pages 272–282, 2000.
6. V. Dalmau. *Computational Complexity of Problems over Generalized Formula.* PhD thesis, Universita Politécnica de Catalunya, 2000.
7. V. Dalmau, R. Gavaldà, P. Tesson, and D. Thérien. Notes on relations closed under expressions over semigroups. Document in preparation, 2004.
8. V. Dalmau and J. Pearson. Closure functions and width 1 problems. In *Principles and Practice of Constraint Programming—CP'99*, volume 1713 of *Lecture Notes in Computer Science*, pages 159–173, 1999.

9. 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.
10. M. Goldmann and A. Russell. The computational complexity of solving equations over finite groups. *Information and Computation*, 178:253–262, 2002.
11. P. Hell and J. Nešetřil. On the complexity of H-coloring. *J. of Combinatorial Theory (Series B)*, 48:92–110, 1990.
12. 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.
13. J. Howie. *Fundamentals of Semigroup Theory*. Claredon Press, Oxford, 1995.
14. P. Jeavons, D. Cohen, and M. Gyssens. Closure properties of constraints. *J. ACM*, 44(4):527–548, 1997.
15. P. Jonsson and G. Nordh. The complexity of counting the number of solutions to systems of equations over finite algebraic structures. Preprint, May 2003.
16. O. Klíma. Complexity of unification and matching problems in the varieties of idempotent semigroups. To appear in *Int. J. of Algebra and Computation*, 2004.
17. O. Klíma. *Unification Modulo Associativity and Idempotency*. PhD thesis, Masaryk University, 2004.
18. B. Larose and L. Zádori. Taylor terms, constraint satisfaction and the complexity of polynomial equations over finite algebras. Submitted for publication, 2004.
19. C. Moore, P. Tesson, and D. Thérien. Satisfiability of systems of equations over finite monoids. In *MFCS'01*, pages 537–547, 2001.
20. J.-E. Pin. *Varieties of formal languages*. North Oxford Academic Publishers Ltd, London, 1986.
21. T. J. Schaefer. The complexity of satisfiability problems. In *Proc. 10^{th} ACM STOC*, pages 216–226, 1978.
22. S. Seif and C. Szabó. Algebra complexity problems involving graph homomorphism, semigroups and the constraint satisfaction problem. *Journal of Complexity*, 19(2):153–160, 2003.
23. P. Tesson. *Computational Complexity Questions Related to Finite Monoids and Semigroups*. PhD thesis, McGill University, 2003.
24. P. Tesson and D. Thérien. Complete classifications for the communication complexity of regular languages. *Theory of Computing Systems*, 2004. Accepted for publication.

Appendix

We present here detailed proofs for the hardness results which were omitted in Section 3 and for the semigroup theoretic lemmas stated throughout the paper.

Additional Hardness Results

Lemma 16. *If S is a union of groups such that $\mathcal{H}$ is not a congruence on S, then* T-EQN$^*_S$ *is NP-complete.*

Proof. If $\mathcal{H}$ is not a congruence, then it either is not a right-congruence or is not a left congruence. We can without loss of generality assume the first case, i.e. that there exist a, b, c such that $a\mathcal{H}b$ but $ac \not\mathcal{H} bc$, since the other case can be handled symmetrically. The reduction that we will use is almost exactly that of Example 1 and this will be made possible by the following technical construction.

Lemma 27. *If S is a union of groups such that $\mathcal{H}$ is not a right congruence on S, then there are $a, e \in S$ with $a >_{\mathcal{J}} e$ such that*

1. *$e^2 = e$ and $a^\omega e = e = ea^\omega$;*
2. *$ae \not\mathcal{H} e$ (and in fact $ae \not\mathcal{R} e$);*
3. *For all x with $a >_{\mathcal{J}} x >_{\mathcal{J}} e$, we have $xe \neq e$.*

Proof. Recall that since S is a union of groups we have $s^{\omega+1} = s$ for all $s \in S$ and $\mathcal{J}$ is a congruence over S. There are $a, b, c \in S$ such that $a \mathcal{H} b$ but $ac \not\mathcal{H} bc$ and we can choose a, b to be $\mathcal{J}$-minimal with this property. Since S is a union of groups we have $a^\omega = b^\omega$ so either $a^\omega c \not\mathcal{H} ac$ or $b^\omega c \not\mathcal{H} bc$ for otherwise $ac\mathcal{H}a^\omega c = b^\omega c\mathcal{H}bc$ contradicting our initial hypothesis. Let us thus assume that $a^\omega c \not\mathcal{H} ac$: because of Green's Lemma, we cannot have $a \mathcal{J} c$ and in fact we can assume that c lies $\mathcal{J}$-below a (otherwise, we can choose $d = a^\omega c$ and obtain $a^\omega d = a^\omega c \not\mathcal{H} ac = a^{\omega+1}c = ad$). Let us pick c as a $\mathcal{J}$-maximal element lying below a with the property $ac \not\mathcal{H} a^\omega c$: because S is a union of groups we in fact have $c \mathcal{J} ac \mathcal{J} a^\omega c$. If f is the idempotent $(a^\omega c)^\omega$ then $f \mathcal{H} a^\omega c$ and by Green's Lemma $af \mathcal{H} ac \not\mathcal{H} a^\omega c \mathcal{H} f$. In fact, since af and $a^\omega f = f$ are $\mathcal{L}$-related they cannot be $\mathcal{R}$-related. Finally, choosing $e = (fa^\omega)^\omega$, we get $a^\omega e = e = ea^\omega$ and, again using Green's Lemma, $ae \mathcal{R} af \not\mathcal{R} f \mathcal{R} e$ so in particular $ae \not\mathcal{R} e$.

If $a >_{\mathcal{J}} x >_{\mathcal{J}} e$ then we have, from our initial choice of c above $ax \mathcal{H} a^\omega x$. Similarly, because we earlier chose a, b to be $\mathcal{J}$-minimal, we must have $axe \mathcal{H} a^\omega xe$. Hence, $xe \neq e$ for otherwise this last expression simplifies to $ae \mathcal{H} e$. □

Note that in Example 1, a and e exactly had these properties. When building our reduction, we work over the inducible subsemigroup of elements lying $\mathcal{J}$-below a and, as in the example, introduce for every Boolean variable X_i the equations

$$\begin{array}{llll} (1) & x_i e = e & (2) & \bar{x}_i e = e \\ (3) & v_i x_i s_i e = e & (4) & v_i a \bar{x}_i t_i e = e \end{array}$$

where v_i, s_i, t_i are constrained to be $\mathcal{H}$-related to a.

Just as before, we also include for each 3SAT clause, e.g. $X_1 \vee \overline{X}_2 \vee X_3$ the equation

$$(5) \quad x_1\bar{x}_2x_3 = e.$$

Given an assignment to the Boolean literals satisfying the 3SAT formula, one can verify that this system has a solution by setting $x_i = e$, $\bar{x}_i = a^\omega$, $t_i = a^{\omega-1}$ and $v_i = s_i = a^\omega$ whenever X_i is TRUE and $x_i = a^\omega$, $\bar{x}_i = e$, $s_i = a$, $t_i = a^\omega$ and $v_i = a^{\omega-1}$ whenever X_i is FALSE. Furthermore, each equation of type (5) will be satisfied since at least one of the three terms is e while the others are a^ω and since we have $e = a^\omega e = ea^\omega$.

Conversely, suppose that there exists a solution to the constructed system. Equations (1,2) show that x_i and $\bar{x}_i$ are $\mathcal{J}$ related either to e or to a because of condition 3 in Lemma 27. On the other hand if *both* values are $\mathcal{J}$-related to e the because of Eqn. (1) we have $x_i \geq_{\mathcal{R}} e$ and thus $x_i \mathrel{\mathcal{R}} e$ since $x_i \mathrel{\mathcal{J}} e$ and similarly $\bar{x}_i \mathcal{R} e$. Also $v_i x_i s_i e = e$ so $v_i x_i \geq_{\mathcal{R}} e$ and in fact get $v_i x_i \mathrel{\mathcal{R}} e$ when $x_i \mathrel{\mathcal{J}} e$. Similarly $v_i a \bar{x}_i \mathrel{\mathcal{R}} e$. Since $\mathcal{R}$ is a left-congruence we have also $v_i e \mathrel{\mathcal{R}} v_i x_i \mathrel{\mathcal{R}} e$ and $v_i ae \mathrel{\mathcal{R}} v_i a \bar{x}_i \mathrel{\mathcal{R}} e$. But this leads to a contradiction for if $ve \mathrel{\mathcal{R}} vae$ then multiplying both sides on the left by $v^{\omega-1}$, we get $e = v^\omega e \mathrel{\mathcal{R}} v^\omega ae = ae$.

Hence, if we set X_i (resp. $\overline{X}_i$) to TRUE when $x_i \mathrel{\mathcal{J}} e$ (resp. $\bar{x}_i \mathrel{\mathcal{J}} e$) then a litteral and its complement are never both true. Finally, if $x_1\bar{x}_2x_3 = e$, one of those three variables must be $\mathcal{J}$-related to e since their product will otherwise lie in the $\mathcal{J}$-class of a. □

Before proving Lemma 17, we recall some useful properties of bands, all of which can be obtained from results mentioned in our introduction, and provide a useful characterization of regular bands.

Lemma 28 ([13]). *Let S be a band and $a, b, c \in S$ arbitrary elements. Then*
(i) $a \leq_{\mathcal{L}} b \iff ab = a$,
(ii) $a \leq_{\mathcal{R}} b \iff ba = a$,
(iii) $a \leq_{\mathcal{J}} b \iff aba = a$,
(iv) $a \leq_{\mathcal{J}} b \implies ac \leq_{\mathcal{J}} bc,\ ca \leq_{\mathcal{J}} cb$.

Lemma 29. *Let S be a band. Then it is a regular band if and only if the following two (dual) conditions are satisfied:*

$$(\forall x, y, z \in S) \quad x \mathrel{\mathcal{R}} y, x \geq_{\mathcal{J}} z \implies xz = yz,$$

$$(\forall x, y, z \in S) \quad x \mathrel{\mathcal{L}} y, x \geq_{\mathcal{J}} z \implies zx = zy.$$

Proof. Let S be a regular band and $x, y, z \in S$ such that $x \mathrel{\mathcal{R}} y, x \geq_{\mathcal{J}} z$. From the assumption $x \mathrel{\mathcal{R}} y$ we have $xy = y$ and from $x \geq_{\mathcal{J}} z$ we can deduce $zxz = z$. Hence $yz = xyzxz$ and if we use regularity (i.e. $xyzx = xyxzx$) we obtain $yz = xyzxz = xyxzxz$. Because $xyx = x$ and S is a band we finally obtain $yz = xzxz = xz$. The second condition can be obtain dually.

Conversely, let S be a band which satisfies both conditions and let $a, b, c \in S$ be arbitrary elements. If we let $x = ab$, $y = aba$ and $z = caabca$ then these elements satisfy the assumptions of the first condition and we have $abca = abca \cdot abca = xz = yz = abaca \cdot abca$. Hence $abaca \mathcal{R} abca$. Dually we obtain $abaca \mathcal{L} abca$ and alltogether we have $abaca \,\mathcal{H}\, abca$ which means that $abaca = abca$ as S is a band. $\square$

Lemma 17. *If S is a band but is not a regular band, then* T-EQN$^*_S$ *is NP-complete.*

Proof. If S is a band but not a regular band, it must violate one of the two conditions of Lemma 29 and we will w.l.o.g. assume that S violates the first one. The following construction will now allow us to present a reduction closely related to the one of Example 2.

Lemma 30. *Let S be a band which does not safisfy the first condition of Lemma 29. Then there exist elements $a, b, c \in S$ such that $ab = b$, $ba = a$ (i.e. $a \,\mathcal{R}\, b$), $ca = c$, $ac \neq bc$, $c <_{\mathcal{J}} a$ and satisfying the following condition*

$$\forall s \in S : \quad c <_{\mathcal{J}} s \leq_{\mathcal{J}} a \implies as = bs. \tag{1}$$

Proof. Let x, y, z be elements which disprove the condition of Lemma 29 and such that z is $\mathcal{J}$-maximal with respect to this property. By Lemma 5 and the aperiodicity of S we can see that $x \,\mathcal{R}\, y \,\mathcal{J}\, z$ implies $xz = xyz = yz$. Hence $z <_{\mathcal{J}} x$ and we can put $a = x$, $b = y$ and $c = zx$. Assume for a moment that $ac = bc$. Then $xyx = ac = bc = yzx$ and if we multiply this equality by z then we obtain $xzxz = yzxz$ which is $xz = yz$ because $zxz = z$. This is a contradiction so $ac \neq bc$. The equalities are easy to see and property (1) is a consequence of the $\mathcal{J}$-maximality of z in the counterexample. $\square$

We can now obtain the following reduction from 3SAT to T-EQN$^*_T$ where T is the (inducible) semigroup of elements lying $\mathcal{J}$-below a. For each Boolean literal X_i in the formula, we introduce the variables $x_i, \bar{x}_i, y_i$ and construct the equations

$$\begin{array}{llll}
(1) & cx_i = c & (2) & c\bar{x}_i = c \\
(3) & ax_ia\bar{x}_i = ac & (4) & bx_ib\bar{x}_i = bc \\
(5) & y_ix_iac = ac & (6) & y_i\bar{x}_ibc = bc \\
(7) & y_ia = a & &
\end{array}$$

Moreover for any q which is $\mathcal{R}$-related to a we add the equations

$$(8) \quad qx_iqc = qc \qquad\qquad (9) \quad q\bar{x}_iqc = qc.$$

Note that in any solution to these equations we know from Eqs. (1,2) that both x_i and $\bar{x}_i$ lie $\mathcal{J}$-above c. Suppose that both lie *strictly* $\mathcal{J}$-above c then

$ax_i = bx_i$ and $a\bar{x}_i = b\bar{x}_i$ by Lemma 30. But then $ax_i a\bar{x}_i = bx_i b\bar{x}_i$ and this contradicts Eqs. (3,4).

Suppose on the other hand that both x_i and $\bar{x}_i$ are $\mathcal{J}$-related to c: by Eqs. (1,2) we get $x_i \,\mathcal{L}\, \bar{x}_i \,\mathcal{L}\, c$. We thus have $x_i = x_i ac$ and in fact $x_i tc = x_i$ (as well as $\bar{x}_i tc = \bar{x}_i$) for any $t \geq_{\mathcal{J}} c$. Since Eq. (7) imposes $y_i \,\mathcal{R}\, a$ we deduce from Eqs. (8,9) that

$$y_i x_i ac = y_i x_i = y_i x_i y_i c = y_i c = y_i \bar{x}_i y_i c = y_i \bar{x}_i bc.$$

This, however contradicts Eqs. (5,6). Hence, exactly one of $x_i, \bar{x}_i$ is $\mathcal{J}$-related to c and the other lies strictly $\mathcal{J}$-above c.

We complete our reduction by introducing, for each of clause of the 3SAT formula, e.g. $X_1 \vee \overline{X}_2 \vee X_3$, the pair of the equations:

$$(10) \quad ax_1 a\bar{x}_2 ax_3 = ac \qquad\qquad (11) \quad bx_1 b\bar{x}_2 bx_3 = bc$$

One can verify that if the 3SAT instance is satisfiable, then we can satisfy the system obtained through our reduction by letting $x_i = c$, $\bar{x}_i = a$, $y_i = a$ whenever X_i is TRUE, and $x_i = a$, $\bar{x}_i = c$, $y_i = b$ whenever X_i is FALSE.

Conversely, suppose the system of the equations is satisfiable. Since exactly one of $x_i, \bar{x}_i$ is $\mathcal{J}$-related to c, we get a consistent truth assignment to the literals by setting X_i (resp. $\overline{X}_i$) to TRUE if and only if $x_i \,\mathcal{J}\, c$ (resp. $\bar{x}_i \,\mathcal{J}\, c$). This assignment satisfies every clause of the original formula for if the variables occurring in Eq. (11) all lie strictly $\mathcal{J}$-above c we have $ax_1 = bx_1 \quad a\bar{x}_2 = b\bar{x}_2$ and $ax_3 = bx_3$ so that

$$ax_1 a\bar{x}_2 ax_3 = bx_1 b\bar{x}_2 bx_3$$

in violation of Eqs. (11,12). □

To present the complete proof of Lemma 18, we need to introduce semigroup theoretic tools which allow a deep understanding of the structure of completely simple semigroups. We seek a refinement of Lemma 5 in order to understand the structure of multiplication within such semigroups.

Let G denote some finite group with multiplication $\circ$ and m, n be positive integers. A *complete Rees matrix* is an m by n matrix R with entries in G and the corresponding *complete Rees semigroup* is the completely simple semigroup with elements in $([n] \times G \times [m])$ and where the multiplication of elements is given by:

$$(i_1, g_1, j_1) \cdot (i_2, g_2, j_2) = (i_1,\; g_1 \circ R_{j_1,i_2} \circ g_2\;, j_2).$$

Note that this semigroup has n $\mathcal{R}$-classes given for each $1 \leq l \leq n$ by $\mathcal{R}_l = \{(l, g, j) | g \in G; 1 \leq j \leq m\}$ and similarly has m $\mathcal{L}$-classes.

Theorem 9. *[13] Every completely simple semigroup with n $\mathcal{R}$-classes and m $\mathcal{L}$-classes is isomorphic to a complete Rees semigroup. The corresponding Rees matrix is $m \times n$ and its first row and first column entries can be assumed to all be the group identity 1_G. Moreover, the semigroup is orthodox if and only if all the Rees matrix entries are 1_G.*

In particular the m idempotents in the first $\mathcal{R}$-class and the n idempotents in the first $\mathcal{L}$-class are $\{(1, 1_G, j) | 1 \leq j \leq m\}$ and $\{(i, 1_G, 1) | 1 \leq i \leq n\}$ respectively. Note also that the egg-box picture of the complete Rees semigroup corresponding to an $m \times n$ complete Rees matrix has n rows ($\mathcal{R}$-classes) and m columns ($\mathcal{L}$-classes).

Lemma 18. *If S contains a $\mathcal{J}$-class T forming a completely simple but unorthodox semigroup then* T-EQN$^*_S$ *is NP-complete.*

Proof. First note that the subsemigroup $T_{\leq}$ of elements lying $\mathcal{J}$-below T is inducible in S. Furthermore, for any $t \in T$, the target-equation $(txt)^\omega = t^\omega$ defines (in $T_{\leq}$) the semigroup T. We can thus assume without loss of generality that S is itself completely simple unorthodox semigroup.

We consider the $m \times n$ complete Rees matrix R associated to S: we can assume that the first row and first column entries of R are all 1_G.

We can recursively reorder the rows and columns of R: suppose row s is such that $R_{s,i} = 1_G$ for every $i \leq t$. We choose the row $(s+1)$ as the one with the most number of 1_G entries among $R_{s+1,i}$ with $i \leq t$ and reorder the columns such that all these entries appear first in the row.

Because we assumed that the semigroup is not orthodox, there is some non-identity entry in R so after reordering, we can, as shown in Figure 3, find indices

$$\begin{pmatrix}
1_G & 1_G & \dots & \dots & \dots & \dots & \dots & \dots & 1_G \\
1_G & 1_G & \dots & \dots & \dots & \dots & \dots & \dots & 1_G \\
1_G & 1_G & \dots & \dots & 1_G & R_{a,b} & * & * & * \\
1_G & 1_G & \dots & \dots & 1_G & * & * & * & * \\
1_G & \dots & 1_G & * & * & R_{c,b} & ? & ? & ? \\
1_G & 1_G & * & ? & ? & ? & ? & ? & ? \\
1_G & * & ? & ? & ? & ? & ? & ? & ?
\end{pmatrix}$$

Fig. 3. Rees matrix of S after reordering: all entries above the dotted line are 1_G. The $*$'s represent entries which *must* be unequal to 1_G.

a, b, c with $1 < b \leq n$ and $1 < a < c \leq m + 1$ and such that
- $R_{a,b} \neq 1_G$;
- if $1 \leq j < a$ then $R_{j,i} = 1_G$ for all $1 \leq i \leq n$;
- if $a \leq j < c$ then $R_{j,i} = 1_G$ if and only if $i < b$.

We now mimic the reduction from 1-3SAT of Example 3: for each Boolean X_k we create variables $x_k, \bar{x}_k$, force them to be idempotent and impose

$$(1)\ x_k \bar{x}_k = (1, 1_G, 1) \qquad\qquad (2)\ x_k (b, 1_G, a) \bar{x}_k = (1, R_{a,b}, 1).$$

In any solution to the system, we must have $x_k = (1, 1_G, s_k)$ for some $1 \leq s_k \leq m$ since it is an idempotent and, by Eq. (1), lies in the first $\mathcal{R}$-class. Similarly, $\bar{x}_k = (t_k, 1_G, 1)$ for some $1 \leq t_k \leq n$. Eq. (1) thus also forces $R_{s_k,t_k} = 1_G$ and from Eq. (2) we have $R_{s_k,b} \cdot R_{a,t_k} = R_{a,b}$. Similarly we require that $R_{s_k,i} = 1_G$ for all $1 \leq i < b$ by using equations of the form:

$$(3) \quad x_k(i, 1_G, 1) = (1, 1_G, 1).$$

We thus have insured that $s_k < c$ and in fact that either $s_k < a$ or $t_k < b$ for otherwise $R_{s_k,t_k} \neq 1_G$.

For a clause $X_1 \vee \overline{X}_2 \vee X_3$ we wish to add the requirement $R_{s_1,b} \cdot R_{a,t_2} \cdot R_{s_3,b} = R_{a,b}$. This can be encoded as an equation such as:

$$(4) \quad x_1(b, 1_G, a)\bar{x}_2(1, 1_G, 1)x_3(b, 1_G, 1) = (1, R_{a,b}, 1).$$

If the 1-3SAT is satisfiable, then the system can be satisfied by setting $x_k = (1, 1_G, a)$ and $\bar{x}_k = (1, 1_G, 1)$ whenever X_k is TRUE and $x_k = (1, 1_G, 1)$ and $\bar{x}_k = (b, 1_G, 1)$ whenever X_k is FALSE.

For the converse, we assume w.l.o.g. (see Example 3) that $R_{a,b}$ does not have order 2. Note that if $s_k < a$ then $R_{s_k,b} = 1_G$ and so $R_{a,t_k} = R_{a,b}$ whereas if $t_k < b$ then $R_{a,t_k} = 1_G$ so $R_{s_k,b} = R_{a,b}$. So we can choose X_k to be TRUE if $R_{s_k,b} = R_{a,b}$ and $R_{a,t_k} = 1_G$ and X_k to be FALSE if $R_{s_k,b} = 1_G$ and $R_{a,t_k} = R_{a,b}$. For any 1-3SAT clause, say $X_1 \vee \overline{X}_2 \vee X_3$, we have $R_{s_1,b} \cdot R_{a,t_2} \cdot R_{s_3,b} = R_{a,b}$ and so exactly one of $R_{s_1,b}, R_{a,t_2}, R_{s_3,b}$ is $R_{a,b}$ and the other two are 1_G so exactly one literal per clause is TRUE. □

Other Technical Results

Lemma 8. *For a semigroup S, a variety of bands $\mathbf{B}$ and a variety of groups $\mathbf{H}$, the following are equivalent:*

1. *S is a strong $\mathbf{B}$-band of $\mathbf{H}$-groups;*
2. *S belongs to $\mathbf{B} \vee \mathbf{H}$.*
3. *S is an orthodox union of groups all of which lie in $\mathbf{H}$, such that $E(S)$ is a band in $\mathbf{B}$ and $\mathcal{H}$ is a congruence. In particular, the idempotents form a subsemigroup and $S/\mathcal{H} \equiv E(S)$.*

Proof. $(1 \Rightarrow 2)$ Suppose S is a strong $\mathbf{B}$-band of $\mathbf{H}$-groups with an underlying band $E = \{e_1, \ldots, e_k\}$ and a family of groups $\{G_e | e \in E\}$. Let $H = \prod_{e \in E} G_e$ and consider the subset T of $E \times H$ consisting of elements $(f, g_{e_1}, \ldots, g_{e_k})$ such that $g_{e_i} = \phi_{f,e_i}(g_f)$ for all $e_i \leq_{\mathcal{J}(E)} f$. One can show that T is subsemigroup of $E \times H$. We claim that S is a morphic image of T. Indeed, define $\psi : T \to S$ as

$$\psi(f, g_1, \ldots, g_k) = g_f.$$

Obviously, ψ is surjective. Moreover, it is a well-defined morphism since we can show that $\psi(f, g_1, \ldots, g_k) \cdot \psi(f', g'_1, \ldots, g'_k)$ is:

$$\begin{aligned}
&= g_f \cdot g'_{f'} \\
&= \phi_{f,ff'}(g_f) \cdot \phi_{f',ff'}(g'_{f'}) \\
&= g_{ff'} \cdot g'_{ff'} \\
&= \psi(ff', g_1 g'_1, \ldots, g_k g'_k)
\end{aligned}$$

Note that we are using the fact that $g_{ff'} = \phi_{f,ff'}(g_f)$.

$(2 \Rightarrow 3)$ If E is a band in $\mathbf{B}$ and G is a group in $\mathbf{H}$ then clearly $E \times G$ satisfies (3). One can easily show that any divisor of S of $E \times G$ is also an orthodox union of groups and that $E(S)$ divides E so that indeed $E(S) \in \mathbf{B}$. It is easy to see that $\mathcal{H}$ is a congruence over $E \times G$ and that this property is preserved by taking morphic images. Suppose S is a submonoid of $E \times G$ with $a, b, c \in S$ and $a \,\mathcal{H}\, b$. Then a and b are also $\mathcal{H}$-related in $E \times G$ so we have $ac\mathcal{H}bc$ in $E \times G$. The latter is a union of groups so we have $(ac)^\omega = (bc)^\omega$. So $bc = ac(ac)^{\omega-1}bc$ and $bc = bc(ac)^{\omega-1}ac$ and since $(ac)^{\omega-1}bc$ and $bc(ac)^{\omega-1}$ lay in S we indeed have $ac \,\mathcal{H}\, bc$ in S.

$(3 \Rightarrow 1)$ For any $a, b \in S$ we have $(ab)^\omega = a^\omega b^\omega$ since $\mathcal{H}$ is a congruence and S is orthodox. We denote $E = E(S)$ and $G_e = \mathcal{H}_e$ for any $e \in E$. For any idempotent $e \in S$ and any $x, y \in S$ with $x^\omega = y^\omega$ we have

$$exeye = exx^\omega e(ye)^\omega ye = exy^\omega ey^\omega eye = exy^\omega eye = exye.$$

Thus if e and f are idempotents with $f \geq_{\mathcal{J}(E)} e$ the map $\phi_{f,e} : G_f \to G_e$ given by $\phi_{f,e}(x) = exe$ is a well defined group homomorphism. Of course, $\phi_{e,e} = id_{G_e}$ and for any idempotents $d \geq_{\mathcal{J}(E)} e \geq_{\mathcal{J}(E)} f$ we have for any $x \in G_d$

$$\phi_{e,f} \circ \phi_{d,e}(x) = fexef = fe(xef)^\omega xef = fex^\omega efxef = fxef$$

since $ede = e$ and $fef = f$ hence

$$\phi_{e,f} \circ \phi_{d,e}(x) = fx(fx)^\omega ef = fxfdef = fxf = \phi_{d,f}(x).$$

Clearly, S is the union of the groups G_e. Multiplication in S for $x \in G_e$ and $y \in G_f$ is thus given by

$$x \cdot y = x^\omega y^\omega xx^\omega y^\omega yx^\omega y^\omega = efxefefyef = \phi_{e,ef}(x) \cdot \phi_{f,ef}(y).$$

□

Lemma 12. *A band S is normal if and only if all its local monoids are semilattices.*

Proof. If S is a normal band then for any a, b, c we have $(aba)(aca) = abaca = acaba = (aca)(aba)$ and so aSa is a semilattice.

Conversely, every band whose local monoids are semilattices is a regular band because we have

$$\begin{aligned} abca = abababcacaca &= (aba)(aba)(abca)(aca)(aca) \quad \text{(Using idempotency)} \\ &= aba(cab)(cab)aca = \qquad \text{(Since } aSa \text{ is commutative)} \\ &= abacaabaca = abaca \quad \text{(By idempotency).} \end{aligned}$$

Thus if S is a band with commutative local monoids, we have

$$abca = abaca = abaaca = acaaba = acaba = acba$$

which proves our claim. $\square$