Complexity of Unification and Matching Problems in the Varieties of Idempotent Semigroups*

Ondřej Klíma

Dept. of Mathematics, Masaryk University, Janáčkovo nám. 2a, 662 95 Brno, Czech Republic, klima@math.muni.cz

Abstract. The complexity of the unification problem in the theory of one associative, idempotent and commutative binary function symbol is known to be polynomial. On the other hand, it was proved that the unification problem is NP-complete in the theory consisting of one binary function symbol which is only associative and idempotent. We describe the complexity of the unification problems for all equational theories of a single binary function symbol consisting of associativity and idempotency together with any set of additional identities. We also describe the complexity of the matching problems for these theories.

Mathematics Subject Classification: 68Q25, 20M05

1 Introduction

The equational unification is a generalization of the syntactic (Robinson) unification in which semantic properties of function symbols are taken into account. Applications of equational unification can be found in areas as term rewriting, logic programming and automated theorem proving. For an overview of unification theory consult e.g. [4, 5].

From the algebraic point of view every equational theory corresponds to a variety of algebras of a given type and equational unification is solving of equations in the free algebra of this variety.

Unification problems for the AI-theory (i.e. the equational theory of one associative and idempotent binary function symbol) have been investigated e.g. in [1, 2, 22], however, the complexity questions were not completely answered. In this paper we extend the results from [16, 17] about complexity aspects. In contrast with [16, 17] we will assume that our function symbol satisfies some additional non-trivial identities. We will show that the unification problem for such a theory is decidable in polynomial time if the function symbol satisfies the identity $xyzx = xzyx$ and it is NP-complete otherwise. An analogous result will be proved for the matching problem and the identity $xyxzx = xyzx$.

* The paper is supported by the Grant Agency of the Czech Republic, grant No. 201/01/0323

In the language of varieties, these results can be formulated in the following way. The unification (resp. matching) problem is decidable in polynomial time for all subvarieties of the variety $\mathcal{NB}$ of all normal idempotent semigroups (resp. of the variety $\mathcal{RB}$ of all regular idempotent semigroups) and it is NP-complete for the other subvarieties of the variety of all idempotent semigroups (see Figure 4).

For applications it is also important to study the situation when monoids are considered instead of semigroups. In this case we obtain the same results about the complexity of the unification and matching problems.

The paper is organized as follows. We introduce the necessary background on idempotent semigroups in section 2. In section 3 we recall the definition of the unification and matching problems and known results. Section 4 consists of formulations and proofs of our main results.

2 Idempotent Semigroups

2.1 Basic Definitions of Invariants

We denote the set of all positive integers and the set of all non-negative integers by $\mathbb{N}$ and $\mathbb{N}_0$ respectively and the cardinality of a given set X by $\|X\|$.

Let Σ be an arbitrary finite alphabet. The free semigroup over Σ will be denoted Σ^+, the free monoid $\Sigma^* = \Sigma^+ \cup \{\epsilon\}$, where ϵ is the empty word. For a word $u \in \Sigma^*$ we call $\mathsf{c}(u)$ the set of all letters occurring in u (the *content* of u). Further $|u|$ stands for the length of the word u and $u_\sharp$ for the cardinality of $\mathsf{c}(u)$ (i.e. the number of letters occurring in the word u).

By an invariant we mean a mapping from the set Σ^* to some set Θ (in this paper the set Θ is always either Σ^* or 2^Σ). We will use the following invariants. For $u \in \Sigma^*$ we define

$\mathsf{h}(u)$ – the first letter of u (*head*),
$\mathsf{t}(u)$ – the last letter of u (*tail*),
$\overrightarrow{u}$ – the sequence of the first occurrences of all letters from $\mathsf{c}(u)$ in u,
$\overleftarrow{u}$ – the sequence of the last occurrences of all letters from $\mathsf{c}(u)$ in u,
$0(u)$ – the longest prefix of u which contains at most $u_\sharp - 1$ different letters,
$1(u)$ – the longest suffix of u which contains at most $u_\sharp - 1$ different letters,

with conventions $\mathsf{h}(\epsilon) = \mathsf{t}(\epsilon) = \overrightarrow{\epsilon} = \overleftarrow{\epsilon} = 0(\epsilon) = 1(\epsilon) = \epsilon$.

Because the invariants 0 and 1 are mappings from the set Σ^* to the same set Σ^*, we can compose them. In this way each element from the set $\{0,1\}^*$ can be viewed as an invariant (so called *iterated invariant*) defined inductively: for an arbitrary $u \in \Sigma^+$ we put $\mathsf{e}(u) = u$ for the empty invariant e (e is a neutral element in $\{0,1\}^*$) and $(0\mu)(u) = 0(\mu(u))$, $(1\mu)(u) = 1(\mu(u))$ for all iterated invariants $\mu \in \{0,1\}^*$.

An arbitrary iterated invariant $\mu \in \{0,1\}^*$ can be written in a unique way as $\mu = 0^{k_0}1^{l_1}0^{k_1}\cdots 0^{k_n}1^{l_{n+1}}$ where $n \geq 0$, $k_i, l_i \in \mathbb{N}$ for all $i \in \{1,\ldots,n\}$ and $k_0, l_{n+1} \in \mathbb{N}_0$. Then we define the *number of blocks* in μ as $\mathrm{block}(\mu) = 2n + \mathrm{sgn(k_0)} + \mathrm{sgn(l_{n+1})}$, where $\mathrm{sgn(m)} = 1$ for all $m \in \mathbb{N}$ and $\mathrm{sgn}(0) = 0$. (The blocks are the iterated invariants 0^{k_0}, 1^{l_1}, …, $1^{l_{n+1}}$.) The last invariant 0 or 1 employed during the application of the iterated invariant $\mu \in \{0,1\}^*$, i.e. the first letter of μ if we consider μ as a word over the alphabet $\{0,1\}$, is denoted by $\mathrm{last}(\mu)$.

For technical reasons we need an invariant n which is defined as $\mathsf{n}(u) = \epsilon$ for an arbitrary word $u \in \Sigma^*$.

For a word $u \in \Sigma^+$ and $n = u_\sharp$ we define $\overset{\Rightarrow}{u} \in \Sigma^+$ in this way: $\overset{\Rightarrow}{u} = a_n b_n \cdots a_1 b_1$ where $a_i, b_i \in \Sigma$, $a_n a_{n-1} \cdots a_1 = \overrightarrow{u}$ and $b_i = \mathsf{t}(0^{i-1}(u))$ for all $i \in \{1, \ldots, n\}$. Dually we define $\overset{\Leftarrow}{u}$.

Example 1. We demonstrate these definitions on the word $u = bdabdbcacbdb$ over the alphabet $\{a,b,c,d\}$. We have $\mathsf{c}(u) = \{a,b,c,d\}$, $u_\sharp = 4$, $|u| = 12$, $\mathsf{h}(u) = b$, $\mathsf{t}(u) = b$, $0(u) = bdabdb$, $1(u) = cbdb$. Further, $10(u) = 1(0(u)) = 1(bdabdb) = bdb$ and analogically $0^2(u) = bd$, $01(u) = cb$, $1^2(u) = bdb$, $10^2(u) = d$ and $010(u) = b$. Finally, $\overrightarrow{u} = bdac$ and $\overleftarrow{u} = acdb$, $\overset{\Rightarrow}{u} = bbddabcb$ and $\overset{\Leftarrow}{u} = baccbdbb$.

2.2 Varieties of Idempotent Semigroups

An idempotent semigroup (in short called a band) is a semigroup where the identity $x^2 = x$ is satisfied. The variety of all bands is denoted as $\mathcal{B}$. The lattice of all varieties of bands $\mathcal{L}(\mathcal{B})$ (see Figure 1) was described independently by Birjukov [6], Fennemore [9] and Gerhard [10] (see [11] for a revised version).

For a given variety of bands $\mathcal{V}$ and a given set Σ we denote by $\sim_{\mathcal{V},\Sigma}$ the (fully invariant) congruence on Σ^+ such that for $p, q \in \Sigma^+$ we have $p \sim_{\mathcal{V},\Sigma} q$ if and only if $p = q$ is an identity satisfied in the variety $\mathcal{V}$. (The indices $\mathcal{V}$ and Σ are omitted if they are obvious from context.)

It appears to be more useful to know a solution of the identity problem for the given variety $\mathcal{V}$, i.e. an effective description of the congruence $\sim_{\mathcal{V}}$, rather than a base of identities for $\mathcal{V}$. In this paper we will use a description due to Polák [20] using the invariants defined above. We will present this description later, now we just explain the appearance of these invariants on Figure 1. We associate to the variety of bands $\mathcal{V}$ the set of invariants $\{\iota_1, \ldots, \iota_n\}$ (as on Figure 1) if the following condition is satisfied: for two words $p, q \in \Sigma^+$ we have $p \sim_{\mathcal{V}} q$ if and only if $\iota_i(p) = \iota_i(q)$ for all $i \in \{1, \ldots, n\}$.

Example 2. For the word $u = bdabdbcacbdb$ from Example 1 we have $u \sim_{\mathcal{LSNB}} bdabcb = v$ because $\overset{\Rightarrow}{u} = \overset{\Rightarrow}{v}$. Analogically, $u \sim_{\mathcal{RB}} bdacdb = w$ because $\overrightarrow{u} = \overrightarrow{w}$ and $\overleftarrow{u} = \overleftarrow{w}$. We have also $u \sim_{\mathcal{LNB}} bacd$ because $\mathsf{c}(u) = \mathsf{c}(bacd)$ and $\mathsf{h}(u) = \mathsf{h}(bacd)$.

In this paper we will meet especially the varieties from the bottom of the lattice $\mathcal{L}(\mathcal{B})$ which were assigned certain names in literature. Here is a list of some of them with the corresponding bases of identities.

$\mathcal{SNB} = \mathrm{Mod}(x^2 = x, xyzuzyxx = xyzxzuzxzyx)$ – all seminormal bands,
$\mathcal{LSNB} = \mathrm{Mod}(x^2 = x, xyzxz = xyz)$ – all left seminormal bands,
$\mathcal{RSNB} = \mathrm{Mod}(x^2 = x, xzxyz = xyz)$ – all right seminormal bands,
$\mathcal{RB} = \mathrm{Mod}(x^2 = x, xyxzx = xyzx)$ – all regular bands,
$\mathcal{LRB} = \mathrm{Mod}(x^2 = x, xyx = xy)$ – all left regular bands,

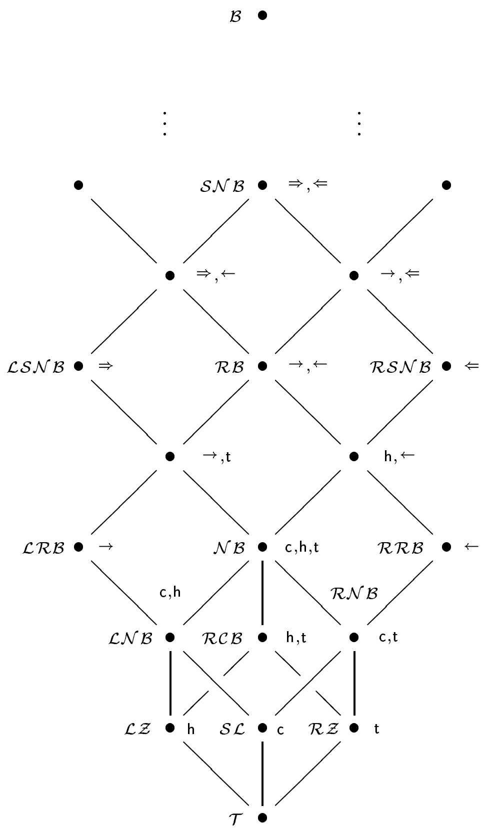

Fig. 1. The lattice of varieties of bands $\mathcal{L}(\mathcal{B})$.

$\mathcal{RRB} = \mathrm{Mod}(x^2 = x, xyx = yx)$ – all right regular bands,
$\mathcal{NB} = \mathrm{Mod}(x^2 = x, xyzx = xzyx)$ – all normal bands,
$\mathcal{LNB} = \mathrm{Mod}(x^2 = x, xyz = xzy)$ – all left normal bands,
$\mathcal{RNB} = \mathrm{Mod}(x^2 = x, xyz = yxz)$ – all right normal bands,
$\mathcal{SL} = \mathrm{Mod}(x^2 = x, xy = yx)$ – all semilattices,
$\mathcal{RCB} = \mathrm{Mod}(x^2 = x, xyz = xz)$ – all rectangular bands,
$\mathcal{LZ} = \mathrm{Mod}(xy = x)$ – all left zero semigroups,
$\mathcal{RZ} = \mathrm{Mod}(xy = y)$ – all right zero semigroups,
$\mathcal{T} = \mathrm{Mod}(x = y)$ – all trivial semigroups.

2.3 Identity Problems for Varieties of Bands

The identity problem for the variety of all bands $\mathcal{B}$ is solvable by the following well known proposition.

Proposition 1 ([12]). *Let $p, q \in \Sigma^+$. If $p_\sharp = 1$ then $p \sim_{\mathcal{B}} q$ if and only if $\mathsf{c}(p) = \mathsf{c}(q)$ and if $p_\sharp > 1$ then $p \sim_{\mathcal{B}} q$ if and only if $\mathsf{c}(p) = \mathsf{c}(q)$, $0(p) \sim_{\mathcal{B}} 0(q)$ and $1(p) \sim_{\mathcal{B}} 1(q)$.*

We can reformulate the previous proposition in the following way. For each word $u \in \Sigma^+$ we define the mapping $\mathrm{inv}_u : \{0,1\}^* \to 2^\Sigma$ by the rule $\mathrm{inv}_u(\mu) = \mathsf{c}(\mu(u))$ for all $\mu \in \{0,1\}^*$. Proposition 1 says that for $p, q \in \Sigma^+$ we have $p \sim_{\mathcal{B}} q$ if and only if the mappings inv_p and inv_q are equal.

The mapping inv_u can be also used for the construction of a short word w which is $\sim_{\mathcal{B}}$-related to the word u (more about it in the last section). This is not usual in literature because there is a natural normal form in the class of words $\sim_{\mathcal{B}}$-related to the given word u based on the term rewriting system for free bands developed in [24]. It is easy to see that this rewriting system gives also a polynomial algorithm solving the identity problem in the variety of all bands $\mathcal{B}$.

We turn our attention to the description of the identity problem for varieties of bands. First of all we define two ordered sets Λ and Ξ (see Figure 2). Ordering on the set $\Lambda = \{0,1\} \times \mathbb{N}$ is given in this way: $(\iota, n) < (\kappa, m)$ iff $n > m$ and ordering on the set of invariants $\Xi = \{\mathsf{n}, \mathsf{h}, \mathsf{t}, \mathsf{c}\}$ is given by $\mathsf{n} < \mathsf{h}, \mathsf{t}$ and $\mathsf{h}, \mathsf{t} < \mathsf{c}$.

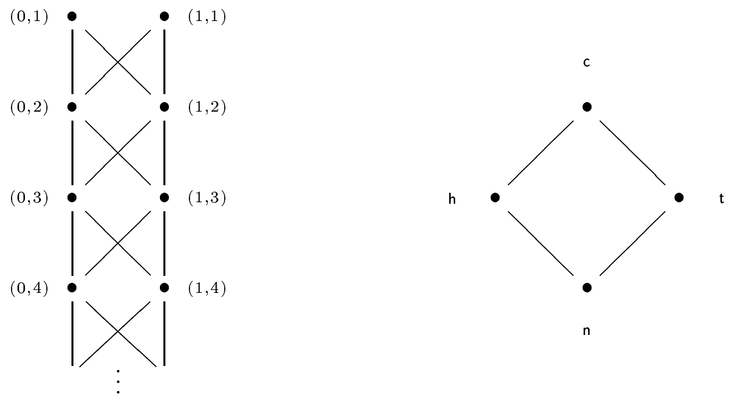

Fig. 2. The ordered sets Λ and Ξ.

Now we consider the ordered set

$$I = \{\varphi : \Lambda \to \Xi \mid \varphi \text{ is isotone}, \varphi(0,n) \neq \mathsf{t}, \varphi(1,n) \neq \mathsf{h} \text{ for } n \in \mathbb{N}\}.$$

Ordering on the set I is given in a usual way: $\varphi \leq \varphi'$ if and only if $\varphi(\iota, n) \leq \varphi'(\iota, n)$ for all $(\iota, n) \in \Lambda$.

The next proposition describes the interval in the lattice $\mathcal{L}(\mathcal{B})$ between the varieties $\mathcal{SL}$ and $\mathcal{B}$.

Proposition 2 ([20]). *$[\mathcal{SL}, \mathcal{B}] \cong I$, where every variety $\mathcal{V}$ corresponds to the mapping $\varphi_{\mathcal{V}} \in I$ such that for two words $p, q \in \Sigma^+$: $p \sim_{\mathcal{V}} q$ if and only if $\mathsf{c}(p) = \mathsf{c}(q)$ and for all iterated invariants $\mu \in \{0,1\}^+$ we have $\iota(\mu(p)) = \iota(\mu(q))$, where $\iota = \varphi_{\mathcal{V}}(\mathrm{last}(\mu), \mathrm{block}(\mu)) \in \Xi$.*

Example 3. The variety $\mathcal{LRB}$ corresponds to $\varphi \in I$ satisfying $\varphi(0,1) = \mathsf{c}$ and $\varphi(\iota, n) = \mathsf{n}$ otherwise. This means that $p \sim_{\mathcal{LRB}} q$ iff $\mathsf{c}(0^i(p)) = \mathsf{c}(0^i(q))$ for all $i \in \mathbb{N}_0$, because an iterated invariant $\mu \in \{0,1\}^+$ satisfies $\mathrm{last}(\mu) = 0$ and $\mathrm{block}(\mu) = 1$ if and only if $\mu = 0^i$ for some $i \in \mathbb{N}$. In other words $p \sim_{\mathcal{LRB}} q$ if and only if $\overrightarrow{p} = \overrightarrow{q}$.

The corresponding mapping φ for the variety $\mathcal{LSNB}$ is defined in the following way: $\varphi(0,1) = \mathsf{c}$, $\varphi(1,1) = \varphi(1,2) = \mathsf{t}$ and $\varphi(\iota, n) = \mathsf{n}$ otherwise. We have $p \sim_{\mathcal{LSNB}} q$ if and only if $\mathsf{c}(p) = \mathsf{c}(q)$, $\overrightarrow{p} = \overrightarrow{q}$ (because $\varphi(0,1) = \mathsf{c}$) and $\mathsf{t}(1^i 0^j(p)) = \mathsf{t}(1^i 0^j(q))$ for all $i \in \mathbb{N}, j \in \mathbb{N}_0$ (because $\varphi(1,1) = \varphi(1,2) = \mathsf{t}$), i.e. $p \sim_{\mathcal{LSNB}} q$ if and only if $\overset{\Rightarrow}{p} = \overset{\Rightarrow}{q}$.

Remark 1. By a proper subvariety $\mathcal{V}$ of the variety of all bands $\mathcal{B}$ we mean each subvariety of $\mathcal{B}$ except $\mathcal{B}$ itself. For each proper subvariety $\mathcal{V} \in [\mathcal{SL}, \mathcal{B}]$ there is an integer k such that $\varphi_{\mathcal{V}}(\iota, n) = \mathsf{n}$ for all $n \geq k$. For the variety of all bands $\mathcal{B}$ the corresponding mapping $\varphi_{\mathcal{B}}$ satisfies $\varphi_{\mathcal{B}}(\iota, n) = \mathsf{c}$ for all $(\iota, n) \in \Lambda$. In general, if we want to know whether $p \sim_{\mathcal{V}} q$ for two words $p, q \in \Sigma^+$, $\mathsf{c}(p) = \mathsf{c}(q)$, then it is enough to check the condition from the previous proposition only for the iterated invariants $\mu \in \{0,1\}^+$ such that $|\mu| < p_\sharp$, because $\mu(p) = \epsilon = \mu(q)$ for all μ, $|\mu| \geq p_\sharp$.

We will work in the last section with some special varieties of bands. The variety $\mathcal{V}_k$ for $k \in \mathbb{N}$ is the variety corresponding to the mapping φ_k defined by the rule $\varphi_k(\iota, n) = \mathsf{c}$ if $n < k$ and $\varphi_k(\iota, n) = \mathsf{n}$ if $n \geq k$. This means that $p \sim_{\mathcal{V}_k} q$ if and only if $\mathsf{c}(\mu(p)) = \mathsf{c}(\mu(q))$ (i.e. $\mathrm{inv}_p(\mu) = \mathrm{inv}_q(\mu)$) for all iterated invariants $\mu \in \{0,1\}^*$ such that $\mathrm{block}(\mu) < k$. So, for example $\mathcal{V}_1 = \mathcal{SL}$ and $\mathcal{V}_2 = \mathcal{RB}$.

2.4 Idempotent Monoids

For any idempotent monoid M with neutral element 1 the subset $M \setminus \{1\}$ forms a subsemigroup. Really, for $a, b \in M$ such that $a \cdot b = 1$ we have $a = a \cdot 1 = a(ab) = a^2 b = ab = 1$. This easy observation is behind the following well-known modifications of the previous results.

We denote $\mathcal{IM}$ the variety of all idempotent monoids. Other important varieties of idempotent monoids are the varieties of all regular, left regular, right regular, commutative and trivial idempotent monoids.

$\mathcal{RIM} = \mathrm{Mod}(x^2 = x, 1 \cdot x = x, x \cdot 1 = x, xyxzx = xyzx)$
$\mathcal{LRIM} = \mathrm{Mod}(x^2 = x, 1 \cdot x = 1, xyx = xy)$
$\mathcal{RRIM} = \mathrm{Mod}(x^2 = x, x \cdot 1 = x, xyx = yx)$
$\mathcal{CIM} = \mathrm{Mod}(x^2 = x, x \cdot 1 = x, xy = yx)$
$\mathcal{TM} = \mathrm{Mod}(x = 1)$

Let $\mathcal{V}$ be a variety of idempotent monoids and let Σ be a set. We denote by $\sim_{\mathcal{V},\Sigma}$ the (fully invariant) congruence on Σ^* corresponding to the variety $\mathcal{V}$. One can prove that its restriction $\sim_{\mathcal{V},\Sigma} \cap(\Sigma^+ \times \Sigma^+)$ is a fully invariant congruence on the semigroup Σ^+. (Notice that $\sim_{\mathcal{V},\Sigma} \cap(\Sigma^+ \times \Sigma^+) = \sim_{\mathcal{V},\Sigma} \setminus\{(\epsilon,\epsilon)\}$ in the case $\mathcal{V} \neq \mathcal{TM}$.) This means that $\mathcal{V}$ determines a variety of idempotent semigroups, denote it $\mathbf{S}(\mathcal{V})$, corresponding to this congruence. For example $\mathbf{S}(\mathcal{TM}) = \mathcal{T}$, $\mathbf{S}(\mathcal{CIM}) = \mathcal{SL}$, $\mathbf{S}(\mathcal{RRIM}) = \mathcal{RRB}$, $\mathbf{S}(\mathcal{LRIM}) = \mathcal{LRB}$, $\mathbf{S}(\mathcal{RIM}) = \mathcal{RB}$. In fact, $\mathbf{S}$ is an embedding of the lattice of varieties of idempotent monoids $\mathcal{L}(\mathcal{IM})$ (see Figure 3) to the lattice $\mathcal{L}(\mathcal{B})$.

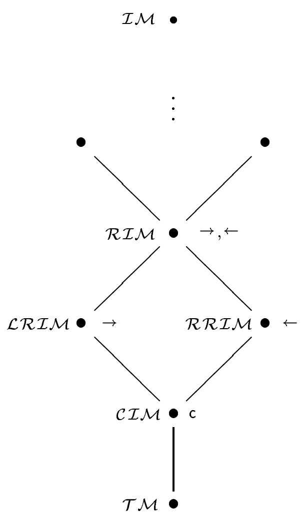

Fig. 3. The lattice of varieties of idempotent monoids $\mathcal{L}(\mathcal{IM})$.

Moreover, the mappings φ corresponding to varieties of bands of the form $\mathbf{S}(\mathcal{V})$ from the interval $[\mathcal{SL},\mathcal{B}]$ satisfy $\varphi(\iota,n) \in \{\mathsf{n},\mathsf{c}\}$ for all $(\iota,n) \in \Lambda$. So, Proposition 2 can be modified to a description of the interval $[\mathcal{CIM},\mathcal{IM}]$ in the lattice of all varieties of idempotent monoids if we consider the ordered set $\{\mathsf{n},\mathsf{c}\}$ with ordering $\mathsf{n} < \mathsf{c}$ instead of the set Ξ. The solution of the identity problem for the variety of idempotent monoids $\mathcal{V}$ is the same as for the variety of idempotent semigroups $\mathbf{S}(\mathcal{V})$; only in addition the empty word ϵ (which forms a one-element $\sim_{\mathcal{V}}$-class in the case $\mathcal{V} \neq \mathcal{TM}$) is allowed.

3 Unification Problems for Idempotent Semigroups

Let C be a finite set of *constants* and X be a finite set of *variables* such that $\mathrm{C} \cap \mathrm{X} = \emptyset$. A *word equation* $L = R$ is a pair $(L,R) \in (\mathrm{C} \cup \mathrm{X})^+ \times (\mathrm{C} \cup \mathrm{X})^+$. A *system of word equations* is a finite set of equations of the form $S = \{L_1 = R_1, \ldots, L_n = R_n\}$ for $n > 0$. A *solution* of the system S in a variety of bands

$\mathcal{V}$ is a homomorphism $\alpha : (\mathrm{C} \cup \mathrm{X})^+ \to \mathrm{C}^+$ which behaves as an identity on the letters from C and equates all the equations of the system, i.e. $\alpha(L_i) \sim_{\mathcal{V},\mathrm{C}} \alpha(R_i)$ for all $i \in \{1, \ldots, n\}$. Such a homomorphism is fully established by a mapping $\alpha|_{\mathrm{X}} \colon \mathrm{X} \to \mathrm{C}^+$. We will use the same symbol α for this mapping and its unique extension to the homomorphism $\alpha : (\mathrm{C} \cup \mathrm{X})^+ \to \mathrm{C}^+$.

Remark 2. If we want to work with idempotent monoids, we consider free monoids $(\mathrm{C} \cup \mathrm{X})^*$ and C^* instead of $(\mathrm{C} \cup \mathrm{X})^+$ and C^+. In this paper we focus on the case of semigroups, because this case is more general in a certain sense as we have seen in the previous section.

We say that a system of word equations is *satisfiable* in a variety $\mathcal{V}$ whenever it has a solution.

Remark 3. Let $\mathcal{V} \subseteq \mathcal{U}$ be two varieties of bands. If the system of word equations S is satisfiable in the variety $\mathcal{U}$ then it is also satisfiable in the variety $\mathcal{V}$, because any solution of the system S in the variety $\mathcal{U}$ is also a solution in the variety $\mathcal{V}$.

Definition 1. *Let $\mathcal{V} \subseteq \mathcal{B}$ be a variety of bands. Given a system of word equations $\{L_1 = R_1, \ldots, L_n = R_n\}$ with a set of variables* X *and a set of constants* C *as an instance of the $\mathcal{V}$-*UNIFICATION *problem, the task is to decide whether this system has a solution in the variety $\mathcal{V}$. The $\mathcal{V}$-*MATCHING *problem is the modification of the $\mathcal{V}$-*UNIFICATION *problem where only systems with $R_i \in \mathrm{C}^+$ are considered. Sometimes we speak about the* UNIFICATION *(resp.* MATCHING*) problem for the variety $\mathcal{V}$ instead of the $\mathcal{V}$-*UNIFICATION *(resp. $\mathcal{V}$-*MATCHING*) problem.*

The $\mathcal{V}$-UNIFICATION problem for any variety of bands $\mathcal{V}$ is decidable because of local finiteness of the variety $\mathcal{V}$ (consequence of Proposition 1). So we can consider two basic natural questions:

1) What is the complexity of the $\mathcal{V}$-UNIFICATION problem?

2) What does the set of all solutions look like?

We concentrate on the first question in this paper, but we recall some results in connection with the second one now. If we want to describe the set of all solutions of a system S, it is useful to assume that a solution is a mapping $\alpha : \mathrm{X} \to (\mathrm{C} \cup \mathrm{Y})^+$, where Y is an infinite set of new variables. On the set of all solutions of the system S we can define a quasiordering $\leq$ in this way: $\alpha \leq \beta$ if and only if there is a mapping $\beta : \mathrm{Y} \to (\mathrm{C} \cup \mathrm{Y})^+$ such that $\beta \circ \alpha = \beta$. A solution α is called *minimal* if for every solution β, $\beta \leq \alpha$ implies $\alpha \leq \beta$. A set M of minimal solutions is called *complete* if for every solution β there is $\alpha \in M$ such that $\alpha \leq \beta$.

We say that a variety $\mathcal{V}$ has unification type *unitary* if every satisfiable system of word equations possesses a one-element complete set of minimal solutions in $\mathcal{V}$. We say that a variety $\mathcal{V}$ has unification type *finitary* if every system of word equations possesses finite complete set of minimal solutions in $\mathcal{V}$. If a complete set of minimal solutions in $\mathcal{V}$ exists for an arbitrary system but there is no finite complete set of minimal solutions for some system, we speak about *infinitary* unification type. If a complete set of minimal solutions does not exist for some system, we speak about *nullary* unification type.

For example, unification type is unitary for the syntactic unification [21]. The UNIFICATION problem for the variety of all commutative semigroups is of finitary type [23]. An example of infinitary unification type is the UNIFICATION problem for the variety of all semigroups, because for instance the equation $xy = yx$ has minimal solutions $\alpha(x) = z^k, \alpha(y) = z^l$ where k and l are relatively prime (see [19]). The first UNIFICATION problem of nullary type was described by Fages and Huet in [8]. Then Schmidt-Schauss [22] and Baader [1] (even for $\mathrm{C} = \emptyset$) proved that the UNIFICATION problem for the variety of all bands is of nullary type. In fact, we have the following characterization of unification type for all varieties of bands.

Proposition 3 ([2]). *A variety of bands $\mathcal{V}$ has unitary unification type if $\mathcal{V} \in [\mathcal{T}, \mathcal{RCB}]$, finitary type if $\mathcal{V} \in [\mathcal{SL}, \mathcal{NB}]$ and nullary type otherwise.*

Now we recall known results about complexity of the UNIFICATION and MATCHING problems for the variety of all bands.

Proposition 4 ([16]). *The $\mathcal{B}$-MATCHING problem is NP-complete.*

Proposition 5 ([17]). *The $\mathcal{B}$-UNIFICATION problem is NP-complete.*

4 Main Results

The main results of the paper are the following two theorems which describe (in the case P≠NP) the borderline between varieties with UNIFICATION (resp. MATCHING) problems decidable in polynomial time and varieties with NP-complete UNIFICATION (resp. MATCHING) problems. This borderline is depicted on Figure 4.

Theorem 1. *Let $\mathcal{V}$ be a variety of bands. The $\mathcal{V}$-UNIFICATION problem is decidable in polynomial time if $\mathcal{V} \subseteq \mathcal{NB}$ and it is NP-complete otherwise.*

Proof. The statement is a consequence of Propositions 6, 8 and 10 below in the case of proper subvarieties of $\mathcal{B}$. For $\mathcal{B}$ it is Propositions 5. □

Theorem 2. *Let $\mathcal{V}$ be a variety of bands. The $\mathcal{V}$-MATCHING problem is decidable in polynomial time if $\mathcal{V} \subseteq \mathcal{RB}$ and it is NP-complete otherwise.*

Proof. It is a consequence of Propositions 6, 7, 11 below and Proposition 4. □

Both theorems can be formulated for the UNIFICATION and MATCHING problems for the varieties of idempotent monoids as well. It is easy to see that the following proofs work also if the empty word is allowed in the case of varieties of the form $\mathbf{S}(\mathcal{V})$ for some variety of idempotent monoids $\mathcal{V}$. The varieties of the form $\mathbf{S}(\mathcal{V})$ are noted by filled black circles on Figure 4.

Corollary 1. *Let $\mathcal{V}$ be an arbitrary variety of idempotent monoids. The $\mathcal{V}$-UNIFICATION problem is decidable in polynomial time if $\mathcal{V} \subseteq \mathcal{CIM}$ and it is NP-complete otherwise. The $\mathcal{V}$-MATCHING problem is decidable in polynomial time if $\mathcal{V} \subseteq \mathcal{RIM}$ and it is NP-complete otherwise.* □

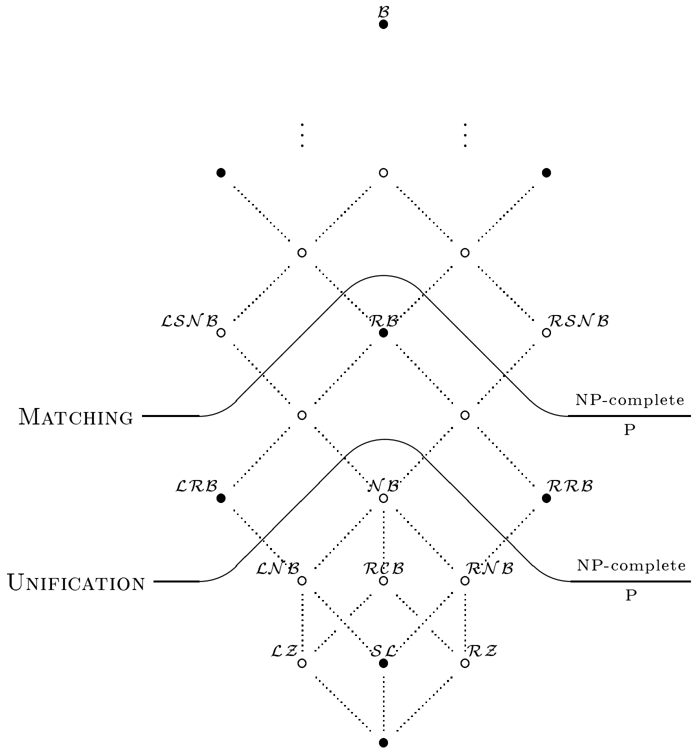

Fig. 4. The complexity of the UNIFICATION and MATCHING problems in the subvarieties of the variety of all bands $\mathcal{B}$.

4.1 NP Algorithms

Lemma 1. *Let k and n be positive integers, $n > 1$. Then the number of all iterated invariants $\mu \in \{0,1\}^+$ such that $|\mu| \leq n$ and* $\mathrm{block}(\mu) = k$ *is equal to* $2\binom{n}{k}$ *and the number of all iterated invariants $\mu \in \{0,1\}^*$ such that $|\mu| \leq n$ and* $\mathrm{block}(\mu) < k$ *is less than* $2n^k$.

Proof. Each invariant $\mu \in \{0,1\}^+$ such that $|\mu| \leq n$ and $\mathrm{block}(\mu) = k$ is determined by $\mathrm{last}(\mu)$ and by the positions of ends of blocks. Hence there are $2\binom{n}{k}$ of them. The number of all invariants $\mu \in \{0,1\}^*$ such that $|\mu| \leq n$ and $\mathrm{block}(\mu) < k$ is $K = 1 + \sum_{l=1}^{k-1} 2\binom{n}{l}$. Because $\binom{n}{l} \leq n^l$, we get $K \leq 1 + 2\sum_{l=1}^{k-1} n^l < 2\sum_{l=0}^{k-1} n^l = 2\frac{n^k-1}{n-1} < 2n^k$ for $n > 1$. □

Remark 4. From Proposition 2 and Lemma 1 we have a polynomial-time algorithm for solving the identity problem for an arbitrary proper subvariety of the variety of all bands, because we must test only polynomially many invariants for given two words (see Remark 1).

Lemma 2. *Let Σ be a finite set and $k \in \mathbb{N}$. Let K denote the number of all iterated invariants $\mu \in \{0,1\}^*$ with the properties $|\mu| < \|\Sigma\|$ and* $\mathrm{block}(\mu) < k$. *Then for an arbitrary word $u \in \Sigma^+$ there exists a word $w \in \Sigma^+$ such that $w \sim_{\mathcal{V}_k} u$ and $|w| < 2K$.*

Proof. Let T be the set of all iterated invariants $\mu \in \{0,1\}^*$ which satisfy the conditions $|\mu| < u_\sharp$ and $\mathrm{block}(\mu) < k$. Then $w \sim_{\mathcal{V}_k} u$ if and only if $\mathrm{inv}_w(\mu) = \mathrm{inv}_u(\mu)$ for all $\mu \in T$, because $\mathrm{inv}_u(\mu) = \emptyset$ for $\mu \in \{0,1\}^*$ such that $|\mu| \geq u_\sharp$. We order the set T in this way: $\mu_1 \leq \mu_2$ if and only if μ_2 is a suffix of μ_1. We can view T as a binary rooted tree with the root e where each node μ has successors 0μ and 1μ if they are in T. Now we define inductively the mapping $\omega : T \to \Sigma^+$. If μ is a leaf, i.e. $|\mu| = u_\sharp - 1$, then the word $\omega(\mu)$ is the unique letter from the set $\mathrm{inv}_u(\mu)$. If the node μ has only one successor 0μ (resp. 1μ) then $\omega(\mu) = \omega(0\mu)a$ (resp. $\omega(\mu) = a\omega(1\mu)$) where a is determined by the condition $a \in \mathrm{inv}_u(\mu) \setminus \mathrm{inv}_u(0\mu)$ (resp. $a \in \mathrm{inv}_u(\mu) \setminus \mathrm{inv}_u(1\mu)$). Finally, if the node μ has both successors 0μ and 1μ then $\omega(\mu) = \omega(0\mu)ab\omega(1\mu)$ where $a \in \mathrm{inv}_u(\mu) \setminus \mathrm{inv}_u(0\mu)$ and $b \in \mathrm{inv}_u(\mu) \setminus \mathrm{inv}_u(1\mu)$.

We prove by induction on the tree structure of T (starting from leaves) that for each $\mu \in T$

$$\mathsf{c}(\omega(\mu)) = \mathrm{inv}_u(\mu). \tag{1}$$

If $\mu \in T$ is a leaf then it is clear. Let $\mu \in T$ be an arbitrary node and assume that (1) holds for all successors of μ. We distinguish between two cases, namely the case when μ has only one successor $\iota\mu$ where $\iota \in \{0,1\}$ and the case when μ has both successors. In the first case $\mathsf{c}(\omega(\iota\mu)) = \mathrm{inv}_u(\iota\mu)$ by the assumption and from the definition of $\omega(\mu)$ we can deduce $\mathsf{c}(\omega(\mu)) = \mathsf{c}(\omega(\iota\mu)) \cup (\mathrm{inv}_u(\mu) \setminus \mathrm{inv}_u(\iota\mu)) = \mathrm{inv}_u(\iota\mu) \cup (\mathrm{inv}_u(\mu) \setminus \mathrm{inv}_u(\iota\mu)) = \mathrm{inv}_u(\mu)$. In the second case we have $\mathsf{c}(\omega(\mu)) = \mathsf{c}(\omega(0\mu)) \cup (\mathrm{inv}_u(\mu) \setminus \mathrm{inv}_u(0\mu)) \cup (\mathrm{inv}_u(\mu) \setminus \mathrm{inv}_u(1\mu)) \cup \mathsf{c}(\omega(1\mu)) = \mathrm{inv}_u(\mu)$. So, (1) holds for every $\mu \in T$. Now from (1) and the definition of $\omega(\mu)$ we can deduce $\iota(\omega(\mu)) = \omega(\iota\mu)$ for any $\mu \in T$ and $\iota \in \{0,1\}$ such that $\iota\mu \in T$. Hence $\lambda(\omega(\mu)) = \omega(\lambda\mu)$ holds for any $\mu \in T$ and $\lambda \in \{0,1\}^*$ such that $\lambda\mu \in T$. Now, for an arbitrary $\mu \in T$ we have $\mathrm{inv}_{\omega(\mathsf{e})}(\mu) = \mathsf{c}(\mu(\omega(\mathsf{e}))) = \mathsf{c}(\omega(\mu)) = \mathrm{inv}_u(\mu)$ by (1). This means $\omega(\mathsf{e}) \sim_{\mathcal{V}_k} u$.

Moreover, one can prove by induction on the tree structure of T that for each $\mu \in T$

$$|\omega(\mu)| < 2 \cdot \|\{\, \lambda \in \{0,1\}^* \mid \lambda\mu \in T\}\|.$$

Hence, for $\mu = \mathsf{e}$ we get $|\omega(\mathsf{e})| < 2 \cdot \|T\| = 2K$. ◻

Proposition 6. *The* Unification *problem is in NP for an arbitrary proper subvariety of the variety $\mathcal{B}$.*

Proof. Let $\mathcal{V}$ be a proper subvariety of the variety $\mathcal{B}$. We denote k the least integer such that $\mathcal{V} \subseteq \mathcal{V}_k$. Let S be a system over a set of variables X and an n-element set of constants C. From Lemmas 1 and 2 we know that for an arbitrary word $u \in \mathrm{C}^+$ there is a word w satisfying $|w| < 4(n-1)^k$ and $w \sim_{\mathcal{V}} u$. So, if there is a solution α of the system S, then we can assume that $|\alpha(x)| < 4n^k$ for all $x \in \mathrm{X}$. Hence we can guess a solution α of this length and then we check whether α solves all equations of the system in polynomial time by Remark 4. So, the $\mathcal{V}$-UNIFICATION problem is in NP. $\square$

4.2 NP-hardness

We will use a reduction from the NP-complete problem 3-SAT (see [18]) to prove NP-hardness of our problems. An instance of 3-SAT is a conjunction $\Phi \equiv \Phi_1 \wedge \Phi_2 \wedge \ldots \wedge \Phi_n$ of clauses and each clause Φ_i, $1 \leq i \leq n$, is of the form $l_1 \vee l_2 \vee l_3$ where l_j, $1 \leq j \leq 3$, is a literal (i.e. l_j is a variable from the set *Var*, possibly negated — we call it positive resp. negative literal). A valuation is a mapping $v : Var \to \{T, F\}$. Every valuation extends naturally to Φ and we say that Φ is satisfiable if and only if there exists a valuation v such that $v(\Phi) = T$.

Proposition 7. *Let $\mathcal{V}$ be a subvariety of the variety of all bands $\mathcal{B}$ which contains some non-regular band, i.e. $\mathcal{V} \nsubseteq \mathcal{RB}$. Then the $\mathcal{V}$-*MATCHING *problem is NP-hard.*

Proof. This proof is a modification of the original proof from [16]. It is enough to show the proposition for $\mathcal{V} \in [\mathcal{LSNB}, \mathcal{B}]$. The rest is a consequence of left-right duality.

We use a reduction from 3-SAT. For the formula Φ we construct a system of word equations S such that the system S is satisfiable in the varieties from the interval $[\mathcal{LSNB}, \mathcal{B}]$ if and only if Φ is satisfiable. By Remark 3 we must prove the following two claims: (i) if the formula Φ is satisfiable then there is a solution of the system S in the variety $\mathcal{B}$ and (ii) if the system S has a solution in the variety $\mathcal{LSNB}$ then the formula Φ is satisfiable.

The system S will consist of the following equations (2) – (6) with constants $\mathrm{C} = \{a, f, t\}$ and variables $\mathrm{X} = \{x, r_1^x, s_1^x, r_2^x, s_2^x \mid x \in Var \cup \overline{Var}\}$ where $\overline{Var} = \{\overline{z} \mid z \in Var\}$ is a disjoint copy of *Var*. We define $\widetilde{z} = z$ and $\widetilde{\neg z} = \overline{z}$ for $z \in Var$ and for every clause $\Phi_i \equiv l_1 \vee l_2 \vee l_3$ we have the equation

$$f\widetilde{l_1}\widetilde{l_2}\widetilde{l_3}f = ftf. \tag{2}$$

For each $z \in Var$ we add the equations

$$tz\overline{z}t = tft, \tag{3}$$
$$fz\overline{z}f = ftf \tag{4}$$

and finally for each $x \in Var \cup \overline{Var}$ we have the following equations:

$$r_1^x x s_1^x = fat, \quad r_1^x a = fa, \tag{5}$$
$$r_2^x x s_2^x = taf, \quad r_2^x a = ta. \tag{6}$$

The intuition behind the construction is the following. If a variable $z \in Var$ is true then $z \sim_{\mathcal{V}} t$ and if z is false then $z \sim_{\mathcal{V}} f$. The first equation ensures that at least one literal in each clause is true and the other equations imply consistency, i.e. a literal and its negation cannot be both true (false). In particular, the equation (3) means that at least one of z and $\overline{z}$ contains f. Similarly for t and (4). The last equations make sure that variables z (and also $\overline{z}$) cannot contain both f and t.

Proof of the claim (i): Suppose that Φ is satisfiable, i.e. there is a valuation v such that $v(\Phi) = T$. Then we can easily define a solution of our system in the variety of all bands $\mathcal{B}$: for all $z \in \mathit{Var}$ such that $v(z) = T$ let $\alpha(z) = t, \alpha(\overline{z}) = f$ and

$$\alpha(r_1^z) = fa, \quad \alpha(s_1^z) = t, \quad \alpha(r_1^{\overline{z}}) = f, \quad \alpha(s_1^{\overline{z}}) = at,$$
$$\alpha(r_2^z) = t, \quad \alpha(s_2^z) = af, \quad \alpha(r_2^{\overline{z}}) = ta, \quad \alpha(s_2^{\overline{z}}) = f,$$

and if $v(z) = F$ then $\alpha(z) = f, \alpha(\overline{z}) = t$ and

$$\alpha(r_1^z) = f, \quad \alpha(s_1^z) = at, \quad \alpha(r_1^{\overline{z}}) = fa, \quad \alpha(s_1^{\overline{z}}) = t,$$
$$\alpha(r_2^z) = ta, \quad \alpha(s_2^z) = f, \quad \alpha(r_2^{\overline{z}}) = t, \quad \alpha(s_2^{\overline{z}}) = af.$$

Proof of the claim (ii): Let us suppose that α is an arbitrary solution of the system S in the variety $\mathcal{LSNB}$ (recall $p \sim_{\mathcal{LSNB}} q$ if and only if $\overrightarrow{\overrightarrow{p}} = \overrightarrow{\overrightarrow{q}}$) and we find a valuation which satisfies Φ. The equation (3) implies $\mathsf{c}(\alpha(x)) \subseteq \{f, t\}$ for all $x \in \mathit{Var} \cup \overline{\mathit{Var}}$. We will conclude that $\mathsf{c}(\alpha(x)) = \{f, t\}$ is not possible. If it is the case, using the equations (5) we get that $\alpha(x)$ does not begin with the constant f. For the moment assume that $\alpha(x)$ begins with f. From (5) we obtain $f \in \mathsf{c}(\alpha(r_1^x)) \subseteq \{f, a\}$. If $\mathsf{c}(\alpha(r_1^x)) = \{f\}$ then $\mathsf{c}(0(\alpha(r_1^x x s_1^x))) = \{f, t\}$ whereas $\mathsf{c}(0(fat)) = \{f, a\}$, which is a contradiction. Otherwise we have $\mathsf{c}(\alpha(r_1^x)) = \{f, a\}$ and we get $\mathsf{t}(0(\alpha(r_1^x x s_1^x))) = f$ but $\mathsf{t}(0(fat)) = a$. By similar arguments and using the equations (6) we obtain that $\alpha(x)$ does not begin with the constant t.

So, there are just two possibilities for $\alpha(x)$, namely $\alpha(x) \sim f$ or $\alpha(x) \sim t$ for any $x \in \mathit{Var} \cup \overline{\mathit{Var}}$. By the equation (3) we have $\alpha(z) \sim f$ or $\alpha(\overline{z}) \sim f$ for all $z \in \mathit{Var}$. The equation (4) implies either $\alpha(z) \sim t$ or $\alpha(\overline{z}) \sim t$. For each clause $\Phi_i \equiv l_1 \vee l_2 \vee l_3$ the equation (2) implies that $\alpha(\widetilde{l_j}) \sim t$ for some j, $1 \leq j \leq 3$. Let us finally define the valuation v as $v(z) = T$ if $\alpha(z) \sim t$ and $v(z) = F$ if $\alpha(z) \sim f$ for each $z \in \mathit{Var}$. The valuation is consistent and $v(\Phi) = T$. □

Proposition 8. *Let $\mathcal{V}$ be a proper subvariety of the variety of all bands $\mathcal{B}$ which contains some non-normal band, i.e. $\mathcal{V} \not\subseteq \mathcal{NB}$. Then the $\mathcal{V}$-*UNIFICATION *problem is NP-hard.*

Proof. It is enough to prove the proposition only for the three varieties from the interval $[\mathcal{LRB}, \mathcal{RB}]$. The rest is a consequence of Proposition 7 and duality.

We use a reduction from 3-SAT again. For the formula Φ we construct a system S such that the formula Φ is satisfiable if and only if the system S is satisfiable in the varieties from the interval $[\mathcal{LRB}, \mathcal{RB}]$. By Remark 3 we must prove the following two claims: (i) if Φ is satisfiable then there is a solution of the system S in the variety $\mathcal{RB}$ and (ii) if the system S has a solution in the variety $\mathcal{LRB}$ then Φ is satisfiable.

The system S will consist of the following equations (7) – (13) with constants $\mathrm{C} = \{a, b, c, t\}$ and variables $\mathrm{X} = \{x, s_x \mid x \in \mathit{Var} \cup \overline{\mathit{Var}}\}$ where $\overline{\mathit{Var}} = \{\overline{z} \mid z \in \mathit{Var}\}$ is a disjoint copy of Var.

We define $\widetilde{z} = z$ and $\widetilde{\neg z} = \overline{z}$ for $z \in \mathit{Var}$ and for every clause $\Phi_i \equiv l_1 \vee l_2 \vee l_3$ we have the equation

$$ab\widetilde{l_1}\widetilde{l_2}\widetilde{l_3}ab = abtab. \tag{7}$$

For each $z \in \mathit{Var}$ we add the equations

$$zt = at, \tag{8}$$
$$\overline{z}t = bt, \tag{9}$$
$$abz\overline{z}ab = abtab, \tag{10}$$
$$bs_z c = bc, \tag{11}$$
$$as_{\overline{z}}c = ac, \tag{12}$$
$$s_z z\overline{z}s_z s_{\overline{z}} = s_{\overline{z}}\overline{z}zs_z s_{\overline{z}}. \tag{13}$$

Proof of the claim (i): Suppose that Φ is satisfiable, i.e. there is a valuation v such that $v(\Phi) = T$. For all $z \in \mathit{Var}$ such that $v(z) = T$, let $\alpha(z) = at$, $\alpha(\overline{z}) = b$, $\alpha(s_z) = cb$, $\alpha(s_{\overline{z}}) = c$ and if $v(z) = F$ then $\alpha(z) = a$, $\alpha(\overline{z}) = bt$, $\alpha(s_z) = c$, $\alpha(s_{\overline{z}}) = ca$. Checking that α is a solution of S in the variety $\mathcal{RB}$ is a routine.

Proof of the claim (ii): Let α be a solution of the system S in the variety $\mathcal{LRB}$ (recall $p \sim_{\mathcal{LRB}} q$ if and only if $\overrightarrow{p} = \overrightarrow{q}$). From (8) and (9) we have $a \in \mathsf{c}(\alpha(z)) \subseteq \{a, t\}$, $b \in \mathsf{c}(\alpha(\overline{z})) \subseteq \{b, t\}$ and from (10) $t \in \mathsf{c}(\alpha(z\overline{z}))$ for each $z \in \mathit{Var}$. The equations $(11 - 13)$ serve for elimination of the case $t \in \mathsf{c}(\alpha(z)) \cap \mathsf{c}(\alpha(\overline{z}))$.

Assume for the moment that $t \in \mathsf{c}(\alpha(z)) \cap \mathsf{c}(\alpha(\overline{z}))$. Then $\mathsf{c}(\alpha(z)) = \{a, t\}$ and $\mathsf{c}(\alpha(\overline{z})) = \{b, t\}$. But (8) and (9) imply $\alpha(z) \sim at$, $\alpha(\overline{z}) \sim bt$, hence the equation (13) can be rephrased as $s_z atbs_z s_{\overline{z}} = s_{\overline{z}}btas_z s_{\overline{z}}$. From (11) and (12) we have $\mathsf{c}(\alpha(s_z)) \subseteq \{b, c\}$ and $\mathsf{c}(\alpha(s_{\overline{z}})) \subseteq \{a, c\}$, so $\alpha(s_z) \sim c$ or $\alpha(s_z) \sim cb$ and $\alpha(s_{\overline{z}}) \sim c$ or $\alpha(s_{\overline{z}}) \sim ca$ because $\mathsf{h}(\alpha(s_z)) = \mathsf{h}(\alpha(s_{\overline{z}}))$ by (13). Hence $\alpha(s_z atbs_z s_{\overline{z}}) \sim catb$ or $\alpha(s_z atbs_z s_{\overline{z}}) \sim cbat$ and $\alpha(s_{\overline{z}}btas_z s_{\overline{z}}) \sim cbta$ or $\alpha(s_{\overline{z}}btas_z s_{\overline{z}}) \sim cabt$ and this is a contradiction.

Now we see that there are two possibilities for each variable $z \in \mathit{Var}$, namely $\alpha(z) \sim at$, $\alpha(\overline{z}) \sim b$ or $\alpha(z) \sim a$, $\alpha(\overline{z}) \sim bt$. In the first case we define $v(z) = T$ and in the second case we define $v(z) = F$. From (7) we obtain $v(\Phi) = T$. $\square$

4.3 Polynomial Algorithms

Proposition 9 ([15]). *The $\mathcal{SL}$-*UNIFICATION *problem is decidable in polynomial time.*

We present the original proof from [15] because we use it in the proof of Proposition 10.

Proof. For the system $S = \{L_1 = R_1, \ldots, L_n = R_n\}$ we construct a set $\mathcal{H}$ of propositional Horn clauses such that $\mathcal{H}$ is satisfiable if and only if S is satisfiable. (See [13] for an overview on Horn clauses.)

For every variable $x \in \mathrm{X}$ and every constant $a \in \mathrm{C}$ we form the propositional variable $P_{x,a}$. The intuition behind the construction is the following: the variable $P_{x,a}$ is true if and only if $a \notin \mathsf{c}(\alpha(x))$.

For each variable x we have the Horn clause

$$\bigwedge_{a \in \mathrm{C}} P_{x,a} \implies F. \tag{14}$$

For every equation $L_i = R_i \in S$ and every constant $a \in (\mathrm{C} \cap \mathsf{c}(R_i)) \setminus \mathsf{c}(L_i)$ we add the clause

$$\bigwedge_{x \in \mathrm{X} \cap \mathsf{c}(L_i)} P_{x,a} \implies F. \tag{15}$$

Analogically, we add the clause

$$\bigwedge_{x \in \mathrm{X} \cap \mathsf{c}(R_i)} P_{x,a} \implies F \tag{16}$$

for $a \in (\mathrm{C} \cap \mathsf{c}(L_i)) \setminus \mathsf{c}(R_i)$.

Finally, for $L_i = R_i \in S$ and $a \in \mathrm{C} \setminus (\mathsf{c}(L_i) \cup \mathsf{c}(R_i))$ we add the clauses

$$\bigwedge_{x \in \mathrm{X} \cap \mathsf{c}(L_i)} P_{x,a} \implies P_{y,a} \quad \text{for all} \quad y \in \mathrm{X} \cap \mathsf{c}(R_i), \tag{17}$$

$$\bigwedge_{x \in \mathrm{X} \cap \mathsf{c}(R_i)} P_{x,a} \implies P_{y,a} \quad \text{for all} \quad y \in \mathrm{X} \cap \mathsf{c}(L_i). \tag{18}$$

Note that by $\bigwedge_{x \in \emptyset} P_{x,a}$ we mean T.

It is easy to see that if there is a valuation v which satisfies $\mathcal{H}$ then we can define a solution α of the system S in the following way: $\alpha(x)$ is a product of all constants $a \in \mathrm{C}$ such that $v(P_{x,a}) = F$. The clause (14) ensures that $\alpha(x)$ is not the empty word and the clauses (15) – (18) ensure that α is a solution.

Since satisfiability of a set of propositional Horn clauses can be checked in linear time [7] and the size of the set $\mathcal{H}$ is polynomial in the size of the original system S, the proof is finished. □

Proposition 10. *The $\mathcal{V}$-*Unification *problem is decidable in polynomial time for each variety $\mathcal{V} \subseteq \mathcal{NB}$.*

Proof. For $\mathcal{LZ}$ (and dually for $\mathcal{RZ}$) it is trivial: we can assume that the system of equations is of the form $\{L_1 = R_1, \ldots, L_n = R_n\}$ with $L_i, R_i \in \mathrm{C} \cup \mathrm{X}$, so it gives a partition of the set $\mathrm{C} \cup \mathrm{X}$. The system is satisfiable if and only if each class of the partition contains at most one constant — if a class M contains a constant a then $\alpha(x) \sim_{\mathcal{LZ}} a$ for all $x \in M \cap \mathrm{X}$ and any solution α. The system is satisfiable in $\mathcal{RCB}$ if and only if it is satisfiable in both $\mathcal{LZ}$ and $\mathcal{RZ}$.

For the $\mathcal{SL}$-Unification problem we use Proposition 9. Moreover, we can slightly modify the proof of Proposition 9 for the varieties $\mathcal{LNB}$ ($\mathcal{RNB}$, $\mathcal{NB}$

respectively). First we test whether the given system is satisfiable in $\mathcal{LZ}$ ($\mathcal{RZ}$, $\mathcal{RCB}$ respectively). If it is the case then for any $a \in \mathrm{C}$ and $x \in \mathrm{X}$ such that $a = \mathsf{h}(\alpha(x))$ or $a = \mathsf{t}(\alpha(x))$ holds for an arbitrary solution α of the system in $\mathcal{LZ}$ ($\mathcal{RZ}$, $\mathcal{RCB}$ respectively), we add the Horn clause $P_{x,a} \implies F$ to the set $\mathcal{H}$ constructed in the proof of Proposition 9. □

The last part of the proof of the main results is about the MATCHING problem for the variety $\mathcal{LRB}$. For technical reasons we will work with systems in the following special form.

Definition 2. *A system of word equations $S = \{L_1 = R_1, \ldots, L_n = R_n\}$ is called a* pattern equation system *if $L_i \in \mathrm{X}^+$ and $R_i \in \mathrm{C}^+$ for all $1 \leq i \leq n$.*

We can restrict ourselves to pattern equation systems since for every instance of the MATCHING problem it is easy to construct an equivalent pattern equation system: just notice that an equation of the form $L_1 a L_2 = A$, where $a \in \mathrm{C}$, can be transformed into equations $L_1 x L_2 = A$ and $x = a$, where x is a new variable.

First of all, we describe the solutions of a pattern equation system in the variety $\mathcal{SL}$. In fact, the free semilattice over the set C is the set of all non-empty subsets of C with the operation of union; we denote it $\mathcal{P}(\mathrm{C})$. So, a solution α of a system can be viewed as a solution in sets, i.e. $\alpha : \mathrm{X} \to \mathcal{P}(\mathrm{C})$. For two such mappings $\alpha, \beta : \mathrm{X} \to \mathcal{P}(\mathrm{C})$ we say that α is contained in β if and only if $\alpha(x) \subseteq \beta(x)$ for all $x \in \mathrm{X}$. The following fact about the maximum solution of a pattern equation system in sets is obvious.

Lemma 3. *The pattern equation system $S = \{L_1 = R_1, \ldots, L_n = R_n\}$ is satisfiable in the variety $\mathcal{SL}$ if and only if the mapping $\gamma : \mathrm{X} \to \mathcal{P}(\mathrm{C}) \cup \{\emptyset\}$ defined by $\gamma(x) = \bigcap_{x \in \mathsf{c}(L_i)} \mathsf{c}(R_i)$ for every $x \in \mathrm{X}$ is a solution of S in sets. Moreover, any solution of the system S in sets is contained in γ.* □

Remark 5. If we consider three mappings $\alpha, \beta, \gamma : \mathrm{X} \to \mathcal{P}(\mathrm{C})$ such that α is contained in β and β is contained in γ and α, γ are solutions of a pattern equation system S in sets, then β is also a solution of the system S in sets.

Now we are going to solve pattern equation systems in the variety $\mathcal{LRB}$ (recall $p \sim_{\mathcal{LRB}} q$ if and only if $\overrightarrow{p} = \overrightarrow{q}$). One can view a solution α of a pattern equation system as a solution in sets with orderings on all sets $\alpha(x)$ as follows.

For a word $u \in \Sigma^+$ and two distinct letters $a, b \in \mathsf{c}(u)$ we write $a <_u b$ if and only if the first occurrence of the letter a in u precedes all occurrences of the letter b in u. Note that $<_u$ is a strict linear ordering on the set $\mathsf{c}(u)$ and $u \sim_{\mathcal{LRB}} v$ if and only if $<_u$ and $<_v$ are the same relations on the set $\mathsf{c}(u) = \mathsf{c}(v)$.

For a mapping $\beta : \mathrm{X} \to \mathcal{P}(\mathrm{C})$, a letter $a \in \mathrm{C}$ and a word $L \in \mathrm{X}^+$ we denote by $f_a(L, \beta)$ the first letter x in L such that $a \in \beta(x)$ if such a letter exists.

Let $S = \{L_1 = R_1, \ldots, L_n = R_n\}$ be a given pattern equation system. For a mapping $\beta : \mathrm{X} \to \mathcal{P}(\mathrm{C})$ we define a binary relation $\prec_x^\beta \subseteq \beta(x) \times \beta(x)$ for all $x \in \mathrm{X}$ in this way: $a \prec_x^\beta b$ if and only if there is $i \in \{1, \ldots, n\}$ such that $f_a(L_i, \beta) = f_b(L_i, \beta) = x$ and $a <_{R_i} b$.

A mapping $\beta : \mathrm{X} \to \mathcal{P}(\mathrm{C})$ is called a *left-compatible* solution of the system S in sets if and only if the following conditions are satisfied:

(i) β is a solution in sets,
(ii) the transitive closure of $\prec_x^\beta$ is an irreflexive relation for all $x \in \mathrm{X}$,
(iii) $\forall i \in \{1, \ldots, n\}$, $\forall a, b \in \mathsf{c}(R_i) : f_a(L_i, \beta) <_{L_i} f_b(L_i, \beta) \implies a <_{R_i} b$.

Lemma 4. *Let $S = \{L_1 = R_1, \ldots, L_n = R_n\}$ be a pattern equation system. Then S is satisfiable in the variety $\mathcal{LRB}$ if and only if there exists a left-compatible solution of S in sets.*

Proof. Let α be a solution of the system S in the variety $\mathcal{LRB}$. We put $\beta(x) = \mathsf{c}(\alpha(x))$. Such β is a left-compatible solution of S in sets. Really, (i) is trivial, (ii) follows from the observation $a \prec_x^\beta b \implies a <_{\alpha(x)} b$ and (iii) is clear because $f_a(L_i, \beta) <_{L_i} f_b(L_i, \beta)$ implies $a <_{\alpha(L_i)} b$.

Conversely, let β be a left-compatible solution of the system in sets. There is a linear ordering on $\beta(x)$ containing $\prec_x^\beta$ for each $x \in \mathrm{X}$, by condition (ii). We denote it $\ll_x$ and define $\alpha(x) = c_1 c_2 \cdots c_m$, where $\{c_1, c_2, \ldots, c_m\} = \beta(x)$ and $c_1 \ll_x c_2 \ll_x \cdots \ll_x c_m$. Such α is a solution of S in the variety $\mathcal{LRB}$. Really, assume for a moment that $a <_{\alpha(L_i)} b$ and $b <_{R_i} a$ for some $i \in \{1, \ldots, n\}$ and $a, b \in \mathrm{C}$. Then $f_a(L_i, \beta) = f_b(L_i, \beta)$ because $f_a(L_i, \beta) <_{L_i} f_b(L_i, \beta)$ implies $a <_{R_i} b$ by condition (iii) and $f_b(L_i, \beta) <_{L_i} f_a(L_i, \beta)$ is a contradiction with $a <_{\alpha(L_i)} b$. If we denote $x = f_a(L_i, \beta) = f_b(L_i, \beta)$ then we have $b \prec_x^\beta a$ and this implies $b <_{\alpha(x)} a$ which is a contradiction with $a <_{\alpha(L_i)} b$. □

Lemma 5. *It is decidable in polynomial time whether a given pattern equation system $S = \{L_1 = R_1, \ldots, L_n = R_n\}$ has a left-compatible solution in sets.*

Proof. We will exhibit a polynomial algorithm which finds a left-compatible solution of S in sets if it exists. First of all, the algorithm computes (in polynomial time) the maximum solution γ of S in sets by Lemma 3 (if it exists). We obtain the statement if we apply the following claim to this solution γ of S in sets. (Any left-compatible solution of S in sets is a solution in sets, so it is contained in the maximum solution of S in sets.)
Claim. Let γ be a solution of S in sets. It is decidable in polynomial time whether there exists a left-compatible solution in sets which is contained in γ. Moreover, if such a solution exists, then there is the maximum left-compatible solution of S in sets which is contained in γ and we can compute it in polynomial time.
Proof of Claim. We construct a polynomial algorithm which finds the maximum left-compatible solution of S in sets contained in γ if it exists. At the beginning let $\delta = \gamma$. In each step the algorithm improves the mapping δ until it becomes a left-compatible solution.

Assume that $\delta : \mathrm{X} \to \mathcal{P}(\mathrm{C}) \cup \{\emptyset\}$ is not a left-compatible solution of S in sets. If (i) is not satisfied for δ, then the algorithm answers that S has no left-compatible solution contained in γ. If (ii) is not satisfied, i.e. there is $x \in \mathrm{X}$ and $a_1 \prec_x^\delta a_2 \prec_x^\delta \ldots \prec_x^\delta a_m \prec_x^\delta a_1$, then the algorithm replaces δ with the mapping δ' such that $\delta'(z) = \delta(z)$ for all variables $z \neq x$ and $\delta'(x) = \delta(x) \setminus \{a_1, \ldots, a_n\}$. If (iii) is not satisfied, i.e. $x = f_a(L_i, \delta) <_{L_i} f_b(L_i, \delta)$, $b <_{R_i} a$ for some $i \in \{1, \ldots, n\}$ and $a, b \in \mathrm{C}$, then the algorithm replaces δ with the mapping δ' such that $\delta'(z) = \delta(z)$ for all variables $z \neq x$ and $\delta'(x) = \delta(x) \setminus \{a\}$.

We must prove that our algorithm stops in polynomial time and that it returns the correct answer. In each step the algorithm computes all relations $\prec_x^\delta$ in polynomial time and it checks the conditions also in polynomial time. Then it reduces the set $\delta(x)$. Because $\delta(x)$ has lower bound $\emptyset$, the algorithm changes each $\delta(x)$ only polynomially many times. So, the algorithm stops in polynomial time.

We show by induction that an arbitrary left-compatible solution β of S in sets which is contained in γ is also contained in every mapping constructed by the algorithm. This is clearly enough to verify the correctness of the algorithm.

Let β be contained in δ which is considered in the algorithm. It is easy to see that δ satisfies (i) by Remark 5. We show that β is contained in δ'. In general, from the fact $\beta(x) \subseteq \delta(x)$ for all $x \in \mathrm{X}$ we have $f_a(L, \delta) = f_a(L, \beta)$ when $a \in \beta(f_a(L, \delta))$ and $f_a(L, \delta) <_L f_a(L, \beta)$ when $a \notin \beta(f_a(L, \delta))$ for all $L \in \mathrm{X}^+$ and $a \in \mathrm{C}$.

Assume that δ does not satisfy (ii), i.e. there is $x \in \mathrm{X}$ and $a_1 \prec_x^\delta a_2 \prec_x^\delta \dots \prec_x^\delta a_m \prec_x^\delta a_{m+1} = a_1$. Assume for a moment that there is $j \in \{1, \dots, m\}$ such that $a_j \notin \beta(x)$ and $a_{j+1} \in \beta(x)$. From the fact $a_j \prec_x^\delta a_{j+1}$ we can deduce the existence of $i \in \{1, \dots, n\}$ satisfying $f_{a_j}(L_i, \delta) = f_{a_{j+1}}(L_i, \delta) = x$ and $a_j <_{R_i} a_{j+1}$. Now $f_{a_{j+1}}(L_i, \beta) = f_{a_{j+1}}(L_i, \delta) = f_{a_j}(L_i, \delta) <_{L_i} f_{a_j}(L_i, \beta)$ and this is a contradiction with condition (iii) for β. So, there are two possibilities $\{a_1, \dots, a_m\} \subseteq \beta(x)$ or $\{a_1, \dots, a_m\} \cap \beta(x) = \emptyset$. The first one implies $a_1 \prec_x^\beta a_2 \prec_x^\beta \dots \prec_x^\beta a_m \prec_x^\beta a_1$ and this is a contradiction with condition (ii) for β. The second one means that β is contained in δ'.

Finally, assume that δ does not satisfy (iii), i.e. $x = f_a(L_i, \delta) <_{L_i} f_b(L_i, \delta)$, $b <_{R_i} a$ for some $i \in \{1, \dots, n\}$ and $a, b \in \mathrm{C}$. Assume for a moment that $a \in \beta(x)$. It is easy to see that $f_a(L_i, \beta) = f_a(L_i, \delta) <_{L_i} f_b(L_i, \beta)$. Hence β does not satisfy condition (iii) and this is a contradiction. So, $a \notin \beta(x)$, i.e. β is contained in δ'. $\square$

Proposition 11. *The $\mathcal{V}$-*Matching *problem is decidable in polynomial time for each variety $\mathcal{V} \subseteq \mathcal{RB}$.*

Proof. Let $S = \{L_1 = R_1, \dots, L_n = R_n\}$ be a pattern equation system. The case $\mathcal{V} \subseteq \mathcal{NB}$ is solved in Proposition 10. Lemmas 4 and 5 verify the case $\mathcal{V} = \mathcal{LRB}$.

For $\mathcal{V} = \mathcal{LRB} \vee \mathcal{NB} = \mathcal{LRB} \vee \mathcal{RNB}$ we test whether the system S is satisfiable in $\mathcal{RNB}$. Then we compute the maximum left-compatible solution δ of S in sets by Lemma 5 and we check whether $\mathsf{t}(R_i) \in \delta(\mathsf{t}(L_i))$ for all $i \in \{1, \dots, n\}$. This is sufficient because $\mathsf{t}(R_i) = \mathsf{t}(R_j)$ whenever $\mathsf{t}(L_i) = \mathsf{t}(L_j)$ thanks to the solvability in $\mathcal{RNB}$.

The cases $\mathcal{V} = \mathcal{RRB}$ and $\mathcal{V} = \mathcal{RRB} \vee \mathcal{NB}$ are dual to $\mathcal{LRB}$ and $\mathcal{LRB} \vee \mathcal{NB}$ respectively; we can consider the notion of a right-compatible solution of a system in sets.

For $\mathcal{V} = \mathcal{RB}$ we can consider the notion of a compatible solution of a system in sets (i.e. solution in sets which is left-compatible and at the same time right-compatible). We construct the algorithm which alternates the two algorithms for $\mathcal{LRB}$ and $\mathcal{RRB}$ in such a way that the left-compatible solution δ from the algorithm for $\mathcal{LRB}$ is the first mapping δ in the algorithm for $\mathcal{RRB}$ and vice-versa. So, the algorithm finds a solution of a solvable system in sets which is left-compatible and at the same time right-compatible. The correctness follows from the claim in the proof of Lemma 5. □

5 Final Remarks

In this paper we have characterized the complexity of the unification problems for all varieties of bands. In unification theory our $\mathcal{V}$-UNIFICATION problem is called the unification problem with constants. Another problems which are considered in unification theory are the elementary unification problem (where only variables occur in the equations) and the general unification problem (where arbitrary uninterpreted function symbols are allowed). The elementary unification problem is uninteresting in the case of idempotent semigroups, because any system is trivially solvable. Kapur and Narendran proved [14] that the general unification problem for $\mathcal{SL}$ is NP-hard. The proof of this result can be modified for any variety $\mathcal{V} \in [\mathcal{SL}, \mathcal{B}]$ and if we use the great result of Baader and Schultz [3] we can conclude that the general unification problem is NP-complete for all varieties $\mathcal{V} \in [\mathcal{SL}, \mathcal{B}]$. On the other hand, it is easy to see that the general unification problem for $\mathcal{LZ}$ (and dually for $\mathcal{RZ}$) is decidable in polynomial time. The same holds for the general unification problem for $\mathcal{RCB}$ but the proof of this fact is not so easy as in the case of $\mathcal{LZ}$.

Acknowledgments. I would like to thank Michal Kunc for his help with preparation of the paper.

References

[1] Baader F.: The Theory of Idempotent Semigroups is of Unification Type Zero, J. Automated Reasoning **2** (1986) 283–286.
[2] Baader F.: Unification in Varieties of Idempotent Semigroups, Semigroup Forum **36** (1987) 127–145.
[3] Baader F., Schulz K. U.: Unification in the Union of Disjoint Equational Theories: Combining Decision Procedures, J. Symbolic Computation **21** (1996) 211–243.
[4] Baader F., Siekmann J. H.: Unification Theory, Handbook of Logic in Artificial Intelligence and Logic Programming, Vol. 2 (1993), Oxford University Press.
[5] Baader F., Snyder W.: Unification Theory, Handbook of Automated Reasoning (2001), Elsevier Science Publishers.
[6] Birjukov A. P.: Varieties of Idempotent Semigroups, Algebra i Logika **9** (1970) 255–273 (Russian).
[7] Dowling W. F., Gallier J. H.: Linear-time Algorithms for Testing the satisfiability of Propositional Horn Formulae, J. Logic Program. **1** (1984) 267–284.
[8] Fages F., Huet G.: Complete Sets of Unifiers and Matchers in Equational Theories, In Proceedings of the 5th Colloquium on Automata, Algebra and Programming, LNCS 159 (1983) 205–220.
[9] Fennemore C. F.: All Varieties of Bands I, II, Math. Nachr. **48** (1971) 237–252, 253–262.

[10] Gerhard J. A.: The Lattice of Equational Classes of Idempotent Semigroups, J. Algebra **15** (1970) 195–224.

[11] Gerhard J. A., Petrich M.: Varieties of Bands Revisited, Proc. Lond. Math. Soc. **58** (1989) 323–350.

[12] Green J. A., Rees D.: On Semigroups in which $x^r = x$, Proc. Camb. Phil. Soc. **48** (1952) 35–40.

[13] Hodges W.: Logical Features of Horn Clauses, Handbook of Logic in Artificial Intelligence and Logic Programming, Vol. 1 (1993), Oxford University Press.

[14] Kapur D., Narendran P.: NP–completeness of the Set Unification and Matching Problems, In Proceedings of CADE'86, LNCS 230 (1986) 489–495.

[15] Kapur D., Narendran P.: Complexity of Unification Problems with Associative-commutative Operators, J. Automated Reasoning **9**, No.2, (1992) 261–288.

[16] Klíma O., Srba J.: Matching Modulo Associativity and Idempotency is NP-complete, In Proceedings of MFCS'00, LNCS 1893 (2000) 456–466.

[17] Klíma O.: Unification Modulo Associativity and Idempotency is NP-complete, http://www.math.muni.cz/∼klima/.

[18] Papadimitriou, C. H.: Computational Complexity (1994), Addison-Wesley Publishing Company.

[19] Plotkin G.: Building in Equational Theories, Machine Intelligence **7** (1972) 73–90.

[20] Polák L.: On Varieties of Completely Regular Semigroups I,II,III, Semigroup Forum **32, 36, 37** (1985,1987,1988) 97–123, 253–284, 1–30.

[21] Robinson J. A.: A Machine-oriented Logic Based on the Resolution Principle, J. Assoc. Comput. Mach. **12**, (1965) 23–41.

[22] Schmidt-Schauss M.: Unification under Associativity and Idempotence is of Type Nullary, J. of Automated Reasoning **2** (1986) 277–281.

[23] Siekmann J.: Unification of Commutative Terms, In Proceedings of EUROSAM '79, LNCS 72 (1979) 531–545.

[24] Siekmann J., Szabó P.: A Noetherian and Confluent Rewrite System for Idempotent Semigroups, Semigroup Forum **25** (1982) 83–110.