AAECC 7, 309–337 (1996)

AAECC
Applicable Algebra in
Engineering, Communication
and Computing

Combination Problems for Commutative/Monoidal Theories or How Algebra Can Help in Equational Unification

Franz Baader[1], **Werner Nutt**[2]

[1] Lehr- und Forschungsgebiet Theoretische Informatik, RWTH Aachen, Ahornstraße 55, D-52074 Aachen, Germany, e-mail: baader@informatik.rwth-aachen.de
[2] German Research Center for Artificial Intelligence (DFKI), Stuhlsatzenhausweg 3, D-66123 Saarbrücken, Germany, e-mail: nutt@dfki.uni-sb.de

Received August 20, 1992; revised version June 13, 1994

Abstract. We study the class of theories for which solving unification problems is equivalent to solving systems of linear equations over a semiring. It encompasses important examples like the theories of Abelian monoids, idempotent Abelian monoids, and Abelian groups. This class has been introduced by the authors independently of each other as "commutative theories" (Baader) and "monoidal theories" (Nutt).

We show that commutative theories and monoidal theories indeed define the same class (modulo a translation of the signature), and we prove that it is undecidable whether a given theory belongs to it. In the remainder of the paper we investigate combinations of commutative/monoidal theories with other theories. We show that finitary commutative/monoidal theories always satisfy the requirements for applying general methods developed for the combination of unification algorithms for disjoint equational theories.

Then we study the adjunction of monoids of homomorphisms to commutative/monoidal theories. This is a special case of a non-disjoint combination, which has an algebraic counterpart in the corresponding semiring. By studying equations over this semiring, we identify a large subclass of commutative/monoidal theories that are of unification type zero. We also show with methods from linear algebra that unitary and finitary commutative/monoidal theories do not change their unification type when they are augmented by a finite monoid of homomorphisms, and how algorithms for the extended theory can be obtained from algorithms for the basic theory.

Keywords: Equational theories, Equational unification, Semirings, Semiadditive categories.

1 Introduction

Equational unification is concerned with solving term equations modulo an equational theory. The theory is called *unitary* (*finitary*) if the solutions of an equation can always be represented by one (finitely many) "most general" solutions. Otherwise the theory is of type *infinitary* or *zero*. Equational theories that are of unification type unitary or finitary play an important rôle in automated theorem provers with built in theories [31, 24, 34, 35], in generalizations of the Knuth-Bendix algorithm [17, 30, 19, 7], and in logic programming with equality [18, 13].

For this reason, determining unification types of equational theories is not only interesting for unification theory but has also consequences for automated reasoning. Of course, for practical applications it is not enough to know that a given theory $\mathscr{E}$ is of type finitary. One also needs a finite $\mathscr{E}$-unification algorithm that computes the finitely many most general solutions. Unfortunately, but not at all surprisingly, there cannot be a general method that determines the unification type of an equational theory [27]; and even if a theory is finitary it is still not clear whether a unification algorithm exists. Consequently, general methods [14] that try to derive such an algorithm from a given set of axioms for the theory are doomed to fail. Without restrictions on the equational theories one obtains procedures that only enumerate complete sets of unifiers, but do not terminate, even if the theory is unitary or finitary.

Nevertheless, it is desirable to have methods – or at least methodologies – for designing unification algorithms for larger classes of theories. Otherwise, whenever a new equational theory comes up in an application one must start completely from scratch when developing a unification algorithm for this theory (we shall come back to this point in Sect. 2). One solution proposed for this problem is to restrict the attention to certain classes of theories that are defined by syntactic properties of the set of axioms (see *e.g.*, [11, 20, 9]). An advantage of such syntactic approaches is that they apply directly to general $\mathscr{E}$-unification problems, *i.e.*, problems where the terms to be unified contain additional free function symbols of arbitrary arity. A disadvantage is that they need not yield a unification algorithm for the theory in question, even though such an algorithm exists. For example, associativity and commutativity of a binary symbol is a syntactic theory (in the sense of [20]), but the general method for unification in syntactic theories usually does not terminate for unification problems modulo this theory.

The syntactic approaches mostly depend on transformations of terms; they usually do not take the properties of the algebras defined by the theory into account. On the other hand, special purpose algorithms designed for theories of practical importance – such as the theory of Abelian monoids (AM), idempotent Abelian monoids (AIM), and Abelian groups (AG) – often depend on algebraic properties of these theories. It turns out that the algebraic methods used for obtaining these unification algorithms can be generalized to larger classes of theories defined by properties of their free algebras. The theories AM, AIM, and AG belong to the class of commutative theories – roughly speaking, theories where the finitely generated free algebras are direct products of the free algebras in one generator [1, 2, 3]. One result shown in the present paper (see Sect. 4 below) is that the class of commutative theories is – modulo a translation of the signature – the same as the class of monoidal theories, developed independently in [25, 28].

Unification in these theories can always be reduced to solving linear equations in certain semirings [25]. On the one hand, this fact can be used to derive general results on unification in commutative/monoidal theories. For example, it can be shown that constant free unification problems are either unitary or of type zero, and the unification type of a theory can be characterized by algebraic properties of the corresponding semiring. These characterizations were used in [25, 2, 28] to determine the unification types of several commutative/monoidal theories. On the other hand, unification algorithms for certain commutative/monoidal theories – for example, the theory of Abelian groups with n commuting homomorphisms – can be derived with the help of well-known algebraic methods for the corresponding semiring – for instance, Buchberger's algorithm for the ring $\mathbf{Z}[X_1, \ldots, X_n]$ of integer polynomials in n indeterminates [3].

An apparent disadvantage of this semantic approach to unification is that it can treat only unification with and without constants, but not general unification. This problem can be solved, however, by using general methods developed for unification in the union of "disjoint equational theories" (*i.e.*, theories over disjoint signatures) [33, 5]. Going from $\mathscr{E}$-unification with constants to general $\mathscr{E}$-unification is an instance of this *combination problem* since it can be seen as the disjoint combination of $\mathscr{E}$ with a free theory. In order to apply the method of [5] to the combination of a commutative/monoidal theory $\mathscr{E}$ with a free theory, one must be able to solve so-called "unification problems with constant restrictions" in $\mathscr{E}$. In Sect. 6, we shall show that a commutative/monoidal theory is finitary for unification with constants if, and only if, it is finitary for unification with constant restrictions. Consequently, the results in [5] imply that a commutative/monoidal theory is finitary for unification with constants if, and only if, it is finitary for general unification.

For disjoint theories, the combination problem can be considered as by and large solved by the work of [33, 5]. The case of non-disjoint signatures is too difficult to be treated in its full generality. In the Sects. 7 and 8, we shall consider a special case of such a more general combination problem. Instead of just taking the (disjoint) union of a commutative/monoidal $\mathscr{E}$ with a certain other theory $\mathscr{H}$, we add equations involving symbols of both signatures to make sure that the resulting theory is again a commutative/monoidal theory. On the corresponding semirings, this operation corresponds to a well-known mathematical construction.

In order to make this more precise, let us consider two of the examples in [2, 3]. Using algebraic properties of the semiring of polynomials with nonnegative integer coefficients, $\mathbf{N}[X]$, it was shown in [3] that the corresponding theory, *i.e.*, the theory of Abelian monoids with a homomorphism, is of unification type zero. In contrast, the theory of Abelian monoids with an involution[1] is unitary (finitary w.r.t. unification with constants). In both cases, the corresponding semiring has a specific structure: it is a monoid semiring $\mathscr{S}\langle H\rangle$, *i.e.*, a semiring $\mathscr{S}$ with an adjoint monoid H. In the first example, the monoid H is the free monoid in one generator, which is an infinite monoid, while in the second example, we have the cyclic group of order two, which is finite. In both examples, the semiring $\mathscr{S}$ is the semiring $\mathbf{N}$ of all nonnegative integers. This semiring corresponds to the theory AM of all commutative monoids, which is a finitary commutative/monoidal theory. The theory of Abelian monoids with an involution is obtained from AM as follows: Take the union of the identities

[1] An involution is a homomorphism h satisfying $h^2(x) = x$

defining Abelian monoids (*i.e.*, $\{x+y \doteq y+x, (x+y)+z \doteq x+(y+z), x+0 \doteq x\}$) with the identities defining an involution (*i.e.*, $\{h(h(x) \doteq x\}$), and add the equations that make sure that the involution acts as a homomorphism (*i.e.*, $\{h(x+y) \doteq h(x)+h(y), h(0) \doteq 0\}$). The combination is non-disjoint because of these additional equations.

In the present paper we shall consider this type of combination more closely. The result for the theory of Abelian monoids with a homomorphism can now be generalized to a whole class of theories as follows. If $\mathcal{S}$ is a *strict semiring* – *i.e.*, a semiring that is not a ring – and H is a *free monoid* then the commutative/monoidal theory corresponding to $\mathcal{S}\langle H\rangle$ is of unification type zero. On the other hand, assume that $\mathcal{S}$ is a semiring such that unification in the corresponding commutative/monoidal theory is unitary (finitary w.r.t unification with constants), and let H be a *finite monoid*. In this case, the theory corresponding to the semiring $\mathcal{S}\langle H\rangle$ is also of unification type unitary (finitary w.r.t. unification with constants). This generalizes the result for the theory of Abelian monoids with an involution. Moreover, a finite unification algorithm for the theory corresponding to $\mathcal{S}$ can be used to derive a finite unification algorithm for the theory corresponding to $\mathcal{S}\langle H\rangle$. These two general results demonstrate the usefulness of the algebraic approach to unification for investigating non-disjoint combination problems. With this approach one can determine the unification types of whole classes of combined theories. It is not at all clear how this could be achieved with a purely syntactical approach.

The paper is organized as follows. First, we shall motivate the need for unification results for whole classes of theories by an extended example. After recalling some basic definitions concerning equational theories, unification theory, and semirings in Sect. 3, we shall introduce commutative theories and monoidal theories in Sect. 4. This section also contains a proof of the equivalence between commutative and monoidal theories. In addition, it will be shown that being commutative/monoidal is an undecidable property of (finitely presented) equational theories. In Sect. 5 we recall the algebraic characterizations of the unification types for commutative/monoidal theories, and give some examples for the results that can be obtained using these characterizations. As a new result we shall show that solvability of unification problems is in general undecidable for commutative/monoidal theories. In Sect. 6, we consider unification with linear constant restrictions in commutative/monoidal theories. The next two sections contain the exact formulations and the proofs of the two results on non-disjoint combination mentioned above. In the conclusion we shall state some interesting open problems in this area.

This article is an improved and extended version of a conference paper [4].

2 A Motivating Example

As mentioned in the introduction, unification modulo equational theories has applications in various areas of automated deduction, such as theorem proving with built-in theories, logic programming with equality, and term rewriting modulo equational theories. In this section, we shall illustrate the third type of application by an example.

Consider the following equational theory $\mathcal{E}_K$, which consists of the axioms for Boolean rings, and two additional identities which state that the unary function symbol b is a homomorphism for the addition and the unit of the Boolean ring:

$$\begin{array}{llll}
(1) & (x+y)+z \doteq x+(y+z) & (2) & x+y \doteq y+x \\
(3) & x+0 \doteq x & (4) & x+(-x) \doteq 0 \\
(5) & x+x \doteq 0 & & \\
(6) & (x+y)*z \doteq (x*z)+(y*z) & & \\
(7) & (x*y)*z \doteq x*(y*z) & (8) & x*y \doteq y*x \\
(9) & x*1 \doteq x & (10) & x*x \doteq x \\
(11) & b(x)*b(y) \doteq b(x*y) & (12) & b(1) \doteq 1.
\end{array}$$

This theory is an axiomatization of equivalence for formulae in the propositional modal logic **K** [22]. Here, "$+$" stands for XOR, "$*$" for conjunction, "1" for truth, "0" for falsity, and "b" for the box operator of the modal logic. Thus, identity (11) expresses that for arbitrary (modal) formulae ϕ and ψ, the formulae $\Box\phi \wedge \Box\psi$ and $\Box(\phi \wedge \psi)$ are equivalent in **K**, *i.e.*, interpreted by the same truth value in any world of any Kripke model [12].

In order to decide the word problem for $\mathscr{E}_K$ (*i.e.*, the equivalence problem for **K**) one can try to construct a canonical rewrite system for this theory. However, commutativity cannot be oriented to a terminating rewrite rule. A possible solution would be to keep some of the axioms of $\mathscr{E}_K$ as unoriented identities, and proceed by using rewriting modulo these identities. But then critical pairs must also be computed by a unification algorithm modulo the unoriented identities.

For example, if we leave the identities (1), (2), (7) and (8) unoriented, and orient the remaining identities from left to right, then the obtained rewrite system is a canonical rewrite system modulo associativity and commutativity of "$+$" and "$*$". It should be noted, however, that the decision procedure for the word problem obtained this way is not very efficient. One could now try to leave even more identities unoriented, and thus leave more of the work to a (special purpose) matching procedure, and less to the actual reduction process. A necessary prerequisite for doing this is that one has a unification algorithm for the set of unoriented identities.[2]

For the addition, one could consider (1), (2) and (3), which is the theory of Abelian monoids, or (1), (2), (3) and (4), which is the theory of Abelian groups. For the multiplication, one could consider (7), (8), (9) and (10), which is the theory of idempotent Abelian monoids, or (7), (8), (9), (11) and (12), which is the theory of Abelian monoids with a homomorphism. All these theories have a very similar structure: they describe properties of a binary function symbol that is associative and commutative and satisfies some additional properties. To facilitate the design of unification algorithms for these theories it would be convenient to have a general framework in which unification modulo such theories can be treated. It turns out that commutative/monoidal theories provide such a framework: all the theories mentioned above belong to this class.

The example can also be used to illustrate why it is important to have a solution for the combination problem for (disjoint) monoidal/commutative theories. For example, assume that we want to use rewriting modulo (1), (2), (3) and (4) for the

[1] We do not claim that finding this unification algorithm is the only problem that must be solved here. For example, one usually also needs a compatible reduction ordering, which is a requirement that cannot be staisfied for any set of identities containing identity (10)

addition, and (7), (8) and (9) for the multiplication. Even if we have unification algorithms for the theory of Abelian groups of AG_+ and for the theory of Abelian monoids AM_*, these cannot directly be used to compute the necessary critical pairs. This is so because the terms to be unified may contain both "+" and "*" as well as the additional free (for AG_+ and AM_*) function symbol b. In other words, one must unify modulo the union of AG_+, AM_*, and the "free" theory $\{b(x) \doteq b(x)\}$. The result in Sect. 6 will show that a unification algorithm modulo this combined theory can effectively be obtained from algorithms for the single theories.

3 Basic Definitions

In the following we assume that the reader is familiar with the basic notions of universal algebra [8, 15]. For more information on unification theory see [6]. The notions from category theory used below are for instance defined in [1], or in any introductory textbook on categories. The composition of mappings is written from left to right, that is, $\phi \circ \psi$ or simply $\phi\psi$ means first ϕ and then ψ. Consequently, we use suffix notation for mappings (but not for function symbols in terms).

3.1 *Equational Theories*

We assume that two disjoint infinite sets of symbols are given, a set of function symbols and a set of variables. A signature Σ is a finite set of function symbols each of which is associated with its arity. Every signature Σ determines a class of *Σ-algebras* and *Σ-homomorphisms*. We define Σ-terms and Σ-substitutions as usual. By $[x_1/t_1, \ldots, x_n/t_n]$ we denote the substitution which replaces the variables x_i by the terms t_i.

An *equational theory* $\mathscr{E}$-(Σ, E) is a pair consisting of a signature Σ and a set of identifies E. The equality of Σ-terms induced by $\mathscr{E}$ will be denoted by $=_{\mathscr{E}}$. Every equational theory $\mathscr{E}$ determines a variety $\mathscr{V}(\mathscr{E})$, the class of all Σ-algebras satisfying each identity of E. For any set of generators X, the variety $\mathscr{V}(\mathscr{E})$ contains a free algebra over $\mathscr{V}(\mathscr{E})$ with generators X, which will be denoted by $\mathscr{F}_{\mathscr{E}}(X)$. Thus any mapping of X into a Σ-algebra A can be uniquely extended to a Σ-homomorphism of $\mathscr{F}_{\mathscr{E}}(X)$ into A.

The following category $\mathscr{C}(\mathscr{E})$ is associated with each equational theory $\mathscr{E} = (\Sigma, E)$: the objects of $\mathscr{C}(\mathscr{E})$ are the free algebras $\mathscr{F}_{\mathscr{E}}(X)$ for finite sets of variables X, the morphisms of $\mathscr{C}(\mathscr{E})$ are the Σ-homomorphisms between free algebras, and the composition of morphisms is the usual composition of mappings. The set of all objects of $\mathscr{C}(\mathscr{E})$ will be denoted by $\mathscr{F}(\mathscr{E})$, and the set of all morphisms from an object $\mathscr{F}_{\mathscr{E}}(X)$ to an object $\mathscr{F}_{\mathscr{E}}(Y)$ by $hom(\mathscr{F}_{\mathscr{E}}(X), \mathscr{F}_{\mathscr{E}}(Y))$. The coproduct of $\mathscr{F}_{\mathscr{E}}(X)$ and $\mathscr{F}_{\mathscr{E}}(Y)$ in $C(\mathscr{E})$ is given by the free algebra $\mathscr{F}_{\mathscr{E}}(X \uplus Y)$, where $\uplus$ denotes disjoint union. If $|X| = |Y|$, then $\mathscr{F}_{\mathscr{E}}(X)$ and $\mathscr{F}_{\mathscr{E}}(Y)$ are isomorphic. Thus $\mathscr{F}_{\mathscr{E}}(X)$ is the coproduct of the isomorphic objects $\mathscr{F}_{\mathscr{E}}(x)$ for $x \in X$, where x is used as abbreviation for the singleton $\{x\}$.

3.2 *Unification*

Let $\mathscr{E} = (\Sigma, E)$ be an equational theory. An *$\mathscr{E}$-unification problem* is a finite sequence of equations $\Gamma = \langle s_i \doteq t_i \mid 1 \leq i \leq n \rangle$, where s_i and t_i are Σ-terms. A substitution δ is called an *$\mathscr{E}$-unifier* of Γ if $s_i\delta =_{\mathscr{E}} t_i\delta$ for each i. The set of all $\mathscr{E}$-unifiers of Γ is denoted by $U_{\mathscr{E}}(\Gamma)$. In general one does not need the set of all $\mathscr{E}$-unifiers. A complete set of $\mathscr{E}$-unifiers, *i.e.*, a set of $\mathscr{E}$-unifiers from which all unifiers may be generated by $\mathscr{E}$-instantiation, is usually sufficient. More precisely, for every set of variables V we extend "$=_{\mathscr{E}}$" to a relation "$=_{\mathscr{E},V}$" between substitutions, and introduce the *$\mathscr{E}$-instantiation quasi-ordering* "$\leqq_{\mathscr{E},V}$" as follows:

- $\delta =_{\mathscr{E},V} \eta$ iff $x\delta =_{\mathscr{E}} x\eta$ for all $x \in V$
- $\delta \leqq_{\mathscr{E},V} \eta$ if there exists a substitution λ such that $\eta =_{\mathscr{E},V} \delta \circ \lambda$.

A set $C \subseteq U_{\mathscr{E}}(\Gamma)$ is a *complete set of $\mathscr{E}$-unifiers* of Γ if for every unifier η of Γ there exists $\delta \in C$ such that $\delta \leqq_{\mathscr{E},V} \eta$, where V is the set of variables occurring in Γ. For reasons of efficiency, this set should be as small as possible. Thus one is interested in *minimal* complete sets of $\mathscr{E}$-unifiers. In minimal complete sets two different elements are not comparable w.r.t. $\mathscr{E}$-instantiation.

The *unification type* of a theory $\mathscr{E}$ is defined with reference to the existence and cardinality of minimal complete sets. The theory $\mathscr{E}$ is *unitary* (*finitary*, *infinitary*, respectively) if minimal complete sets of $\mathscr{E}$-unifiers always exist, and their cardinality is at most one (always finite, at least once infinite, respectively). The theory $\mathscr{E}$ is of *unification type zero* if there exists an $\mathscr{E}$-unification problem without a minimal complete set of $\mathscr{E}$-unifiers.

If the terms in the unification problems may contain free constants, we talk about *unification with constants*, otherwise we talk about *unification without constants*. In many applications, the unification problems that occur contain not only free constants, but also additional free function symbols of arity larger than 0. Such problems will be called *general unification problems*. If nothing else is specified, "unification" will mean "unification without constants."

An $\mathscr{E}$-unification problem $\Gamma = \langle s_1 \doteq t_1, \ldots, s_n \doteq t_n \rangle$ can be reformulated as a problem for morphisms in the category $\mathscr{C}(\mathscr{E})$. Let Y be the finite set of variables occurring in some s_i or t_i. Evidently, we can consider s_i and t_i as elements of $\mathscr{F}_{\mathscr{E}}(Y)$. Since we do not distinguish between $=_{\mathscr{E}}$-equivalent unifiers, any $\mathscr{E}$-unifier can be regarded as a Σ-homomorphism from $\mathscr{F}_{\varepsilon}(Y)$ into $\mathscr{F}_{\mathscr{E}}(Z)$ for some finite set of variables Z. Let $X = \{x_1, \ldots, x_n\}$ be a set of cardinality n. We define Σ-homomorphisms $\sigma, \tau : \mathscr{F}_{\mathscr{E}}(X) \to \mathscr{F}_{\mathscr{E}}(Y)$ by $x_i\sigma := s_i$ and $x_i\tau := t_i$. Now, $\delta : \mathscr{F}_{\mathscr{E}}(Y) \to \mathscr{F}_{\mathscr{E}}(Z)$ is an $\mathscr{E}$-unifier of Γ iff $x_i\sigma\delta = s_i\delta = t_i\delta = x_i\tau\delta$ for all i, that is, iff $\sigma\delta = \tau\delta$. This observation justifies to conceive $\mathscr{E}$-unification as a problem involving only morphisms of the category $\mathscr{C}(\mathscr{E})$: given $\sigma, \tau : \mathscr{F}_{\mathscr{E}}(X) \to \mathscr{F}_{\mathscr{E}}(Y)$, find a $\delta : \mathscr{F}_{\mathscr{E}}(Y) \to \mathscr{F}_{\mathscr{E}}(Z)$ such that $\sigma\delta = \tau\delta$.

3.3 Semirings

A *semiring* $\mathscr{S}$ is a tuple $(\mathscr{S}, +, 0, \cdot, 1)$ such that $(\mathscr{S}, +, 0)$ is an Abelian monoid, $(\mathscr{S}, \cdot, 1)$ is a monoid, and all $q, r, s \in \mathscr{S}$ satisfy the equalities

$$(1)\quad (q + r)\cdot s = q\cdot s + r\cdot s \qquad (2)\quad q\cdot(r + s) = q\cdot r + q\cdot s$$

$$(3)\quad 0\cdot s = 0 \qquad (4)\quad s\cdot 0 = 0.$$

The elements 0 and 1 are called *zero* and *unit*. Semirings are different from rings in that they need not be groups w.r.t. addition. Obviously, any ring is a semiring. A prominent example for a semiring which is not a ring is the semiring $\mathbf{N}$ of nonnegative integers.

Similar to the construction of polynomial rings over a given ring, one can use a semiring $\mathscr{S}$ and a monoid H to construct a new semiring, namely the *monoid semiring* $\mathscr{S}\langle H\rangle$. As for polynomials, the elements of the monoid semiring may be represented as sums of the form $\sum_{h\in H} s_h\cdot h$ where only finitely many of the coefficients $s_h\in\mathscr{S}$ are nonzero. The zero elements of $\mathscr{S}\langle H\rangle$ is the sum where all the coefficients are zero, and the unit element is the sum where only the unit of H has a coefficient different from zero and this coefficient is the unit element of $\mathscr{S}$. Addition and multiplication in $\mathscr{S}\langle H\rangle$ are defined as follows:

$$\sum_{h\in H} s_h\cdot h + \sum_{h\in H} t_h\cdot h = \sum_{h\in H} (s_h + t_h)\cdot h$$

$$\sum_{f\in H} s_f\cdot f\cdot \sum_{g\in H} t_g\cdot g = \sum_{h\in H}\left(\sum_{h=fg} s_f\cdot t_g\right)\cdot h$$

Polynomial semirings are special cases of monoid semirings. For example, the ring $\mathbf{Z}[X_1,\ldots,X_n]$ of integer polynomials in n indeterminates is the monoid semiring $\mathbf{Z}\langle \mathrm{FAM}_n\rangle$ where FAM_n denotes the free Abelian monoid in n generators.

As mentioned in the introduction, unification in commutative/monoidal theories can be reduced to solving systems of linear equations in certain semirings. Similar to unification in Abelian monoids [23], problems without constants will correspond to systems of homogeneous equations. For problems with constants one has to solve in addition systems of inhomogeneous equations.

Modules over semirings are a generalization of vector spaces over fields. Since $(\mathscr{S}, \cdot, 1)$ need not be commutative, we have to distinguish between left and right $\mathscr{S}$-modules. Solutions of homogeneous systems form right $\mathscr{S}$-modules. The unification type of a theory will depend on whether these modules are finitely generated or not. A subset M of the n-fold Cartesian product $\mathscr{S}^n$ is a *finitely generated* right $\mathscr{S}$-module if there exist finitely many $x_1,\ldots,x_k\in\mathscr{S}^n$ such that $M=\{x_1s_1+\cdots+x_ks_k\,|\,s_1,\ldots,s_k\in\mathscr{S}\}$.

Solutions of inhomogeneous systems do not form right modules, but unions of cosets of right modules. For the unification type it will be crucial how many cosets are needed to represent all solutions. If $M\subseteq\mathscr{S}^n$ is a right $\mathscr{S}$-module, and N is a subset of $\mathscr{S}^n$, then N is a *coset* of M if there exists some $y\in\mathscr{S}^n$ such that $N=\{y+x\,|\,x\in M\}$. Consequently, the set N is a *finite union of cosets* of M iff there exist finitely many $y_1,\ldots,y_k\in\mathscr{S}^n$ such that $N=\bigcup_{i=1}^k\{y_i+x\,|\,x\in M\}$.

4 Commutative and Monoidal Theories

In this section we shall give the definitions of commutative and monoidal theories, and show in what sense these two notions are equivalent.

4.1 Definitions and Examples

Motivated by the categorical reformulation of $\mathscr{E}$-unification (see Subsect. 3.2), the class of commutative theories is defined by properties of the category $\mathscr{C}(\mathscr{E})$ of finitely generated $\mathscr{E}$-free algebras as follows: an equational theory $\mathscr{E}$ is commutative if the corresponding category $\mathscr{C}(\mathscr{E})$ is semiadditive (see [16, 1] for the definition and for properties of semiadditive categories). In order to give a more algebraic definition we need some additional notation from universal algebra.

Let $\mathscr{E} = (\Sigma, E)$ be an equational theory. A constant symbol e of the signature Σ is called *idempotent in* $\mathscr{E}$ if for all symbols $f \in \Sigma$ we have $f(e, \ldots, e) =_{\mathscr{E}} e$. Note that for nullary f this means $f =_{\mathscr{E}} e$.

Let $\mathscr{K}$ be a class of Σ-algebras. An *n-ary implicit operation* in $\mathscr{K}$ is a family $o = \{o_A | A \in \mathscr{K}\}$ of mappings $o_A: A^n \to A$ which is compatible with all homomorphisms, *i.e.*, for all homomorphisms ω: $A \to B$ with $A, B \in \mathscr{K}$ and all $a_1, \ldots, a_n \in A$, we have $(o_A(a_1, \ldots, a_n))\omega = o_B(a_1\omega, \ldots, a_n\omega)$. In the sequel we shall omit the index and just write o in the place of o_A. Σ-terms induce implicit operations on any class of Σ-algebras in the following way: let t be a Σ-term and let $x_1, \ldots, x_n$ be a sequence of variables such that all the variables occurring in t are contained in this sequence. The n-ary implicit operation $(t; x_1, \ldots, x_n)$ is defined by

$$(a_1, \ldots, a_n) \mapsto t[x_1/a_1, \ldots, x_n/a_n].$$

For example, assume that the signature consists of a binary symbol "$\cdot$" and a unary symbol "$^{-1}$", and let $\mathscr{K}$ be the class of all groups. Then the binary implicit operation $(x \cdot y^{-1}; x, y)$ expresses division in a group. If we apply this operation to a pair of group elements a, b, we obtain the quotient $a \cdot b^{-1}$. For the classes $\mathscr{V}(\mathscr{E})$ and $\mathscr{F}(\mathscr{E})$ all implicit operations can be defined by Σ-terms [21].

We are now ready to give an algebraic definition of commutative theories. An equational theory $\mathscr{E} = (\Sigma, E)$ is called *commutative* if the following holds:

1. the signature Σ contains a constant symbol e that is idempotent in $\mathscr{E}$,
2. there is a binary implicit operation "$*$" in $\mathscr{F}(\mathscr{E})$ such that
   (a) the constant e is a neutral element for "$*$" in any algebra $\mathscr{F}_{\mathscr{E}}(X) \in \mathscr{F}(\mathscr{E})$,
   (b) for any n-ary function symbol $f \in \Sigma$, any algebra $\mathscr{F}_{\mathscr{E}}(X) \in \mathscr{F}(\mathscr{E})$, and any $s_1, \ldots, s_n, t_1, \ldots, t_n \in \mathscr{F}(\mathscr{E})$ we have $f(s_1 * t_1, \ldots, s_n * t_n) = f(s_1, \ldots, s_n) * f(t_1, \ldots, t_n)$.

Though it is not explicitly required by the definition, the implicit operation "$*$" turns out to be associative and commutative (see [1], Corollary 5.4). This justifies the name "commutative theory." An obvious consequence of the definition of an idempotent constant is that the initial algebra, *i.e.*, $\mathscr{F}_{\mathscr{E}}(\emptyset)$, is of cardinality one for any commutative theory $\mathscr{E}$.

Well-known examples of commutative theories are the theory AM of Abelian monoids (sometimes called AC1 in the literature), the theory AIM of idempotent Abelian monoids, and the theory AG of Abelian groups (see [1]). In these theories, the implicit operation "$*$" is given by the explicit binary operation in the signature. An example for a commutative theory where "$*$" is really implicit can also be found in [1] (Example 5.1). We shall now consider examples of commutative theories where the signature contains some additional function symbols (see [29, 3] for more examples).

Example 4.1 We consider the following signatures: $\Sigma := \{+, 0, h\}$, where "$+$" is binary, 0 is nullary, and h is unary; $\Delta := \{+, 0, f\}$, where "$+$" is binary, 0 is nullary, and f is binary; and $\Omega := \{+, 0, -, i\}$, where "$+$" is binary, 0 is nullary, and $-$ and i are unary.

AMH $= (\Sigma, E_{\text{AMH}})$, the theory of *Abelian monoids with a homomorphism.* E_{AMH} consists of the identities which state that "+" is associative, commutative with neutral element 0, and the identities which state that h is a homomorphism, *i.e.*, the identities $h(x+y) \doteq h(x)+h(y)$, $h(0) \doteq 0$.

AMIn $= (\Sigma, E_{\text{AMIn}})$, the theory of *Abelian monoids with an involution.* E_{AMIn} consists of the identities of E_{AMH}, and the additional identity $h(h(x)) \doteq x$, which states that h is an involution.

COM $= (\Delta, E_{\text{COM}})$. E_{COM} consists of the identities which state that "+" is associative, commutative with neutral element 0, and the identities $f(x+x', y+y') \doteq f(x,y)+f(x',y')$ and $f(0,0) \doteq 0$ which ensure that COM is really commutative.

GAUSS $= (\Omega, E_{\text{GAUSS}})$. E_{GAUSS} consists of the identities which state that "+" is the binary operation of an Abelian group with neutral element 0 and inverse $-$, and the additional identity $x + i(i(x)) \doteq 0$.

With the exception of the third example, the additional function symbols – *i.e.*, the function symbols apart from the binary symbol yielding the implicit operation, and the idempotent constant symbol – are all unary symbols. This motivatives the definition of monoidal theories. An equational theory $\mathscr{E} = (\Sigma, E)$ is *monoidal* if

1. Σ contains a constant symbol 0, a binary function symbol "+", and all the other symbols in Σ are unary
2. "+" is associative and commutative
3. 0 is the neutral element for "+", that is, $0 + x =_{\mathscr{E}} x + 0 =_{\mathscr{E}} x$
4. every unary symbol h is a homomorphism for "+" and 0, that is, $h(x+y) =_{\mathscr{E}} h(x) + h(y)$ and $h(0) =_{\mathscr{E}} 0$.

It is easy to see that monoidal theories are always commutative theories. Obviously, the theories AM, AIM, AG, AMH, AMIn, and GAUSS are monoidal. The theory COM is not monoidal, since its signature contains an additional *binary* function symbol. However, we shall see in the next subsection that COM may also be regarded as monoidal theory if the signature is translated appropriately.

4.2 *Commutative and Monoidal Theories are Equivalent*

Next we show that by means of a signature transformation every commutative theory can be turned into a monoidal theory that, from the viewpoint of unification, is equivalent.

Let Σ and Σ' be signatures. A *signature transformationfrom Σ' to Σ* is a mapping θ that associates to every Σ'-term a Σ-term such that

1. $x\theta = x$ for every variable x
2. $f(t_1, \ldots, t_n)\theta = (f(x_1, \ldots, x_n)\theta)[x_1/t_1\theta, \ldots, x_n/t_n\theta]$ if f is an n-ary symbol and $x_1, \ldots, x_n$ are n distinct variables.

It follows from the definition that θ is completely defined by the images of the flat terms $f(x_1, \ldots, x_n)$ where f ranges over Σ'. Intuitively, θ interprets every Σ'-symbol by a Σ-term, and then extends this interpretation consistently to arbitrary Σ'-terms.

To every commutative theory $\mathscr{E} = (\Sigma, E)$ we associate a theory $\hat{\mathscr{E}} = (\hat{\Sigma}, \hat{E})$ and a signature transformation θ from $\hat{\Sigma}$ to Σ as follows. The signature $\hat{\Sigma}$ consists of a constant 0, a binary symbol "+", and unary symbols $f_1, \ldots, f_n$ for every n-ary symbol $f \in \Sigma$, where $n \geqq 1$. To define the set of identities $\hat{E}$ we need the transformation θ. Let e be the idempotent constant in $\mathscr{E}$ and let $(t_*; x, y)$ be the pair corresponding to the implicit operation "$*$" in $\mathscr{E}$. We define θ by $0\theta := e$, $(x+y)\theta := t_*$, and $f_i(x)\theta := f(e, \ldots, x, \ldots, e)$, where $f(e, \ldots, x, \ldots, e)$ has the variable x in the i-th argument position and the constant e in the other positions. Now, with the help of this signature transformation we define $\hat{E}$ as $\hat{E}\colon \{\hat{s} \doteq \hat{t} \mid \hat{s}\theta =_{\mathscr{E}} \hat{t}\theta\}$. That is, $\hat{E}$ is the preimage of "$=_{\mathscr{E}}$" under θ.

Proposition 4.2 *Let $\mathscr{E} = (\Sigma, E)$ be a commutative theory with associated theory $\hat{\mathscr{E}} = (\hat{\Sigma}, \hat{E})$ and signature transformation θ. Then:*

1. *$\hat{\mathscr{E}}$ is a monoidal theory*
2. *$\hat{s} =_{\hat{\mathscr{E}}} \hat{t}$ implies $\hat{s}\theta =_{\mathscr{E}} \hat{t}\theta$ for all $\hat{\Sigma}$-terms $\hat{s}, \hat{t}$.*

Proof. 1. Since the implicit operation "$*$" is associative and commutative, the same is true for "+". From part (2.b) of the definition of commutative theories we conclude that every f_i is a homomorphism for "+". Finally, since e is neutral for "$*$", we have that 0 is a zero for "+", and since e is idempotent, we conclude that 0 is a zero for the homomorphism f_i.

2. The claim follows from the definition of $\hat{E}$ and the fact that $\hat{E}$ is a stable congruence, *i.e.*, a congruence that is invariant under substitution. □

Let $\mathscr{E} = (\Sigma, E)$ and $\mathscr{E}' = (\Sigma', E')$ be equational theories. We say that $\mathscr{E}$ and $\mathscr{E}'$ are *equivalent* if there exist signature transformation θ' from Σ to Σ' and θ from Σ' to Σ such that

1. $s =_{\mathscr{E}} t$ implies $s\theta' =_{\mathscr{E}'} t\theta'$ for all Σ-terms s and t and $s' =_{\mathscr{E}'} t'$ implies $s'\theta =_{\mathscr{E}} t'\theta$ for all Σ'-terms s' and t'
2. $s\theta'\theta =_{\mathscr{E}} s$ for all Σ-terms s, and $s'\theta\theta' =_{\mathscr{E}'} s'$ for all Σ'-terms s'.

The first condition means that θ and θ' can be seen as mappings on equivalence classes of terms. The second says that θ and θ' are inverses of each other modulo the equational theories.

One of the most prominent examples of equivalent theories are boolean rings and boolean algebras. If two theories are equivalent they describe essentially the same structures. More precisely, if $\mathscr{E}$ and $\mathscr{E}'$ are equivalent, then the categories $\mathscr{C}(\mathscr{E})$ and $\mathscr{C}(\mathscr{E}')$ are isomorphic, and so are the varieties of $\mathscr{E}$ and $\mathscr{E}'$ [36]. Since unification properties of a theory $\mathscr{E}$ depend on the category $\mathscr{C}(\mathscr{E})$, it follows that equivalent theories share the same unification properties.

Theorem 4.3 *Let $\mathscr{E} = (\Sigma, E)$ be a commutative theory with associated theory $\hat{\mathscr{E}} = (\hat{\Sigma}, \hat{E})$. Then $\mathscr{E}$ and $\hat{\mathscr{E}}$ are equivalent.*

Proof. Let θ be the signature transformation from $\hat{\Sigma}$ to Σ. To show the equivalence of $\mathscr{E}$ and $\hat{\mathscr{E}}$ we exhibit a signature transformation $\hat{\theta}$ from Σ to $\hat{\Sigma}$ and show that θ and $\hat{\theta}$ have the required properties. We define $\hat{\theta}$ by $e\hat{\theta} = 0$, and $f(x_1, \ldots, x_n)\hat{\theta} = f_1(x_1) + \cdots + f_n(x_n)$ for every n-ary symbol f in Σ.

By Proposition 4.2 we already know that $\hat{s} =_{\hat{\mathscr{E}}} \hat{t}$ implies $\hat{s}\theta =_{\mathscr{E}} \hat{t}\theta$ for all $\hat{\Sigma}$-terms $\hat{s}, \hat{t}$.

Next we prove that $s\hat{\theta}\theta =_{\mathscr{E}} s$ for every Σ-term s. For this purpose it suffices to show the claim for flat terms of the form $f(x_1, \ldots, x_n)$. For such terms we have

$$\begin{aligned} f(x_1,\dots,x_n)\hat{\theta}\theta &= (f_1(x_1)+\cdots+f_n(x_n))\theta \\ &= f(x_1,e,\dots)*\cdots*f(\dots,e,x_n) \\ &=_{\mathscr{E}} f(x_1*e*\cdots*e,\dots,e*\cdots*e*x_n) \\ &=_{\mathscr{E}} f(x_1,\dots,x_n), \end{aligned}$$

where the first two equalities follow from the definition of $\hat{\theta}$ and θ, and the last two equalities follow from parts (2.b) and (2.a) of the definition of commutative theories.

To show that $\hat{s}\theta\hat{\theta} =_{\hat{\mathscr{E}}} \hat{s}$ for every $\hat{\Sigma}$-term $\hat{s}$, it suffices by the definition of $\hat{E}$ to show that $\hat{s}\theta\hat{\theta}\theta =_{\mathscr{E}} \hat{s}\theta$, which is a consequence of the fact that $s\hat{\theta}\theta =_{\mathscr{E}} s$ for every Σ-term s.

Finally, we show that for all Σ-terms s, t we have that $s =_{\mathscr{E}} t$ implies $s\hat{\theta} =_{\hat{\mathscr{E}}} t\hat{\theta}$. But this follows again from the definition of $\hat{E}$, since $s\hat{\theta}\theta =_{\mathscr{E}} s =_{\mathscr{E}} t =_{\mathscr{E}} t\hat{\theta}\theta$ then yields $s\hat{\theta} =_{\hat{\mathscr{E}}} t\hat{\theta}$. □

From this result it follows that from the viewpoint of unification there is no difference between commutative and monoidal theories.

4.3 *Adding Monoids of Homomorphisms*

There is an interesting difference between the theory GAUSS on the one hand, and the theories AMH and AMIn on the other hand. The additional identity $x+i(i(x)) \doteq 0$ in the theory GAUSS establishes a closer connection between the unary symbol i and the binary symbol "+" than just the fact that i is a homomorphism for "+". This is not the case for the additional identity $h(h(x)) \doteq x$ in AMIn which says something about h alone. This observation will now be put into a more general setting.

Let $\mathscr{E}=(\Sigma, E)$ be a monoidal theory, and let H be a monoid generated by the finitely many elements $h_1,\dots,h_n$. Since composition of unary functions is associative, one may consider the generators of H as unary function symbols, and represent the monoid H by an equational theory $\mathscr{E}_H$ that has these unary symbols as signature, and the set

$$E_H := \{h_{i_1}(\dots h_{i_k}(x)\dots) \doteq h_{j_1}(\dots(x)\dots) \mid h_{i_1}\cdots h_{i_k} = h_{j_1}\cdots h_{j_l} \text{ holds in } H\}$$

as its set of identities.

We define the augmented theory $\mathscr{E}\langle H\rangle = (\Sigma', E')$ as follows: the signature Σ' extends Σ by the unary function symbols $h_1,\dots,h_n$; the set of identities E' is the union of $E \cup E_H$ with the identities which state that $h_1,\dots,h_n$ are homomorphisms and that these homomorphisms commute with the unary functions in Σ, *i.e.*,

$$\begin{aligned} E' = E \cup E_H \cup \{h_i(x+y) \doteq h_i(x)+h_i(y), h_i(0) \doteq 0 \mid i=1,\dots,n\} \\ \cup \{h_i(f(x)) \doteq f(h_i(x)) \mid f \text{ is a unary symbol in } \Sigma, i=1,..,h\}. \end{aligned}$$

In Sects. 7 and 8 we shall study unification in theories of the form $\mathscr{E}\langle H\rangle$.

The theory AMH is AM$\langle h^*\rangle$ where h^* stands for the free monoid in one generator, and AMIn is AM$\langle Z_2\rangle$ where Z_2 stands for the cyclic group of order 2, *i.e.*, Z_2 consists of two elements e and h, and the multiplication in Z_2 is defined as $e\cdot e=e$, $h\cdot e=e\cdot h=h$, and $h\cdot h=e$. On the other hand, one can prove that GAUSS cannot be represented in the form AG$\langle H\rangle$ because of the interaction between i and "+" stated by $x+i(i(x)) \doteq 0$.

4.4 *Being Commutative/Monoidal is an Undecidable Property*

The goal of this subsection is to prove the undecidability result stated in the next theorem:

Theorem 4.4 *The following problem is in general undecidable*:

- *Given a finite signature Σ and a finite set of identities E.*
- *Is (Σ, E) monoidal (resp. commutative)?*

We shall reduce a known undecidable problem for monoids and groups to this problem. It is well-known (see, *e.g.*, [10], Corollary 3.8) that for finitely presented groups it is in general undecidable whether the group is trivial (*i.e.*, consists of only one element) or not. Since finitely presented groups are a special case of finitely presented monoids, this undecidability result holds for monoids as well.

Now assume that a finite presentation of a monoid is given. This presentation consists of a set of generators $\Delta = \{h_1, \ldots, h_n\}$ and a finite set of relations $R = \{u_1 = v_1, \ldots, u_m = v_m\}$, where the u_i and v_i are words over Δ. Obviously, the monoid presented this way is trivial if, and only if, for all generators h_i, the relation $h_i = \varepsilon$ follows from R (where ε denotes the empty word).

As in the previous subsection, we consider the set of generators Δ as a set of unary function symbols. The set of relations R can then be turned into a set of identities over this signature: $E_R := \{u_1(x) = v_1(x), \ldots, u_m(x) = v_m(x)\}$[3]. The monoid presented by Δ and R is trivial if, and only if, the equational theory $\mathscr{E}_R = (\Delta, E_R)$ satisfies $h_i(x) =_{\mathscr{E}_R} x$ for all i, $i = 1, \ldots, n$.

Let $\Sigma = \{+, 0\}$ for a binary symbol "$+$" and a constant symbol 0, and let AM be the theory that says that "$+$" is associative and commutative and that 0 is a neutral element for "$+$". As mentioned before, this theory is both commutative and monoidal.

The theorem is now an immediate consequence of the next lemma:

Lemma 4.5 *The monoid presented by Δ and R is trivial if, and only if, the equational theory $\mathscr{E} = (\Sigma \cup \Delta, E_{\mathrm{AM}} \cup E_R)$ is monoidal (resp. commutative).*

Proof. If the monoid is trivial then we have $h_i(x) =_{\mathscr{E}_R} x$ for all $i, i = 1, \ldots, n$. This implies that the h_i behave like homomorphisms for "$+$" and 0 in $\mathscr{E}$, since the identity is obviously a homomorphism. Consequently, $\mathscr{E}$ is monoidal, and thus also commutative.

Conversely, assume that the monoid is not trivial. Thus, there exists a generator h_i such that $h_i(x) \neq_{\mathscr{E}_R} x$. We claim that this implies $h_i(0) \neq_{\mathscr{E}} 0$. In fact, since the theory $\mathscr{E}$ is obtained as disjoint union of $\mathscr{E}_R$ and AM, one can use the Abstraction Lemma (Lemma 4.1) of [5] to show that $h_i(0) =_{\mathscr{E}} 0$ implies $h_i(x) =_{\mathscr{E}_R} x$.[4]

Obviously, $h_i(0) \neq_{\mathscr{E}} 0$ shows that $\mathscr{E}$ is not monoidal. In addition, this inequality also implies that the initial $\mathscr{E}$-algebra is of cardinality greater than one, and thus $\mathscr{E}$ cannot be commutative (since there is no idempotent constant). □

[3] For $u = h_{i_1} \cdots h_{i_k}$, we use $u(x)$ as an abbreviation for $h_{i_1}(\cdots h_{i_k}(x) \cdots)$.

[4] Note that one can without loss of generality assume that the substitution $[x/0]$ satisfies the normalization requirement in Lemma 4.1 of [5], since one can choose an ordering in which 0 is minimal.

5 Unification in Commutative/Monoidal Theories

We first show how to construct in a canonical way a semiring from a commutative/monoidal theory and how to use it for solving unification problems with and without constants. Then we exhibit a commutative/monoidal theory where unification with constants and the existence of nontrivial solutions for unification without constants are undecidable.

5.1 Commutative/Monoidal Theories and Semirings

In [1] the following properties for a commutative theory $\mathscr{E}$ are shown within the categorical framework, using well-known results for semiadditive categories.

1. The implicit operation "$*$" required in the definition of commutative theories induces a binary operation "$+$" on any morphism set $hom(\mathscr{F}_{\mathscr{E}}(X), \mathscr{F}_{\mathscr{E}}(Y))$ as follows: for $\sigma, \tau: \mathscr{F}_{\mathscr{E}}(X) \to \mathscr{F}_{\mathscr{E}}(Y)$ we define $\sigma + \tau$ by $t(\sigma + \tau) := (t\sigma) * (t\tau)$ for all $t \in \mathscr{F}_{\mathscr{E}}(X)$. This operation is associative and commutative, and it distributes with the composition of morphisms. The morphism $0: \mathscr{F}_{\mathscr{E}}(X) \to \mathscr{F}_{\mathscr{E}}(Y)$ defined by $x \mapsto e$ for all $x \in X$, where e is the idempotent constant required in the definition of commutative theories, is a neutral element for "$+$" on $hom(\mathscr{F}_{\mathscr{E}}(X), \mathscr{F}_{\mathscr{E}}(Y))$.

2. The Cartesian product of $\mathscr{F}_{\mathscr{E}}(X)$ and $\mathscr{F}_{\mathscr{E}}(Y)$ is also a product in the categorical sense. Furthermore, the product is isomorphic to the coproduct, that is $\mathscr{F}_{\mathscr{E}}(X \uplus Y) \simeq \mathscr{F}_{\mathscr{E}}(X) \times \mathscr{F}_{\mathscr{E}}(Y)$. The canonical injection $\iota_X: \mathscr{F}_{\mathscr{E}}(X) \to \mathscr{F}_{\mathscr{E}}(X \uplus Y)$ is given by $x\iota_X := x$ for any $x \in X$. The canonical projection $\pi_X: \mathscr{F}_{\mathscr{E}}(X \uplus Y) \to \mathscr{F}_{\mathscr{E}}(X)$ is given by $x\pi_X := x$ for $x \in X$ and $y\pi_{\mathscr{E}} := 0$ for $y \in Y$. If $X = \{x\}$ is a singleton, we write ι_X and π_x instead of $\iota_{\{x\}}$, and $\pi_{\{x\}}$, respectively.

3. Consider $\sigma: \mathscr{F}_{\mathscr{E}}(X) \to \mathscr{F}_{\mathscr{E}}(Y)$. Let ι_X for $x \in X$ be the injections of the coproduct $\mathscr{F}_{\mathscr{E}}(X) = \bigoplus_{x \in X} \mathscr{F}_{\mathscr{E}}(x)$ and π_y for $y \in Y$ be the projections of the product $\mathscr{F}_{\mathscr{E}}(Y) = \bigotimes_{y \in Y} \mathscr{F}_{\mathscr{E}}(y)$. Then σ is uniquely determined by the matrix $M_\sigma := (\iota_x \sigma \pi_y)_{x \in X, y \in Y}$. For $\sigma, \tau: \mathscr{F}_{\mathscr{E}}(X) \to \mathscr{F}_{\mathscr{E}}(Y)$ and $\delta: \mathscr{F}_{\mathscr{E}}(Y) \to \mathscr{F}_{\mathscr{E}}(Z)$, we have $M_{\sigma+\tau} = M_\sigma + M_\tau$, and $M_{\sigma\delta} = M_\sigma M_\delta$.

As an example, consider the morphism $\sigma = [x_1/h(y_1), x_2/y_1 + h^2(y_2)]$ from $\mathscr{F}_{\mathrm{AMH}}(x_1, x_2)$ to $\mathscr{F}_{\mathrm{AMH}}(y_1, y_2)$. Then σ is determined by the matrix.

$$M_\sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} [x_1/h(y_1)] & [x_1/0] \\ [x_2/y_1] & [x_2/h^2(y_2)] \end{pmatrix}.$$

Let $\mathbf{1}$ be an arbitrary set of cardinality one. Property 1 from above yields that the set $hom(\mathscr{F}_{\mathscr{E}}(\mathbf{1}), \mathscr{F}_{\mathscr{E}}(\mathbf{1}))$ with addition "$+$" and composition as multiplication is a semiring, which will be denoted by $\mathscr{S}_{\mathscr{E}}$. Any $\mathscr{F}_{\mathscr{E}}(x)$ is isomorphic to $\mathscr{F}_{\mathscr{E}}(\mathbf{1})$, and thus, for $|X| = n$, $\mathscr{F}_{\mathscr{E}}(X)$ is the n-th power and copower of $\mathscr{F}_{\mathscr{E}}(\mathbf{1})$. Consequently, for $\sigma: \mathscr{F}_{\mathscr{E}}(X) \to \mathscr{F}_{\mathscr{E}}(Y)$, the entries $\iota_x \sigma \pi_y$ of the $|X| \times |Y|$-matrix M_σ may all be considered as elements of $\mathscr{S}_{\mathscr{E}}$.[5] That means that all morphisms in $\mathscr{C}(\mathscr{E})$ can be written as matrices over the semiring $\mathscr{S}_{\mathscr{E}}$. Addition and multiplication of matrices correspond to addition and composition of morphisms, as stated in Property 3 above. Conversely, any $|X| \times |Y|$-matrix over $\mathscr{S}_{\mathscr{E}}$ gives rise to a morphism $\sigma: \mathscr{F}_{\mathscr{E}}(X) \to \mathscr{F}_{\mathscr{E}}(Y)$.

[5] If no order on X and Y is specified, we refer to $\iota_x \sigma \pi_y$ as the entry in the x-th row and the y-th column.

As an example, consider an arbitrary morphism $\gamma{:}\mathscr{F}_{\mathrm{AMH}}(y)\rightarrow\mathscr{F}_{\mathrm{AMH}}(y)$. Then there exist $a_0,\ldots,a_k\in\mathbf{N}$ such that $y\gamma=_{\mathrm{AMH}}a_0y+a_1h(y)+\cdots+a_kh^k(y)$, where multiplication of a term by an element of $\mathbf{N}$ stands for repeated addition of the term to itself. We associate with the morphism γ the polynomial $a_0+a_1X+\cdots+a_kX^k$, which is an element of semiring $\mathbf{N}[X]$ of polynomials in one indeterminate X with nonnegative integer coefficients.

The morphism $\sigma=[x_1/h(y_1),\ x_2/y_1+h^2(y_2)]$ from above and the morphism $\delta=[y_1/h(z),y_2/2z]$ can be expressed by the matrices

$$M_\sigma=\begin{pmatrix} X & 0\\ 1 & X^2\end{pmatrix}\quad\text{and}\quad M_\delta=\begin{pmatrix} X\\ 2\end{pmatrix}$$

over $\mathbf{N}[X]$. An easy calculation shows that the morphism $\sigma\delta=[x_1/h^2(z),\ x_2/h(z)+2h^2(z)]$ corresponds to the matrix $M_\sigma M_\delta$.

Example 5.1 The theories of Example 4.1 yield the following semirings (see [28, 3]).

$\mathscr{S}_{\mathrm{AMH}}$, the semiring corresponding to the theory AMH of Abelian monoids with a homomorphism, is isomorphic to $\mathbf{N}[X]$, the semiring of polynomials in one indeterminate X with nonnegative integer coefficients.

$\mathscr{S}_{\mathrm{AMIn}}$, which corresponds to the theory of Abelian monoids with an involution, is the monoid semiring $\mathbf{N}\langle Z_2\rangle$, where Z_2 denotes the cyclic group of order 2.

$\mathscr{S}_{\mathrm{COM}}$, the semiring corresponding to the theory COM, is isomorphic to $\mathbf{N}\langle X,Y\rangle$, the semiring of polynomials in two *noncommuting* indeterminates X,Y with nonnegative integer coefficients. Note that $\mathbf{N}\langle X,Y\rangle$ is the monoid semiring $\mathbf{N}\langle\{X,Y\}^*\rangle$, where $\{X,Y\}^*$ denotes the free monoid in two generators X,Y.

$\mathscr{S}_{\mathrm{GAUSS}}$ is isomorphic to the ring of Gaussian numbers $\mathbf{Z}\oplus i\mathbf{Z}$, consisting of the complex numbers $m+in$, where $m,n\in\mathbf{Z}$.

The first two examples suggest that there is a close connection between augmenting a commutative/monoidal theory by a monoid (as defined at the end of Subsect. 4.3) and adjoining a monoid to the corresponding semiring (as defined in Subsect. 3.3). For AMIn $=$ AM$\langle Z_2\rangle$, for instance, one verifies that the semirings $\mathscr{S}_{\mathrm{AM}\langle Z_2\rangle}$ and $\mathscr{S}_{\mathrm{AM}}\langle Z_2\rangle$ are isomorphic. It is easy to see that this kind of connection holds in general.

Theorem 5.2 *Let $\mathscr{E}$ be a commutative/monoidal theory, and let H be a finitely generated monoid. Then $\mathscr{S}_{\mathscr{E}\langle H\rangle}$, the semiring corresponding to $\mathscr{E}$ augmented by H, and the monoid semiring $\mathscr{S}_{\mathscr{E}}\langle H\rangle$ are isomorphic.*

Proof. Let $\mathscr{E}=(\Sigma,E)$ be a commutative/monoidal theory and H be a monoid generated by the finitely many elements $h_1,\ldots,h_n$. Then $\mathscr{E}\langle H\rangle$ has the signature $\Sigma'=\Sigma\cup\{h_1,\ldots,h_n\}$.

We shall construct a semiring isomorphism that maps every element $\gamma\in\mathscr{S}_{\mathscr{E}\langle H\rangle}$ to an element $\hat{\gamma}\in\mathscr{S}_{\mathscr{E}}\langle H\rangle$. Recall that the elements of $\mathscr{S}_{\mathscr{E}\langle H\rangle}$ are the Σ'-homomorphisms from $\mathscr{F}_{\mathscr{E}\langle H\rangle}(\mathbf{1})$ to $\mathscr{F}_{\mathscr{E}\langle H\rangle}(\mathbf{1})$ where $\mathbf{1}=\{x\}$ is a singleton. Let γ be such a Σ'-homomorphism. Then γ is uniquely determined by the element $x\gamma$. Since all h_i are homomorphisms for "$+$" and commute with every homomorphism in Σ, we can assume without loss of generality that $x\gamma=_{\mathscr{E}\langle H\rangle}\sum_{i=1}^m h_{i1}(\ldots(h_{in_i}(t_i))\ldots)$ where the t_i's are Σ-terms. For every $i=1,\ldots,m$ let γ_i be the Σ-homomorphism from $\mathscr{F}_{\mathscr{E}}(\mathbf{1})$ to $\mathscr{F}_{\mathscr{E}}(\mathbf{1})$ defined by $x\gamma_i{:=}t_i$. Then we have $\gamma_i\in\mathscr{S}_{\mathscr{E}}$. We define $\hat{\gamma}$ as $\hat{\gamma}{:=}\sum_{i=1}^m\gamma_i\cdot h_{i1}\cdots h_{in_i}\in\mathscr{S}_{\mathscr{E}}\langle H\rangle$.

One can verify that the definition of $\hat{\gamma}$ does not depend on the particular presentation of γ and that the mapping "$\hat{}$" is bijective. Exploiting the fact that $h_1, \ldots, h_n$ are homomorphisms in $\mathscr{E}\langle H\rangle$ one shows that "$\hat{}$" is compatible with the semiring operations and hence is a semiring isomorphism. □

The isomorphisms of $\mathscr{S}_{\mathscr{E}\langle H\rangle}$ will be used in Sects. 7 and 8 to study the unification problem for $\mathscr{E}\langle H\rangle$ in an algebraic setting.

5.2 Unification without Constants

In Subsect. 3.2 we have seen that $\mathscr{E}$-unification can be reformulated as unification in the category $\mathscr{C}(\mathscr{E})$. A unification problem in $\mathscr{C}(\mathscr{E})$ is given by a pair of morphisms σ, τ, and a unifier is a morphism δ such that $\sigma\delta = \tau\delta$. If we translate the morphisms into matrices over $\mathscr{S}_\mathscr{E}$, this means that an $\mathscr{E}$-unifier corresponds to a matrix M over $\mathscr{S}_\mathscr{E}$ such that $M_\sigma M = M_\tau M$. This correspondence is used in [25, 28, 3] to characterize the unification types of commutative/monoidal theories by algebraic properties of the corresponding semirings.

Theorem 5.3 *A commutative/monoidal theory $\mathscr{E}$ is unitary w.r.t. unification without constants if, and only if, $\mathscr{S}_\mathscr{E}$ satisfies the following condition: for any pair M_1, M_2 of $m \times n$-matrices over $\mathscr{S}_\mathscr{E}$ the set*

$$\mathscr{U}(M_1, M_2) := \{x \in \mathscr{S}_\mathscr{E}^n \mid M_1 x = M_2 x\}$$

is a finitely generated right $\mathscr{S}_\mathscr{E}$-module.

If $\mathscr{U}(M_\sigma, M_\tau)$ is generated by $x_1, \ldots, x_k \in \mathscr{S}_\mathscr{E}^n$, then the matrix which has $x_1, \ldots, x_k$ as columns corresponds to a most general $\mathscr{E}$-unifier of σ and τ.

Since constant-free unification problems in commutative/monoidal theories are either unitary or of type zero [25, 1, 28], the theorem yields that the theory $\mathscr{E}$ is of type zero iff there exist matrices M_1, M_2 over $\mathscr{S}_\mathscr{E}$ such that the right $\mathscr{S}_\mathscr{E}$-module $\mathscr{U}(M_1, M_2)$ is not finitely generated. Using this characterization, it can be shown that the theories AMH and COM are of type zero (see [1, 3]). The theories AMIn and GAUSS are unitary w.r.t. unification without constants (see [1] for the first, and [28] for the second result).

5.3 Unification with Constants

Following [29], we also reformulate unification with constants as unification in $\mathscr{C}(\mathscr{E})$. To this end we view constants as special variables that are always substituted by themselves. Then a unification problem with constants from a finite set C gives rise to morphisms $\sigma, \tau : \mathscr{F}_\mathscr{E}(X \cup C) \to \mathscr{F}_\mathscr{E}(Y \cup C)$ with the property that $c\sigma = c\tau = c$ for all $c \in C$.[6] We say that such a morphism *respects constants.*

If σ respects constants, then the matrix M_σ has a special form:

$$M_\sigma = \begin{pmatrix} M_\sigma^h & M_\sigma^i \\ 0 & I \end{pmatrix},$$

[6] This idea first appeared in [1].

where M^h_σ is an $|X| \times |Y|$-matrix, M^i_σ is an $|X| \times |C|$-matrix, 0 is the $|C| \times |Y|$-matrix with all entries 0, and I is the $|C| \times |C|$-unit matrix. The 0-submatrix is due to the fact that σ does not substitute terms with variables for constant. The unit matrix expresses that σ maps any constant to itself.[7] If M_σ is composed from M^h_σ and M^i_σ like above, we write $M_\sigma = \langle M^h_\sigma, M^i_\sigma \rangle$. Obviously, any such matrix corresponds to a morphism that respects constants.

Consider the unification problem with constants given by σ, τ. A unifier of σ and τ is a morphism $\delta: \mathcal{F}_\mathcal{E}(Y \cup C) \rightarrow \mathcal{F}_\mathcal{E}(Z \cup C)$ that respects constants such that $\sigma\delta = \tau\delta$. Because of the correspondence between morphisms and matrices, unification of σ and τ is equivalent to finding a matrix $M = \langle M^h, M^i \rangle$ such that $M_\sigma M = M_\tau M$. Taking into account the particular shape of the matrices we obtain

$$M_\sigma M = \begin{pmatrix} M^h_\sigma & M^i_\sigma \\ 0 & I \end{pmatrix} \begin{pmatrix} M^h & M^i \\ 0 & I \end{pmatrix} = \begin{pmatrix} M^h_\sigma & M^h_\sigma M^i + M^i \\ 0 & I \end{pmatrix},$$

and an analogous representation of $M_\tau M$. Thus, $M_\sigma M = M_\tau M$ holds if, and only if,

$$M^h_\sigma M^h = M^h_\tau M^h \tag{1}$$

$$M^h_\sigma M^i + M^i_\sigma = M^h_\tau M^i + M^i_\tau \tag{2}$$

hold. Equation (1) means that the columns of M^h have to satisfy the homogeneous linear equation system $M^h_\sigma x = M^h_\tau x$, which is a problem that we already encountered when solving unification problems without constants. To translate equation (2), let a_c, b_c be the c-th column of M^i_σ, M^i_τ, respectively. Then (2) holds iff for every $c \in C$ the c-th column of M^i satisfies the equation

$$M^h_\sigma x + a_c = M^h_\tau x + b_c. \tag{3}$$

Now, we relate unifiers to solutions of linear equation systems. Let I_c denote the set of all solutions to (3). A set of vectors $K_c \subseteq I_c$ is a *cover* if I_c can be represented as the union of cosets

$$I_c = \bigcup_{g \in K_c} y + \mathcal{U}(M^h_\sigma, M^h_\tau).$$

A cover K_c represents I_c in the sense that for every $y \in I_c$ there is a $y' \in K_c$ and a $y'' \in \mathcal{U}(M^h_\sigma, M^h_\tau)$ such that $y = y' + y''$. Obviously, a cover always exists because I_c itself is a cover.

Suppose that M^h is a matrix whose columns generate $\mathcal{U}(M^h_\sigma, M^h_\tau)$, and that for every $c \in C$ we have a cover K_c of I_c. Then we can construct a set $\mathcal{M}$ of matrices that correspond to a complete set of unifiers: $\mathcal{M}$ consists of all matrices $\langle M^h, N \rangle$, where N is a $|Y| \times |C|$-matrix whose c-th column is an element of K_c. Note that this construction generalizes the unification method for AM described in [23]. Obviously, if each K_c is a singleton (finite), then $\mathcal{M}$ is a singleton (finite).

Conversely, one can also translate arbitrary inhomogeneous linear equation systems into unification problems for $\mathcal{E}$. Thus, we can characterize the type of unification with constants in algebraic terms.

Theorem 5.4 *Let $\mathcal{E}$ be a commutative/monoidal theory which is unitary w.r.t. unification without constants. Then $\mathcal{E}$ is unitary (finitary) w.r.t. unification with constants if,*

[7] The superscripts $\cdot^h$ and $\cdot^i$ are chosen to indicate that in unification problems M^h_σ and M^i_σ will give rise to homogeneous and inhomogeneous linear equations, respectively.

and only if, $\mathcal{S}_{\mathscr{E}}$ satisfies the following condition: for any pair M_1, M_2 of $m \times n$-matrices over $\mathcal{S}_{\mathscr{E}}$, and any pair $a, b \in \mathcal{S}_{\mathscr{E}}^m$ the set

$$\{x \in \mathcal{S}_{\mathscr{E}}^n \mid M_1 x + a = M_2 x + b\}$$

is a coset (finite union of cosets) of the right $\mathcal{S}_{\mathscr{E}}$-module $\mathscr{U}(M_1, M_2)$.

This characterization has been used to show that AMIn is finitary w.r.t. unification with constants. The theory GAUSS is even unitary w.r.t. unification with constants. This is due to the fact that $\mathcal{S}_{\text{GAUSS}} \simeq \mathbf{Z} \oplus i\mathbf{Z}$ is a ring, and not only a semiring. In fact, let $\mathcal{S}_{\mathscr{E}}$ be a ring, and let x_0 be an arbitrary solution of the equation $M_1 x + a = M_2 x + b$. We show that $\{x_0\}$ is a cover. For any solution y of the inhomogeneous equation system, the difference $y - x_0$ is a solution of the homogeneous equation system $M_1 x = M_2 x$. This shows that any solution y of the inhomogeneous equation is an element of the coset $x_0 + \mathscr{U}(M_1, M_2)$. Conversely, any element of this coset is a solution of the inhomogeneous equation.

5.4 The Unification Problem is Undecidable

The goal of this subsection is to show that there exists a commutative/monoidal theory $\mathscr{E}$ such that it is in general undecidable whether an $\mathscr{E}$-unification problem with constants, Γ, has a solution or not. For unification problems without constants, there is always the trivial solution that substitutes all variables by the idempotent constant of the commutative theory. On the semiring level, this corresponds to the fact that the zero vector is always a solution of a homogeneous linear equation. It is still undecidable, however, whether there exists a nontrivial solution. In order to show these undecidability results, we shall use the undecidability of the word problem for groups in the following (slightly modified) formulation.

Proposition 5.5 *There exists a finitely presented group with an undecidable word problem such that the only element of finite order is the identity.*

Actually, such a group was exhibited by Boone and Collins (see [32] for a description of the construction and a proof that this group has an undecidable word problem; the fact that any element of this group different from the identity is of infinite order was proved in [27]). Now, assume that G is such a (finitely presented) group, generated by Δ. For words u, v over Δ, we write $u =_G v$ iff u and v belong to the same equivalence class in the presentation of G. We shall reduce the word problem for G (*i.e.*, the question whether $u =_G v$ for given words u, v) to solvability of unification problems in the commutative/monoidal theory AM$\langle G \rangle$.

Lemma 5.6 *Let u and v be words over Δ. Then $u =_G v$ iff the* AM$\langle G \rangle$*-unification problem $\langle u(x) \doteq v(x) \rangle$ has a non-trivial solution.*

Proof. As shown above, unification modulo AM$\langle G \rangle$ corresponds to solving linear equations in the group semiring $\mathbf{N}\langle G \rangle$, where $\mathbf{N}$ is the semiring of nonnegative integers. The unification problem $\langle u(x) \doteq v(x) \rangle$ is translated into the equation $u \cdot x = v \cdot x$, which must be solved by an element $s \in \mathbf{N}\langle G \rangle$.

First, assume that $u =_G v$. Thus, u is equal to v in $\mathbf{N}\langle G \rangle$, which means that the unit of the semiring is a (obviously non-trivial) solution of $u \cdot x = v \cdot x$.

Now, assume that $u \cdot x = v \cdot x$ has a non-trivial solution $s \in \mathbf{N}\langle G \rangle$. Thus, $s = a_1 \cdot h_1 + \cdots + a_k \cdot h_k$ for $k \geqq 1$ positive integers $a_1, \ldots, a_k$ and $h_1, \ldots, h_k \in G$, and

$$a_1 \cdot uh_1 + \cdots + a_k \cdot uh_k = us = vs = a_1 \cdot vh_1 + \cdots + a_k \cdot vh_k.$$

This equality in $\mathbf{N}\langle G\rangle$ implies that any monom that occurs on the left-hand side must occur on the right-hand side as well, and vice versa. Let i_1 be an arbitrary index between 1 and k. For i_1 there exists in index i_2 such that $uh_{i_1} =_G vh_{i_2}$. Since G is a group, this equality can be rewritten to $h_{i_2} =_G v^{-1}vh_{i_1}$. Assume that index i_j is already defined. For this index, there exists an index i_{j+1} with $uh_{i_j} =_G vh_{j+1}$, *i.e.*, $h_{i_{j+1}} =_G v^{-1}uh_{i_j}$. Since there are only finitely many possible indices, there exists $j < l$ such that $i_j = i_l$.

Consequently, we obtain $h_{i_j} = h_{i_l} = (v^{-1}u)^{l-j}h_{i_j}$, which implies that $(v^{-1}u)^{l-j}$ is the unit of the group G. This shows that $v^{-1}u$ is of finite order in G. By our assumption on G, only the unit of G has finite order, and thus $u =_G v$. □

Since the word problem for G is undecidable, the lemma implies that the existence of a non-trivial solution for unification problems without constants modulo the monoidal theory $\mathrm{AM}\langle G\rangle$ is in general undecidable. For unification with constants, undecidability follows from the next lemma.

Lemma 5.7 *Let u and v be words over Δ, and let c be a free constant. Then $u =_G v$ iff the $\mathrm{AM}\langle G\rangle$-unification problem $\langle u(c) \doteq v(c)\rangle$ has a solution.*

Proof. As described in the previous subsection, unification modulo $\mathrm{AM}\langle G\rangle$ corresponds to solving linear equations in the group semiring $\mathbf{N}\langle G\rangle$. The unification problem $\langle u(c) \doteq v(c)\rangle$ is translated into the inhomogeneous equation (without variables) $u = v$. Obviously, this equation is solvable iff $u = v$ holds in $\mathbf{N}\langle G\rangle$, and this is the case of iff $u = v$ holds in G. □

Theorem 5.8 *There exists a commutative/monoidal theory $\mathscr{E}$ such that it is in general undecidable whether an $\mathscr{E}$-unification problem with constants has a solution or not. In addition, it is in general undecidable whether an $\mathscr{E}$-unification problem without constants has a non-trivial solution or not.*

6 Unification with Constant Restrictions

Baader and Schulz [5] showed that one can devise a unification algorithm for arbitrary combinations of disjoint theories if one is given algorithms that solve unification problems with free constants and so-called constant restrictions in the individual theories. In this section we extend our techniques to tackle such problems in commutative/monoidal theories. We show that they correspond to particular systems of inhomogeneous linear equations.

If C is a finite set of free constants, then a *constant restriction* over C is a family $\mathbf{Y} = (Y_c)_{c \in C}$ of finite sets of variables Y_c. A substitution δ *satisfies* $\mathbf{Y}$ if for each $c \in C$ the constant c does not occur in any term $y\delta$ with $y \in Y_c$. For a unification problem Γ and a constant restriction $\mathbf{Y}$ we denote with $U_{\mathscr{E}}(\Gamma, \mathbf{Y})$ the set of $\mathscr{E}$-unifiers of Γ satisfying $\mathbf{Y}$. Complete subsets of $U_{\mathscr{E}}(\Gamma, \mathbf{Y})$ are defined similarly as in the case of arbitrary unifiers. We say that $\mathscr{E}$ is finitary (unitary) for unification with constant restrictions if for any Γ and $\mathbf{Y}$ there is a complete subset of $U_{\mathscr{E}}(\Gamma, \mathbf{Y})$ which is finite (a singleton). The combination algorithm in [5] constructs complete sets of unifiers for a problem in a combined theory from complete sets of unifiers satisfying certain unification problems with constant restrictions in the individual theories. The resulting set is finite if the input sets are finite. Thus it is important to know whether a theory is finitary for unification with constant restrictions.

General Assumption. *In the following, we assume that $\mathscr{E}=(\Sigma,E)$ is a commutative/monoidal theory, C is a finite set of five constants, and* $\mathbf{Y}=(Y_c)_{c\in C}$ *is a constant restriction. By "+" we denote the associative-commutative binary operation of this theory, and by* 0 *the corresponding neutral element.*

In the categorical framework we consider terms as elements of free algebras and substitutions as morphisms. Hence, we do not distinguish between $\mathscr{E}$-equal terms and substitutions. This means that t, considered as an element of $\mathscr{F}_\varepsilon(Z\cup C)$, does not contain $c\in C$ if there is a term t' with $t={}_\varepsilon t'$ such that c does not occur in t', and that δ, considered as a morphism, satisfies $\mathbf{Y}$ iff there is a substitution δ' which satisfies $\mathbf{Y}$ in the sense defined above.

In unification problems we view free constants as special variables that are not moved. Thus, if we characterize the absence of variables in terms we cover also the case of free constants. Recall that for a variable x the canonical projection π_x is the morphism that keeps x fixed and maps every variable distinct from x to 0.

Lemma 6.1 *Let t be a Σ-term and $x_1,\ldots,x_n$ be the variables occurring in t. Then* $t={}_{\mathscr{E}}t\pi_{x_1}+\cdots+t\pi_{x_n}$.

Proof. The mapping $\pi:\mathscr{F}_{\mathscr{E}}(X)\to\mathscr{F}_{\mathscr{E}}(X)$ defined by $t\pi:=t_{x_1}+\cdots+t\pi_{x_n}$ is the sum of the projections π_{x_i}. Each π_{x_i} is a morphism. Because $\mathscr{E}$ is commutative/monoidal, π is also a morphism (see Subsect. 5.1). The morphism π satisfies $x_i\pi={}_{\mathscr{E}}x_i$ for all i, $i=1,\ldots,n$. Hence, we have that π is $\mathscr{E}$-equal to the identity morphism, which implies the claim. □

Lemma 6.2 *Let t be a Σ-term and x be a variable. Then $t\pi_x={}_{\mathscr{E}}0$ if, and only if, there exists a term t' with $t={}_{\mathscr{E}}t'$ such that x does not occur in t'.*

Proof. "$\Rightarrow$" If x does not occur in t, then we are done. Otherwise, let $\{x_0,\ldots,x_n\}$ be the set of variables occurring in t, where $x=x_0$. Define $t_i:=t\pi_{x_i}$, *i.e.*, t_i is obtained from t by replacing every variable other than x_i with 0 and keeping x_i in its place. Define $t':=t_1+\cdots+t_n$. Obviously, t' does not contain x. Moreover, we have

$$t'={}_{\mathscr{E}}0+t'={}_{\mathscr{E}}t\pi_{x_0}+t\pi_{x_1}+\cdots+t\pi_{x_n}={}_{\mathscr{E}}t,$$

where the last identity holds because of the preceding lemma.

"$\Leftarrow$" Suppose that $t={}_{\mathscr{E}}t'$ for some term t' that does not contain x. The term $t'\pi_x$ is obtained from t' by replacing every variable $x'\neq x$ by 0. Since x does not occur in t', it follows that $t\pi_x$ is a ground term. Since $\mathscr{E}$ is commutative/monoidal, this yields $t'\pi_x={}_{\mathscr{E}}0$, which implies $t\pi_x={}_{\mathscr{E}}0$. □

The next lemma shows that a morphism satisfies $\mathbf{Y}$ iff certain entries in the matrix M_δ are zero.

Lemma 6.3 *Let $\delta:\mathscr{F}_{\mathscr{E}}(Y\cup C)\to\mathscr{F}_{\mathscr{E}}(Z\cup C)$ be a morphism that respects constants and $M_\delta=(\delta_{uv})_{u\in Y\cup C,v\in Z\cup C}$ be the corresponding matrix. Then δ satisfies $(Y_c)_{c\in C}$ if, and only if, $\delta_{yc}=0$ for all $c\in C$* and *$y\in Y_c$.*

Proof. Recall that $\delta_{uv}=t_u\delta\pi_v$, *i.e.*, the entry at position uv of δ is given by the substitution $[u/u\delta\pi_v]$ (see Sect. 5).

Now, δ satisfies $(Y_c)_{c \in C}$, iff for all $c \in C$ and every $y \in Y_c$, $y\delta$ does not contain c, iff for all $c \in C$ and $y \in Y_c$ we have $y\delta\pi_c =_{\mathscr{E}} 0$ by Lemma 6.2, iff for all $c \in C$ and $y \in Y_c$ the morphism $\iota_y \delta \pi_c$ is $\mathscr{E}$-equal to the 0-morphism. □

Obviously, unification with constants is a special case of unification with constant restrictions because it corresponds to the case where $Y_c = \emptyset$ for every c. Thus, the "only if"-part of the following theorem is obvious.

Theorem 6.4 *The theory $\mathscr{E}$ is finitary for unification with constant restrictions if, and only if, $\mathscr{E}$ is finitary for unification with constants.*

In the rest of the section we assume that $\mathscr{E}$ is finitary for unification with constants. We shall prove Theorem 6.4 by showing how to compute a complete set of unifiers by solving an appropriate set of linear equation systems.

First, observe that $\mathscr{E}$ is also finitary w.r.t. unification without constants – because this is a special case of unification with constants – and thus, since the only possible types are unitary or zero (see [25, 28, 3]), even unitary w.r.t. unification without constants.

We consider a unification problem given by morphisms σ, $\tau: \mathscr{F}_{\mathscr{E}}(X \cup C) \to \mathscr{F}_{\mathscr{E}}(Y \cup C)$ that respect constants. We want to describe those matrices M that correspond to unifiers of σ and τ satisfying $\mathbf{Y}$. By the results of Sect. 5.3, such a matrix has the form $M = \langle M^h, M^i \rangle$ and the equations (1) $M^h_\sigma M^h = M^h_\tau M^h$ and (2) $M^h_\sigma M^i + M^i_\sigma = M^h_\tau M^i + M^i_\tau$ hold. Equation (1) means that each column of M^h satisfies

$$M^h_\sigma x = M^h_\tau x. \tag{4}$$

Equation (2) is equivalent to the requirement that for every $c \in C$ we have $M^h_\sigma m_c + a_c = M^h_\tau m_c + b_c$, where m_c, a_c, b_c is the c-th column of $M^i, M^i_\sigma, M^i_\tau$, respectively.

To describe the impact of the constant restriction $\mathbf{Y}$, we define for every $c \in C$ the matrix $N^{\mathbf{Y}}_c = (v_{yy'})_{y,y' \in Y}$ by $v_{yy'} := 1$ iff $y = y'$ and $y \in Y_c$, and $v_{yy'} := 0$ otherwise. Then the entries at position (y, c) in M are 0 for any $y \in Y_c$ if, and only if, $N^{\mathbf{Y}}_c m_c = 0$. Combined with equation (2), this yields that each column m_c of M^i must satisfy the equation

$$\begin{pmatrix} M^h_b \\ N^{\mathbf{Y}}_c \end{pmatrix} x + \begin{pmatrix} b_c \\ 0 \end{pmatrix} = \begin{pmatrix} M^h_\tau \\ 0 \end{pmatrix} x + \begin{pmatrix} d_c \\ 0 \end{pmatrix}. \tag{5}$$

Now, we construct a set of matrices that corresponds to a complete set of unifiers of σ and τ that satisfy $\mathbf{Y}$. Since $\mathscr{E}$ is unitary w.r.t. unification *without* constants, the set of solutions to the equation (4) is finitely generated. We construct M^h by choosing a finite set of generators and placing them as columns into a matrix.

Since $\mathscr{E}$ is finitary w.r.t. unification *with* constants, there is a family $\mathbf{K} = (K_c)_{c \in C}$, where K_c is a finite cover for the solutions of equation (5) (see Subsect. 5.3). We define $\mathscr{M}_{\mathbf{K}}$ as the set of all matrices of the form

$$\begin{pmatrix} M^h & N \\ 0 & 1 \end{pmatrix}$$

where N is a $|Y| \times |C|$-matrix whose c-th column is an element of K_c.[8]

[8] Note the similarity to the construction in Sect. 5.3.

Lemma 6.5 *The set of matrices $\mathscr{M}_{\mathrm{K}}$ corresponds to a complete set of unifiers of σ, τ satisfying the constant restriction* **Y**.

Proof. Obviously, all elements of $\mathscr{M}_{\mathrm{K}}$ correspond to unifiers of σ, τ satisfying **Y**. In order to show completeness, let η be an arbitrary unifier of σ, τ satisfying **Y**. We show that there exist matrices $M = \langle M^h, N \rangle \in \mathscr{M}_{\mathrm{K}}$ and $L = \langle L^h, L^i \rangle$ such that $M_\eta = ML$. Then $\eta = \delta\lambda$, where δ, λ are the morphisms corresponding to M and L.

Since η is a unifier, we have $M^h_\sigma M^h_\eta = M^h_\tau M^h_\eta$. This means that each column of M^h_η is a solution of equation (4). By construction, the columns of M_h generate all solutions to this equation. Therefore, each column of M^h_η can be represented as a linear combination of the columns of M^h. We construct a matrix L^h by placing into the z-th column the coefficients of such a representation of the z-th column of M^h_η. Then $M^h_\eta = M^h L^h$.

Now let $c \in C$. The c-th column e_c of M^i_η is a solution of the equation (5). Since K_c is a cover of the solutions to (5), there is a vector $n_c \in K_c$ such that $e_c = x_c + n_c$ for some solution x_c of the corresponding homogeneous equation system. This system is an extension of (4), and thus x_c is a solution to (4) as well. Again, since the columns of M^h generate all solutions to (4), there is a vector l_c such that $x_c = M^h l_c$, which implies that $e_c = M^h l_c + n_c$. Now, let N be the matrix whose c-th column is n_c and L^i be the matrix whose c-th column is l_c. Then we have $M^i_\eta = M^h L^i + N$. Since also $M^h_\eta = M^h L^h$, it follows that $M_\eta = \langle M^h_\eta, M^i_\eta \rangle = \langle M^h L^h, M^h L^i + N \rangle = ML$, and we have shown the claim. □

For any equational theory, an algorithm for general unification, *i.e.*, unification where terms may contain additional free function symbols, is obtainable from an algorithm for unification with constant restrictions [5]. It returns finite complete sets of general $\mathscr{E}$-unifiers if the intermediate problems with constant restrictions produced by it have finite complete sets of unifiers.

Corollary 6.6 *There is an algorithm for general $\mathscr{E}$-unification if, and only if, there is an algorithm to solve inhomogeneous linear equations over $\mathscr{S}_{\mathscr{E}}$. Moreover, $\mathscr{E}$ is finitary for general unification if, and only if, it is finitary for unification with constants.*

7 A Sufficient Condition for Unification Type Zero

In this section we shall generalize the "type zero" result for the theory AMH to a whole class of commutative/monoidal theories. This class will be defined by properties of the corresponding semiring. Before we can do that, we need one more notation.

Let $\mathscr{S}$ be a semiring which is not a ring. That means that the Abelian monoid $(\mathscr{S}, +, 0)$ is not a group, *i.e.*, there exists an element $p \in \mathscr{S}$ such that, for all $q \in \mathscr{S}$, we have $p + q \neq 0$. We shall call such an element p of $\mathscr{S}$ *non-invertible*. An element $s \in \mathscr{S}$ which has an inverse w.r.t. "$+$" is called *invertible*. For the semiring $\mathbf{N}$, all elements different from 0 are non-invertible. For the direct product $\mathbf{N} \times \mathbf{Z}$, an element (n, z) is invertible iff $n = 0$. Here are some trivial facts about invertible and non-invertible elements.

1. The elements $s_1, \ldots, s_k$ of $\mathscr{S}$ are invertible if, and only if, their sum $s_1 + \cdots + s_k$ is invertible.
2. The element $\sum_{h \in H} s_h \cdot h$ of the monoid semiring $\mathscr{S}\langle H \rangle$ is non-invertible if, and only, if there exists $h \in H$ such that s_h is non-invertible in $\mathscr{S}$. Thus, if $\mathscr{S}$ is not a ring, then $\mathscr{S}\langle H \rangle$ is not a ring for any monoid H.

Recall that the theory AMH corresponds to the semiring $\mathbf{N}[X]$ of polynomials in one indeterminate X with nonnegative integer coefficients. That means that we have a monoid semiring $\mathscr{S}\langle H\rangle$ where *all the nonzero elements* of $\mathscr{S}$ are non-invertible, and where the monoid H is the free monoid X^* in one generator. The "type zero" result for AMH can now be generalized to the case where $\mathscr{S}$ contains *at least one* non-invertible element.

Theorem 7.1 *Let $\mathscr{E}$ be a commutative/monoidal theory such that the corresponding semiring $\mathscr{S}_{\mathscr{E}}$ is isomorphic to a monoid semiring $\mathscr{S}\langle X^*\rangle$. If $\mathscr{S}$ is not a ring, i.e., if $\mathscr{S}$ contains at least one non-invertible element, then $\mathscr{E}$ is of unification type zero.*

As mentioned before the monoid semiring $\mathscr{S}\langle X^*\rangle$ is just the polynomial semiring $\mathscr{S}[X]$. The theorem is proved if we can find matrices M_σ, M_τ over $\mathscr{S}[X]$ such that the right $\mathscr{S}[X]$-module $\mathscr{U}(M_\sigma, M_\tau)$ is not finitely generated.

In the following we shall show that the 1×3-matrices $M_\sigma := (X, X, 0)$ and $M_\tau := (0, 1, X^2)$ have the required property. Thus we consider the homogeneous linear equation

$$X\cdot x_1 + X\cdot x_2 = x_2 + X^2\cdot x_3 \tag{6}$$

which has to be solved by a vector $L \in \mathscr{S}[X]^3$. If L is such a vector, we denote its components by $L^{(1)}$, $L^{(2)}$, $L^{(3)}$.

Let p be a non-invertible element in $\mathscr{S}$. Obviously, for any $n \geqq 1$, the vector L_n which consists of the components $L_n^{(1)} := p$, $L_n^{(2)} := pX + \cdots + pX^{n+1}$, $L_n^{(3)} := pX^n$ is a solution of (6).

Now assume that $\mathscr{U}(M_\sigma, M_\tau)$ is finitely generated, *i.e.*, there exist finitely many solutions $G_1, \ldots, G_m$ of (6) which generate all the solutions of (6). Let $n \geqq 1$ be arbitrary but fixed. Since L_n is a solution of (6) there exist $l_1, \ldots, l_m \in \mathscr{S}[X]$ such that

$$L_n = \sum_{i=1}^{m} G_i l_i. \tag{7}$$

If we consider (7) in the first component, we get $p = \sum_{i=1}^{m} G_i^{(1)} l_i$. For $i = 1, \ldots, m$, let $p_i \in \mathscr{S}$ be the constant coefficient of the polynomial $G_i^{(1)}$, and $h_i \in \mathscr{S}$ be the constant coefficient of l_i. The last equation implies that $p = \sum_{i=1}^{m} p_i h_i$. Since p is non-invertible, there exists some j with $1 \leqq j \leqq m$ such that $p_j h_j$ is non-invertible.

Lemma 7.2 *The polynomial $G_j^{(3)}$ is of degree at least n.*

Proof. Assume that the degree of $G_j^{(3)}$ is less than n. Since G_j is a solution of (6), we know that $G_j h_j$ is also a solution, that is,

$$X\cdot G_j^{(1)} h_j + X\cdot G_j^{(2)} h_j = G_j^{(2)} h_j + X^2\cdot G_j^{(3)} h_j. \tag{8}$$

The components of the solution $G_j h_j$ satisfy the following properties:

- The constant coefficient of the polynomial $G_j^{(1)} h_j$ is $e_1 := p_j h_j$. Thus we know by the choice of j that e_1 is non-invertible.
- The polynomial $G_j^{(2)} h_j$ has constant coefficient 0. This is an immediate consequence of the equation (8).
- All the coefficients of $G_j^{(3)} h_j$ are invertible. This can be seen by considering equation (7) in the third component, which yields $pX^n = \sum_{i=1}^{m} G_i^{(3)} l_i$. Since $G_j^{(3)} h_j$ contains only monomials of degree less than n, all these monomials vanish during the summation. Consequently, all the coefficients of these monomials have to be invertible.

From the fact that the coefficient of X in $X\cdot G_j^{(1)}h_j$ is e_1 and in $X\cdot G_j^{(2)}h_j$ is 0 we get by (8) that the coefficient of X in $G_j^{(2)}h_j+X^2\cdot G_j^{(3)}h_j$ is also e_1. Hence, the coefficient of X in $G_j^{(2)}h_j$ is e_1.

Starting with the fact the coefficient e_1 of X in $G_j^{(2)}h_j$ is non-invertible, we shall now deduce that the coefficient of X^2 in $G_j^{(2)}h_j$ is also non-invertible. Since the coefficient of X in $G_j^{(2)}h_j$ is e_1, the coefficient of X^2 in $X\cdot G_j^{(2)}h_j$ is also e_1. Thus the coefficient of X^2 on the left hand side of (8) is $e':=e_1+e$ for some e. The coefficient e' is non-invertible because otherwise e_1 could not be non-invertible. By (8), the coefficient of X^2 in $G_j^{(2)}h_j+X^2\cdot G_j^{(3)}h_j$ is also e'. Since all the coefficients of $X^2\cdot G_j^{(3)}h_j$ are invertible, this finally shows that the coefficient e_2 of X^2 in $G_j^{(2)}h_j$ is non-invertible.

This argument can be iterated to show that, for all $k\geqq 1$, the coefficient e_k of X^k in $G_j^{(2)}h_j$ is non-invertible. This is a contradiction to the fact that the polynomial $G_j^{(2)}h_j$ has only finitely many nonzero coefficients. □

We have just shown that, for any $n\geqq 1$, there exists a j such that $G_j^{(3)}$ is of degree at last n. This is a contradiction to our assumption that there are finitely many generators G_j of all solutions of (6). This completes the proof of the theorem.

8 Adding Finite Monoids of Homomorphisms

We now investigate commutative/monoidal theories that are augmented with finite monoids of homomorphisms. In contrast to the case of free monoids, which was treated in the previous section, we can derive the positive result that adding finite monoids does not change the unification type and that algorithms for the original theory can be used to solve problems in the augmented theory.

An example for such a theory is AMIn, the theory of Abelian monoids with an involution. Recall that AMIn can be written as AM$\langle Z_2\rangle$, and that the corresponding semiring is $\mathrm{N}\langle Z_2\rangle$.

General Assumption. *In this section $\mathscr{E}$ is a commutative/monoidal theory and H is a finite monoid.*

Since unification problems in $\mathscr{E}\langle H\rangle$ are equivalent to systems of linear equations over $\mathscr{S}_{\mathscr{E}}\langle H\rangle$, our basic technique will be to reduce such systems to systems of linear equations over $\mathscr{S}_{\mathscr{E}}$. As a first step we shall establish a one-to-one correspondence between vectors.

Every vector $x\in\mathscr{S}_{\mathscr{E}}\langle H\rangle^n$ has a unique representation as $x=\sum_{h\in H}x_h\cdot h$ where $x_h\in\mathscr{S}_{\mathscr{E}}^n$. As an example the vector

$$x=\begin{pmatrix}1+2h\\ h\end{pmatrix}\in\mathrm{N}\langle Z_2\rangle$$

can be written as

$$x=\begin{pmatrix}1\cdot e+2\cdot h\\ 0\cdot e+1\cdot h\end{pmatrix}=\begin{pmatrix}1\\0\end{pmatrix}\cdot e+\begin{pmatrix}2\\1\end{pmatrix}\cdot h.$$

We can formally justify this notation if we consider $\mathscr{S}_{\mathscr{E}}$ and H as subsets of $\mathscr{S}_{\mathscr{E}}\langle H\rangle$, This can be done by identifying every element $s\in\mathscr{S}_{\mathscr{E}}$ with $s\cdot e\in\mathscr{S}_{\mathscr{E}}\langle H\rangle$, where e is the unit in H, and every element $h\in H$ with $1\cdot h\in\mathscr{S}_{\mathscr{E}}\langle H\rangle$.

Suppose the elements of H are numbered as $h_1, \ldots, h_{|H|}$. If $x \in \mathcal{S}_{\mathcal{E}}\langle H\rangle^n$ has a representation as $x = x_{h_1} \cdot h_1 + \cdots + x_{h_{|H|}} \cdot h_{|H|}$, we define

$$\hat{x} = \begin{pmatrix} x_{h_1} \\ \vdots \\ x_{h_{|H|}} \end{pmatrix} \in \mathcal{S}_{\mathcal{E}}^{n|H|}$$

as the vector obtained from x by writing the vectors x_h one below another. Continuing our example from above we have

$$\hat{x} = \begin{pmatrix} 1 \\ 0 \\ 2 \\ 1 \end{pmatrix}.$$

We thus obtain a bijection between $\mathcal{S}_{\mathcal{E}}\langle H\rangle^n$ and $\mathcal{S}_{\mathcal{E}}^{n|H|}$. In particular, every vector in $\mathcal{S}_{\mathcal{E}}^{n|H|}$ has a representation as $\hat{x}$ for some $x \in \mathcal{S}_{\mathcal{E}}\langle H\rangle^n$. Obviously, for all $x, y \in \mathcal{S}_{\mathcal{E}}^{n|H|}$ and all $s \in \mathcal{S}_{\mathcal{E}}$ we have

$$\widehat{x+y} = \hat{x} + \hat{y} \quad \text{and} \quad \hat{x} \cdot s = \widehat{x \cdot s}. \tag{9}$$

In algebraic terms we can rephrase these equalities by saying that the mapping “$\hat{\cdot}$” is a right $\mathcal{S}_{\mathcal{E}}$-module isomorphism.

Next we will associate to every $m \times n$-matrix M with entries in $\mathcal{S}_{\mathcal{E}}\langle H\rangle$ an $m|H| \times n|H|$-matrix $\hat{M}$ with entries in $\mathcal{S}_{\mathcal{E}}$, such that $\widehat{Mx} = \hat{M}\hat{x}$ holds for every $x \in S_E\langle H\rangle^n$. To derive an appropriate definiton of $\hat{M}$, observe that, similar to a vector, the matrix M has a unique representation $M = \sum_{h\in H} M_h \cdot h$, where the M_h are matrices with entries in $\mathcal{S}_{\mathcal{E}}$. Applying M to a vector x yields

$$\begin{aligned} \widehat{Mx} &= \left(\sum_{f\in H} M_f \cdot f\right)\left(\sum_{g\in H} x_g \cdot g\right) = \sum_{f,g\in H} M_f x_g \cdot f \cdot g \\ &= \sum_{h\in H}\left(\sum_{h=f\cdot g} M_f x_g\right)\cdot h = \sum_{h\in H}\left(\sum_{g\in H}\left(\sum_{h=f\cdot g} M_f\right) x_g\right)\cdot h. \end{aligned}$$

This series of equalities says that the component of the vector $\widehat{Mx}$ corresponding to the element h is obtained by summing over all g the products $(\sum_{h=f\cdot g} M_f) x_g$. This shows that we have to define $\hat{M}$ as the $m|H| \times n|H|$-matrix consisting of the submatrices

$$\hat{M}_{i,j} = \sum_{\substack{h\in H \\ h_i = h}} M_h,$$

where a sum over an empty set of indices is to be understood as the zero matrix. With this definition we obtain

$$\widehat{Ma} = \hat{M}\hat{a}. \tag{10}$$

Returning to our example theory AMIn, consider a matrix M over $\mathbb{N}\langle Z_2\rangle$. If $M = M_e \cdot e + M_h \cdot h$, then the associated matrix is

$$\hat{M} = \begin{pmatrix} M_e & M_h \\ M_h & M_e \end{pmatrix}.$$

Thus, our general approach gives us the same representation of unification problems in AMIn as the one derived in [1].

Next we apply our transformation technique to unification problems without constants.

Proposition 8.1. *Let M_1, M_2 be* m × *n-matrices over $\mathscr{S}_{\mathscr{E}}\langle H\rangle$, and* $x\in\mathscr{S}_{\mathscr{E}}\langle H\rangle^n$. *Then:*

1. $x\in\mathscr{U}(M_1,M_2)$ *if, and only if* $\hat{x}\in\mathscr{U}(\hat{M}_1,\hat{M}_2)$
2. $\mathscr{U}(M_1,M_2)$ *is generated by* $x_1,\dots,x_k$ *if* $\mathscr{U}(\hat{M}_1,\hat{M}_2)$ *is generated by* $\hat{x}_1,\dots,\hat{x}_k$.

Proof. 1. Let $x = \mathscr{S}_{\mathscr{E}}\langle H\rangle^n$. Then we have $x\in\mathscr{U}(M_1,M_2)$ if and only if $M_1x = M_2x$ if and only if $\widehat{M_1x} = \widehat{M_2x}$ if and only if $\hat{M}_\sigma\hat{x} = \hat{M}_\tau\hat{x}$ if and only if $\hat{x}\in\mathscr{U}(\hat{M}_1,\hat{M}_2)$.

2. It suffices to show that every $x\in\mathscr{U}(M_1,M_2)$ is a linear combination of $x_1,\dots,x_k$. If $x\in\mathscr{U}(M_1,M_2)$, then $\hat{x}\in\mathscr{U}(\hat{M}_1,\hat{M}_2)$ by part (1). Hence, $\hat{x} = \hat{x}_1\cdot s_1 + \cdots + \hat{x}_k\cdot s_k$. Using equalities (9), we conclude that $x = x_1\cdot s_1 + \cdots + x_k\cdot s_k$, Thus, x is a linear combination of $x_1,\dots,x_k$. □

If $\mathscr{E}$ is unitary w.r.t. unification without constants, then for all matrices M_1, M_2 with entries from $\mathscr{S}_{\mathscr{E}}\langle H\rangle$ the right $\mathscr{S}_{\mathscr{E}}$-module $\mathscr{U}(\hat{M}_1,\hat{M}_2)$ is finitely generated, and by the preceding proposition, $\mathscr{U}(M_1,M_2)$ is finitely generated. Together with Theorem 5.3 this proves our next theorem.

Theorem 8.2 *If $\mathscr{E}$ is unitary w.r.t. unification without constants, then $\mathscr{E}\langle H\rangle$ is unitary w.r.t. unification without constants.*

The approach to unification problems with constants again consists in reducing a problem for $\mathscr{E}\langle H\rangle$ to a problem for $\mathscr{E}$. Speaking in terms of semirings, we shall reduce inhomogeneous linear equations over $\mathscr{S}_{\mathscr{E}}\langle H\rangle$ to inhomogeneous linear equations over $\mathscr{S}_{\mathscr{E}}$.

For a set $S\subseteq\mathscr{S}_{\mathscr{E}}\langle H\rangle^n$ let $\hat{S}:=\{\hat{x}\mid x\in S\}$.

Proposition 8.3 *Let M_1, M_2 be $m\times n$-matrices with entries in $\mathscr{S}_{\mathscr{E}}\langle H\rangle$ and a, $b\in\mathscr{S}_{\mathscr{E}}\langle H\rangle^m$. Let $N:=\{x\in\mathscr{S}_{\mathscr{E}}\langle H\rangle^n\mid M_1x + a = M_2x + b\}$. Then*:

1. $\hat{N} = \{y\in\mathscr{S}_{\mathscr{E}}^{n|H|}\mid \hat{M}_\sigma y + \hat{a} = \hat{M}_\tau y + \hat{b}\}$
2. N is a coset (finite union of cosets) of $\mathscr{U}(M_1,M_2)$, if $\hat{N}$ is a coset (finite union of cosets) of $\mathscr{U}(\hat{M}_1,\hat{M}_2)$.

Proof. 1. By equalities (9) and (10) it follows that for all $x\in\mathscr{S}_{\mathscr{E}}\langle H\rangle^n$ we have $M_1x + a = M_2x + b$ if and only if $\hat{M}_\sigma\hat{x} + \hat{a} = \hat{M}_\tau\hat{x} + \hat{b}$. Since for every $y\in\mathscr{S}_{\mathscr{E}}^{n|H|}$ there is a unique $x\in\mathscr{S}_{\mathscr{E}}\langle H\rangle^n$ such that $y = \hat{x}$, this yields the claim.

2. If $\hat{N}$ is a coset of $\mathscr{U}(\hat{M}_1,\hat{M}_2)$, then there exists a vector $x\in\mathscr{S}_{\mathscr{E}}\langle H\rangle^n$ such that $\hat{N} = \{\hat{x} + y\mid y\in\mathscr{U}(\hat{M}_1,\hat{M}_2)\}$. Using equality (9) and Proposition 8.1 we conclude that $N = \{x + z\mid z\in\mathscr{U}(M_1,M_2)\}$.

For the case that $\hat{N}$ is a finite union of cosets, the argument has to be slightly generalized. □

By Theorem 5.4, the preceding result gives us a condition for $\mathscr{E}\langle H\rangle$ to be unitary or finitary.

Theorem 8.4 *Suppose $\mathscr{E}$ is unitary w.r.t. unification without constants. If $\mathscr{E}$ is unitary (finitary) w.r.t. unification with constants, then $\mathscr{E}\langle H\rangle$ is unitary (finitary) w.r.t. unification with constants, if $\mathscr{E}$ is unitary (finitary) w.r.t. unification with constants.*

Propositions 8.1 and 8.3 tell us how we can use an algorithm for $\mathscr{E}$ to solve problems in $\mathscr{E}\langle H\rangle$. An $\mathscr{E}\langle H\rangle$-unification problem without constants is given by $m\times n$-matrices M_1, M_2 with entries in $\mathscr{S}_{\mathscr{E}\langle H\rangle}\simeq\mathscr{S}_{\mathscr{E}}\langle H\rangle$. We compute the transforms

$\hat{M}_\sigma$ and $\hat{M}_\tau$ and solve the equation $\hat{M}_\sigma y = \hat{M}_{\tau y}$ over $\mathcal{S}_{\mathcal{E}}$, which we can do with the algorithm for $\mathcal{E}$. If the set of solutions of the matrix equation over $\mathcal{S}_{\mathcal{E}}$ is generated by vectors $y_1, \dots, y_k \in \mathcal{S}_{\mathcal{E}}^{n|H|}$, we compute $x_1, \dots, x_k \in \mathcal{S}_{\mathcal{E}}\langle H \rangle^n$ such that $\hat{x}_i = y_i$. Then the set of solutions of the original equation is generated by $x_1, \dots, x_k$ and the matrix M_δ that has $x_1, \dots, x_k$ as columns represents a most general unifier of the given problem.

Since inhomogeneous linear equations over $\mathcal{S}_{\mathcal{E}\langle H \rangle} \simeq \mathcal{S}_{\mathcal{E}}\langle H \rangle$ can be transformed into inhomogeneous equations over $\mathcal{S}_{\mathcal{E}}$, an algorithm for $\mathcal{E}$ can be used in a similar way as in the constant free case to solve unification problems with constants in $\mathcal{E}\langle H \rangle$.

9 Conclusion

Two approaches to solving unification problems can be distinguished. The first, which might be called the "syntactic approach," relies heavily on the syntactic structure of the identities that define the equational theory (see for instance [14, 26, 20]). The second, which we may characterize as the "semantic approach," exploits the structure of the algebras that satisfy the theory. If little or nothing is known of the algebras involved, the first approach is useful, whereas the second is applicable to theories that describe algebraic structures which have been investigated in mathematics.

With this paper we pursue the semantic approach to unification. We have combined techniques for commutative and monoidal theories that had been developed independently. We have shown that both classes of theories are essentially the same in that every monoidal theory is commutative, and every commutative theory can be turned into a monoidal theory by a signature transformation.

One of the major topics of research in unification in recent years was to construct algorithms for the combination of equational theories. This problem has been solved – at least in principle – for theories with disjoint signatures [33, 5]. The result in Sect. 6 show that the combination method developed in [5] can always be applied for finitary commutative/monoidal theories. Of course, the case where signatures are not disjoint is too difficult to be treated in full generality. We concentrated on a special case, namely the combination of a commutative/ monoidal theory with a monoid of homomorphisms. By exploiting the algebraic structure of the canonical semiring associated to such a theory, we have found combinations that are of unification type zero, and others that are of type unitary or finitary. For the latter case we have pointed out how a unification algorithm can be derived.

There still remain open questions for this kind of combination. We have augmented a given theory either by free monoids or by finite monoids, but we do not know what happens with infinite monoids that are not free.

The only commutative/monoidal theories of unification type zero that we know are those described in this paper. They all have canonical semirings that are not rings. It would be interesting to know whether there exist theories of unification type zero for which the canonical semiring is a ring. Since every semiring can be obtained from a commutative/monoidal theory this question can be posed in purely algebraic terms: is there a ring such that the set of solutions for some system of homogeneous linear equations is not finitely generated?

It is not known whether (1) there exists a unitary or finitary equational theory that is infinitary or of type zero for unification w.r.t. constants, or whether (2) there exists a theory that is unitary or finitary for unification w.r.t. constants, but infinitary or of type zero for general unification or unification with constant restrictions. This question has been raised in the context of combining theories with disjoint signatures. We have shown that the second case cannot occur for commutative/monoidal theories. The first question can be reformulated for commutative/monoidal theories as an algebraic problem: does there exist a semiring such that for every system of homogeneous equations the set of solutions is a finitely generated right module, but there is a system of inhomogeneous equations such that the corresponding set of solutions is not a finite union of cosets? Given the substantial body of results in linear algebra, it is conceivable to find a semiring satisfying this condition. Such a semiring would then give us an example of an equational theory with the above property.

References

1. Baader, F.: Unification in Commutative Theories. J. Symb. Computation **8**, 479–497 (1989)
2. Baader, F.: Unification Properties of Commutative Theories: A Categorical Treatment. In: Pitt, D.H., Rydeheard, D.E., Dybjer, P., Pitts, A.M., Poigné, A. (eds.) Proceedings of the Conference on Category Theory and Computer Science. Lecture Notes in Computer Science, Vol. 389. Berlin, Heidelberg, New York: Springer 1989
3. Baader, F.: Unification in Commutative Theories, Hilbert's Basis Theorem, and Gröbner Bases. J. ACM **40**, 477–503 (1993)
4. Baader, F., Nutt, W.: Adding Homomorphisms to Commutative/Monoidal Theories or How Algebra Can Help in Equational Unification. In: Book, R. (ed.) Proceedings of the 4th International Conference on Rewriting Techniques and Applications. Lecture Notes in Computer Science, Vol 488. Berlin, Heidelberg, New York: Springer 1991
5. Baader, F., Schulz, K. U.: Unification in the Union of Disjoint Equational Theories: Combining Decision Procedures. In: Kapur, D. (ed.) Proceedings of the 11th International Conference on Automated Deduction. Lecture Notes in Computer Science, Vol. 607. Berlin, Heidelberg, New York: Springer 1992
6. Baader, F., Siekmann, J. H.: Unification Theory. In: Gabbay, D. M., Hogger, C. J., Robinson, J. A. (eds.) Handbook of Logic in Artificial Intelligence and Logic Programming. Oxford: Oxford University Press 1994
7. Bachmair, L.: Canonical Equational Proofs. Boston, Basel, Berlin: Birkhäuser 1991
8. Cohn, P. M.: Universal Algebra. New York: Harper and Row 1965
9. Comon, H., Haberstrau, M., Jouannaud, J.-P.: Decidable Problems in Shallow Equational Theories. In: Scedrov, A. (ed.) Proceedings of the 7th Annual IEEE Symposium on Logic in Computer Science. Washington D.C.: IEEE Computer Society Press 1992
10. Davis, M.: Unsolvable Problems. In: Barwise, J. (ed.) Handbook of Mathematical Logic. Amsterdam: Elsevier Science Publishers 1977
11. Fay, M.: First-order Unification in an Equational Theory. In: Proceedings 4th Workshop on Automated Deduction 1979
12. Fitting, M.: Basic Modal Logic. In: Gabbay, D. M., Hogger, C. J., Robinson, J. A. (eds.) Handbook of Logic in Artificial Intelligence and Logic Programming. Oxford: Oxford University Press 1993
13. Gallier, J., Raatz, S.: SLD-Resolution Methods for Horn Clauses with Equality Based on *E*-Unification. In: Keller, R. M. (ed.) Proceedings of the 3rd IEEE Symposium on Logic Programming. Washington D.C.: IEEE Computer Society Press 1986
14. Gallier, J., Snyder, S.: Complete Sets of Transformations for General *E*-Unification. Theor. Comput. Sci. **27**, 203–260 (1989)
15. Grätzer, G.: Universal Algebra. Princeton: Van Nostrand 1968
16. Herrlich, H., Strecker, G. E.: Category Theory. Boston: Allyn and Bacon 1973
17. Huet, G.: Confluent Reductions: Abstract Properties and Applications to Term Rewriting Systems. J. ACM **27**, 797–821 (1980)

18. Jaffar, J., Lassez, J. L., Maher, M.: A Theory of Complete Logic Programs with Equality. J. Logic Programming **1**, 211–224 (1984)
19. Jouannaud, J. P., Kirchner, H.: Completion of a Set of Rules Modulo a Set of Equations. SIAM J. Comp. **15**, 1155–1194 (1986)
20. Kirchner, C., Klay, F.: Syntactic Theories and Unification. In: Mitchell, J. (ed.) Proceedings of the 5th Annual IEEE Symposium on Logic in Computer Science. Washington D.C.: IEEE Computer Society Press 1990
21. Lawvere, F. W.: Functional Semantics of Algebraic Theories. Ph.D. Thesis, Columbia University, New York (1963)
22. Lemmon, E. J.: Algebraic Semantics for Modal Logics I. J. Symbolic Logic **31**, 46–65 (1966)
23. Livesey, M., Siekmann, J.: Unification of Bags and Sets. SEKI-MEMO 76-II, Institut für Informatik I, Universität Karlsruhe (1976)
24. Nevins, A. J.: A Human Oriented Logic for Automated Theorem Proving. J. ACM **21**, 606–621 (1974)
25. Nutt, W.: Unification in Monoidal Theories, Presentation at the Second Workshop on Unification. Val d'Ajol, France, (1988)
26. Nutt, W., Réty, P., Smolka, G.: Basic Narrowing Revisited. J. Symb. Computation **7**, 295–317 (1989)
27. Nutt, W.: The Unification Hierarchy is Undecidable. J. Automated Reasoning **7**, 369–381 (1991)
28. Nutt, W.: Unification in Monoidal Theories. In: Stickel, M. (ed.) Proceedings 10th International Conference on Automated Deduction. Lecture Notes in Computer Science, Vol. 499. Berlin, Heidelberg, New York: Springer 1990
29. Nutt, W.: Unification in Monoidal Theories is Solving Linear Equations over Semirings. Technical Report RR-92-01, DFKI Saarbrücken (1992)
30. Peterson, G., Stickel, M.: Complete Sets of Reductions for Some Equational Theories. J. ACM **28**, 233–264 (1981)
31. Plotkin, G.: Building in Equational Theories. Machine Intelligence **7**, 73–90 (1972)
32. Rotman, J. J.: The Theory of Groups. Boston: Allyn and Bacon 1973
33. Schmidt-Schauß, M.: Combination of Unification Algorithms. J. Symbolic Computation **8**, 51–99 (1989)
34. Slagle, J. R.: Automated Theorem Proving for Theories with Simplifiers, Commutativity and Associativity. J. ACM **21**, 622–642 (1974)
35. Stickel, M.: Automated Deduction by Theory Resolution. J. Automated Reasoning **1**, 333–355 (1985)
36. Taylor, W.: Equational Logic. Houston J. Math. **5**, 1–51 (1979)