AN INTRODUCTION TO TAME CONGRUENCE THEORY

EMIL W. KISS
Eötvös University
Department of Algebra and Number Theory
1088 Budapest, Múzeum krt. 6–8
Hungary

Abstract. In the past twenty years one could witness a nice and fruitful interaction between two special areas of algebra and model theory. An important problem in each of these areas has been solved by methods coming from the other territory. To achieve this end, a deep theory, called tame congruence theory has been developed, which has lots of other applications as well. The aim of this paper is to give a non-technical introduction to this topic for people having a general background in mathematics. We approach the subject by presenting the two problems first, and then give a historical account of the process of how these problems were solved in a broader and broader context.

1. Introduction

Here is how this paper is organized. First we describe briefly the broad outline of tame congruence theory in Section 2. The two problems mentioned in the Abstract will be introduced in Section 3. In the next three sections we look at the classical results on these problems. This discussion helps to understand the problems themselves, and at the same time we shall be able to introduce some basic concepts (like boolean products, discriminator varieties, abelian algebras, the modular commutator) for the benefit of those who do not yet have much experience with general algebraic models. Thus the paper can be read with only a basic knowledge of algebra. These concepts will enable us to formulate the results obtained using tame congruence theory, in later sections.

Sections 7 and 8 introduce tame congruence theory in more depth. In Section 9 we present our first application, which has as a corollary the

Givant-Palyutin Theorem characterizing ω-categorical equational classes. In Section 10 we give an overview of the latest results concerning the two problems. The paper is concluded with a guide describing references for people who want to learn more about these topics.

The author is greatly indebted to K. Kearnes, Á. Szendrei, M. Valeriote, and R. Willard for numerous suggestions that helped to improve the paper.

2. What is tame congruence theory?

We focus on algebraic models, or algebras for short. These are models having no relation symbols, and so are given by an underlying set, and a system of finitary basic operations. Tame congruence theory has been developed by Ralph McKenzie and David Hobby. It is a deep structure theory of finite algebras. The main idea is to first understand these algebras *locally*, and then to put the information obtained together to get their global structure.

During this localization process we shall concentrate on the polynomial structure of an algebra $\mathbf{A}$. To define the concept of a polynomial, recall that a term of an algebra is any composition of its basic operations. If we allow also constant unary functions while forming these compositions, we get the polynomials of $\mathbf{A}$. Thus the polynomials of $\mathbf{A}$ are the operations of the form $p(x_1, \dots, x_n) = t(x_1, \dots, x_m, \mathbf{a})$ where t is a term of $\mathbf{A}$ and $\mathbf{a}$ is a sequence of parameters from $\mathbf{A}$.

Next consider any subset $N \subseteq A$. In order to localize to N, we create a new algebra on N by letting its basic operations be the restrictions of those polynomials p of $\mathbf{A}$ (of arbitrary arity) that can be restricted to N, that is, which satisfy $p(N, \dots, N) \subseteq N$. This new algebra is called the *induced algebra on* N and is denoted by $\mathbf{A}|_N$.

We wish to approximate $\mathbf{A}$ by the induced algebras on its subsets. However, not every subset works equally well. It may happen that there are very few polynomials that can be restricted to a given subset N, and therefore the induced algebra on N will not reflect the structure of $\mathbf{A}$ locally at all. How can we find "good" subsets for local approximations?

Call a unary polynomial e of $\mathbf{A}$ *idempotent* if it satisfies $e(e(a)) = e(a)$ for every element a of A. That is, e is the identity function on its range $e(A)$. The ranges of the idempotent non-constant unary polynomials of $\mathbf{A}$ are called the *neighborhoods* of $\mathbf{A}$.

The induced algebras on the neighborhoods already reflect the local polynomial structure of $\mathbf{A}$, and are therefore suitable for localization. Indeed, let the neighborhood N be the range of an idempotent unary polynomial e. If p is any polynomial of $\mathbf{A}$, then the composition ep can be restricted to N, and hence this restriction is a basic operation of $\mathbf{A}|_N$. On the other hand, e is the identity map on N, so the action of p and ep on N is

the same. Thus we do not lose information about how polynomials behave on N when we restrict to N.

Although the localization process works for any algebra, the theory we describe is for finite algebras. One advantage of finiteness is that it guarantees the existence of minimal neighborhoods. A most surprising fact is that *the induced algebras on minimal (finite) neighborhoods can be completely classified.* To make our presentation simpler we assume for the rest of this section that $\mathbf{A}$ is a finite simple algebra (simplicity means that it has only injective or constant homomorphisms into any other algebra). The induced algebra on any minimal neighborhood must be one of the following five types:

(1) a primitive permutation group;
(2) a one-dimensional vector space;
(3) the two-element boolean algebra;
(4) the two-element lattice;
(5) the two-element semilattice.

The two-element semilattice is the algebra having the meet operation of the two-element boolean algebra as its single basic operation. A permutation group acting on a set N is considered as an algebra by adding to N the permutations induced by the group elements, as unary operations. The precise statement is that if N is a minimal neighborhood, then the induced algebra $\mathbf{A}|_N$ is *polynomially equivalent* to one of the algebras listed above, that is, they have the same set of polynomials.

Each finite simple algebra $\mathbf{A}$ must also be *uniform* in the sense that the induced algebras on any two minimal neighborhoods are isomorphic. This allows us to assign one of the types **1**, **2**, **3**, **4**, and **5** to $\mathbf{A}$. One can show that in many situations $\mathbf{A}$ behaves similarly as the minimal algebra corresponding to its type. The reason for this is not just the fact that the local approximating algebras have a known structure, and so we can calculate in an effective way locally. We must be able to relate the local and global polynomial structure of the algebra. There are two important properties that help in doing so.

First, if any minimal neighborhood N is given, then any two different elements of $\mathbf{A}$ can be separated by a unary polynomial mapping into N. This property is called *separation*, and, in a sense, it embeds $\mathbf{A}$ into a power of N. This property allows us to transfer many global properties of $\mathbf{A}$ to local properties.

Second, any two elements of A can be connected by a sequence of overlapping minimal neighborhoods. This property is called *connectedness*. The fact that we can calculate well in each link of the chain often is sufficient to transfer local properties back into global properties.

The system of minimal neighborhoods forms a kind of geometry on **A**. The "points" of the geometry are the elements of **A**, and the one-dimensional subspaces or "lines" are the minimal neighborhoods. This geometry usually has higher dimensional "subspaces", which are neighborhoods with a tight induced structure. Another important example of approximating with non-minimal neighborhoods is when the algebra in question is not simple. In this case we describe the structure of those neighborhoods that are minimal only in a certain sense. We shall explain these methods, and will give more details and examples in Sections 7–9.

3. Two problems

Since Gödel's famous work on the undecidability of arithmetic it was clear that classes of models having a rich enough structure must have an undecidable first order theory. For example, the first order theory of abelian groups is decidable. On the other hand, if a class of groups is defined by group identities, and contains a nonabelian finite group, then its first order theory is undecidable. To substantiate this feeling, it is a natural idea to try to determine all classes of algebras that have a decidable first order theory, and, if possible, the structure of the members of these classes. To avoid complications, it is usual to assume that the algebras considered are of finite signature, that is, they have finitely many basic operations.

The most natural classes of algebras to look at are the ones defined by sets of identities. These are called equational classes, or *varieties.* Varieties can be characterized algebraically: by Birkhoff's famous theorem a class of algebras is a variety if and only if it is closed under the formation of direct products, subalgebras, and homomorphic images. So the first problem, more precisely stated, is this.

PROBLEM 3.1 (Decidability Problem) Determine all varieties of algebras that have a decidable first order theory.

The decidability problem has been solved under a finiteness assumption by S. Burris, R. McKenzie and M. Valeriote. They decomposed these varieties into three components, each having well-understood structure. We shall present the details of this theorem as we define the required concepts along our way.

The second problem is algebraic in nature, and, as it turned out, its solution involves model theoretic methods. The problem has to do with structural decompositions. Let us investigate some examples first. Consider the variety $\mathcal{A}_m$ of all abelian groups satisfying the identity $mx = 0$ for a given positive integer m. It is well-known that every member of this variety (even the infinite ones) can be written as a direct sum of cyclic groups

of prime power order. Thus there is a structure theorem for the members of this class. Another way of putting this result is that the only directly indecomposable members of $\mathcal{A}_m$ are the cyclic groups of prime power order, and every member is the direct sum of directly indecomposable members.

Now let us look at a different example. If $\mathbf{2}$ denotes the two element boolean algebra, then it is of course directly indecomposable, and every finite boolean algebra is a direct power of $\mathbf{2}$. By Stone's representation theorem, every boolean algebra is isomorphic to a field of sets, that is, a *subalgebra* of a direct power of $\mathbf{2}$.

So the first step in proving a structure theorem for a variety $\mathcal{V}$ of algebras may be to find a reasonably small class $\mathcal{K}$ of its members so that every element of the variety can be embedded into a direct product of elements of $\mathcal{K}$. In notation, $\mathcal{V} \subseteq \mathsf{SP}(\mathcal{K})$ (here P stands for direct products, and S for subalgebras). Such a class $\mathcal{K}$ is $\{\mathbf{2}\}$ in the variety of boolean algebras, and the class of cyclic groups of prime power order in the variety $\mathcal{A}_m$. At the next stage we could endeavor to describe the structure of the members of $\mathcal{K}$, and at the same time try to restrict ourselves to certain, well-behaved subalgebras of these direct products (like direct sums in the case of abelian groups above).

One can encounter difficulties even in the first step of this process. Let us consider the variety $\mathcal{V}$ of groups defined by the identities $x^4 = 1$ and $x^2y = yx^2$. The eight element dihedral group $\mathbf{D}_4$ and the eight element quaternion group $\mathbf{Q}$ are typical members of this variety. In fact, $\mathcal{V} = \mathsf{HSP}(\mathbf{D}_4) = \mathsf{HSP}(\mathbf{Q})$, where H stands for homomorphic images. We shall express this fact by saying that $\mathcal{V}$ is generated by both $\mathbf{D}_4$ and $\mathbf{Q}$.

Surprisingly, no set $\mathcal{K}$ works for this variety. That is, if $\mathcal{K} \subseteq \mathcal{V}$, and $\mathcal{V} \subseteq \mathsf{SP}(\mathcal{K})$, then $\mathcal{K}$ must have so many members that it has to be a proper class, it cannot be a set. In other words, $\mathcal{K}$ must contain groups of arbitrarily big cardinality. We call a variety *residually small* if there exists a *set* $\mathcal{K}$ with the above properties. Our first goal in the plan above, therefore, is to characterize residually small varieties.

A theorem of Birkhoff paves the road for the first steps in this process. The idea is to try to find a minimal class $\mathcal{K}$ that works by looking at those algebras that are "indecomposable" in the above sense. Suppose that an algebra $\mathbf{S}$ has been obtained as a subalgebra of a direct product $\mathbf{P} = \prod \mathbf{A}_i$, and let $\pi_i : \mathbf{P} \to \mathbf{A}_i$ denote the projection homomorphisms. Then $\pi_i(S)$ is a subalgebra of A_i for every i, and if we are interested only in the structure of $\mathbf{S}$, then we may as well assume that $\pi_i(S) = A_i$ for every i. In this case we shall say that $\mathbf{S}$ is a subdirect product of the algebras $\mathbf{A}_i$. An algebra $\mathbf{S}$ is called *subdirectly irreducible* if it has no nontrivial subdirect decompositions, that is, in every subdirect decomposition, one of the projections π_i is injective on S (and thus $\mathbf{S}$ is isomorphic to the corresponding $\mathbf{A}_i$).

Birkhoff's theorem states that every algebra is a subdirect product of subdirectly irreducible algebras (that are factors of the original algebra). Therefore a variety is residually small if and only if, up to isomorphism, there is only a set of subdirectly irreducible algebras in the variety.

PROBLEM 3.2 (RS problem) Investigate the distribution of subdirectly irreducible algebras in varieties. Characterize residually small varieties.

As the example of the quaternion group described above shows, even in a variety $\mathcal{V}$ generated by a single finite algebra it is possible that subdirectly irreducible algebras of unbounded cardinality exist. It can of course happen (as in the case of boolean algebras or abelian groups) that there are only very few subdirectly irreducibles in such a variety. Surprisingly, it is hard to find "mediocre cases", as we shall see later. Also, there is a feeling, which can be substantiated, that algebras with a "nice" structure tend to generate residually small varieties. At least, we should be able to find an algorithm that computes if a given finite algebra generates a residually small variety. We shall explore these questions in the forthcoming sections.

4. A natural type of decidable varieties

To prove a structure theorem for the members of a general variety, it is useful to know which are the "best-behaved" varieties from a structural point of view.

Consider a finite algebra $\mathbf{A}$ with richest possible structure. That is, assume that every (finitary) function on its underlying set is a term function (which means that it can be expressed as a composition of the basic operations). Finite algebras with this property are called *primal*. The two element boolean algebra, or finite fields of prime order are examples of primal algebras.

If $\mathbf{A}$ is a (finite) primal algebra, then the only subdirectly irreducible algebra in the variety $\mathcal{V}$ generated by $\mathbf{A}$ is $\mathbf{A}$ itself. Every finite member of $\mathcal{V}$ is a direct power of $\mathbf{A}$. This does not hold for the infinite algebras in the variety. Nevertheless, the variety $\mathcal{V}$ is decidable, because every infinite algebra in this variety can be described using a boolean algebra in the following way.

From Stone's representation theorem we know that if $\mathbf{B}$ is a boolean algebra, then the set X of all homomorphisms from $\mathbf{B}$ to $\mathbf{2}$ can be endowed with a topology, which is Hausdorff, compact, and zero-dimensional, that is, it has a base of clopen sets. Such topological spaces are called *boolean spaces*. There is a correspondence between boolean algebras and boolean spaces, as the original boolean algebra $\mathbf{B}$ can be recovered as the set of all clopen subsets of the boolean space $\mathbf{X}$. In other words, if for each boolean space $\mathbf{X}$ we consider the algebra consisting of all elements f of the direct

power $\mathbf{2}^X$ satisfying that both $\{x \in X \mid f(x) = 0\}$ and $\{x \in X \mid f(x) = 1\}$ are clopen subsets of $\mathbf{X}$, then we shall obtain all boolean algebras.

Now do this construction with an arbitrary algebra $\mathbf{A}$ in place of $\mathbf{2}$. For a boolean space $\mathbf{X}$ consider all elements f of $\mathbf{A}^X$ satisfying that the set $\{x \in X \mid f(x) = a\}$ is a clopen subset of $\mathbf{X}$ for every $a \in A$. These elements form a subdirect power of $\mathbf{A}$, which is called a boolean power of $\mathbf{A}$. It can be shown that every member of the variety generated by a primal algebra $\mathbf{A}$ is isomorphic to a boolean power of $\mathbf{A}$.

A similar representation theorem holds for more general algebras. For example, let $\mathbf{F}$ be an arbitrary finite field. It is not primal unless it is of prime order. The finite algebras here are direct products of subfields of $\mathbf{F}$. The infinite algebras can still be described using boolean algebras, but we have to allow non-isomorphic components. Thus, let $\mathbf{X}$ be a boolean space, and consider any direct product $\prod \mathbf{A}_x$ for some algebras $\mathbf{A}_x$, $x \in X$. A subdirect subalgebra $\mathbf{S}$ of this product is called a *boolean product*, if the following two conditions are satisfied:

(1) For any $f, g \in S$, their equalizer $[\![f = g]\!] = \{x \in X \mid f(x) = g(x)\}$ is a clopen subset of $\mathbf{X}$.
(2) (Patchwork property) Let $N \subseteq X$ be clopen, $f, g \in S$, and define h to be equal to f on N, and equal to g on the complement of N. Then h must be in S for every choice of N, f, and g.

It can be shown that every algebra in $\mathsf{V}(\mathbf{F})$ is a boolean product of subalgebras of $\mathbf{F}$.

Call a variety $\mathcal{V}$ finitely boolean representable, if there is a class $\mathcal{K}$ of finitely many finite algebras in $\mathcal{V}$ such that every member of $\mathcal{V}$ is a boolean product of elements of $\mathcal{K}$. This similarity with boolean algebras is still sufficient to prove that *finitely boolean representable varieties of finite signature have a decidable first order theory.*

Notice that the variety $\mathcal{A}_m$ of abelian groups defined in the previous section is also finitely boolean representable. In fact, every direct sum $\bigoplus M_i$ $(i \in I)$ of modules can be considered as a boolean product in the following way. Define a topology on I by making all finite and cofinite sets open, and let $X = I \cup \{0\}$ be the one-point compactification of this space. At index 0 add the one-element module to this direct sum as a new summand. It is easy to see that in this setting the direct sum is indeed a boolean product. Thus if $\mathbf{R}$ is a finite ring such that there are only a finite number of directly irreducible $\mathbf{R}$-modules (like the ring of integers modulo m), then the variety of all $\mathbf{R}$-modules is boolean representable, hence, decidable.

Finitely boolean representable varieties are not far removed from the two special cases considered above. They can be decomposed (in a manner to be defined later) to two varieties. One is related to a variety of modules over a ring, it is called the abelian component, and will be discussed in

Section 6. The other one is a generalization of varieties generated by primal algebras. Consider the following function on a set A:

$$t(x,y,z) = \begin{cases} z & \text{if } x = y \\ x & \text{if } x \neq y. \end{cases}$$

The algebra (A,t) is not primal, since it is easy to check that every subset of this algebra is a subalgebra (and every permutation of A is an automorphism). However, it is not far from being primal, because adding all constants as new operations we obtain a primal algebra.

This function t is called the *discriminator* function of $\mathbf{A}$. A variety $\mathcal{V}$ is called a discriminator variety, if it has a term t which is the discriminator on every subdirectly irreducible algebra of the variety. These subdirectly irreducible algebras must in fact be simple, as an easy calculation with the discriminator function shows. The reader can verify that the variety generated by finitely many finite fields is always a discriminator variety.

Every discriminator variety is boolean representable by the class of its subdirectly irreducible (hence simple) algebras, and conversely, the non-abelian component of every finitely boolean representable variety is a discriminator variety. Discriminator varieties also come up in the characterization of decidable varieties, as we shall see.

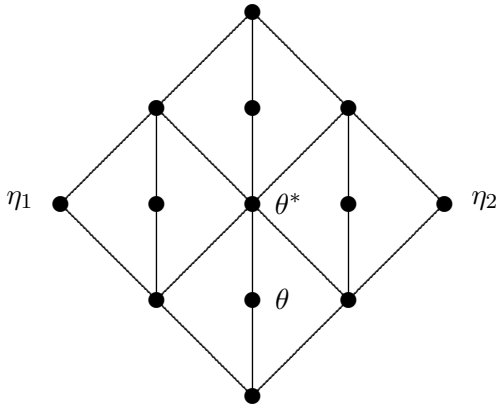
5. Controlling subdirectly irreducibles

How do we know that the only subdirectly irreducible algebra in a variety generated by a primal algebra is this algebra itself? How can we construct discriminator varieties? This is a nontrivial question, since in the definition we have to consider all subdirectly irreducible algebras in the variety. How to recognize subdirectly irreducible algebras at all?

Let us recall the concept of a congruence. If a homomorphism $\varphi : \mathbf{A} \to \mathbf{B}$ is given, then its kernel is defined to be the equivalence relation on A, where two elements are equivalent if and only if φ collapses them. A *congruence* of $\mathbf{A}$ is a partition of A that is a kernel of some homomorphism (to some $\mathbf{B}$). Congruences can be characterized internally: they are equivalence relations that satisfy the substitution property with respect to all basic operations f, which says that $a_i \equiv b_i$ for all i must imply that $f(a_1,\ldots,a_n) \equiv f(b_1,\ldots,b_n)$. This condition is equivalent to saying that the set of pairs (a,b) with $a \equiv b$ is a subalgebra of $\mathbf{A} \times \mathbf{A}$. The congruences of $\mathbf{A}$ are ordered by inclusion: the bigger congruence contains more pairs. This ordering defines a lattice that is called the congruence lattice of $\mathbf{A}$ and is denoted by Con$\mathbf{A}$. This lattice has a largest element 1_A (which has only one congruence block), and a smallest one called 0_A (in which all elements form a one-element class).

Now it is not hard to show that an algebra $\mathbf{S}$ is subdirectly irreducible if and only if among its nonzero congruences there is a smallest one (which is contained by all other congruences). This congruence is called the monolith of $\mathbf{S}$ and is denoted by $\mu(\mathbf{S})$. Thus the quaternion group is subdirectly irreducible, its monolith is the congruence that collapses each element g with $-g$. It corresponds to the normal subgroup $\{1, -1\}$. Simple algebras are also subdirectly irreducible.

How can new subdirectly irreducibles come up in a variety? The example of the quaternion group shows this as well, but it is simpler to present this phenomenon using the example of the ring $\mathbf{R} = 2\mathbf{Z}_8$. This ring consists of the numbers $\{0, 2, 4, 6\}$, and the operations are addition and multiplication modulo 8. The variety generated by $\mathbf{R}$ is residually large. Here is the picture of the congruence (ideal) lattice of $\mathbf{R} \times \mathbf{R}$.



The two projection kernels are denoted by η_1 and η_2, respectively. The congruence θ corresponds to the ideal $\{(0,0),(4,4)\}$, and the congruence θ^* corresponds to the ideal $\{0,4\} \times \{0,4\}$. It is clear that the factor modulo θ is subdirectly irreducible, with monolith θ^*/θ, because every congruence strictly above θ contains θ^*.

This lattice is not distributive, because we have $\theta \vee (\eta_1 \wedge \eta_2) = \theta$, but $(\theta \vee \eta_1) \wedge (\theta \vee \eta_2) = \theta^*$. This observation holds for every newly created subdirectly irreducible algebra in any subdirect square. Indeed, let $\mathbf{B}$ be a subdirect square of two algebras $\mathbf{A}_1$ and $\mathbf{A}_2$ with projection kernels η_1 and η_2 (so $\mathbf{B}/\eta_i \cong \mathbf{A}_i$), and let $\mathbf{B}/\theta$ be subdirectly irreducible with monolith θ^*/θ. Then $\eta_1 \wedge \eta_2 = 0$ (since if two pairs agree in both coordinates, then they are equal), so $\theta \vee (\eta_1 \wedge \eta_2) = \theta$. On the other hand, $\theta \vee \eta_1$ properly contains θ, since otherwise $\eta_1 \leq \theta$, so $\mathbf{B}/\theta$ is already a factor of $\mathbf{B}/\eta_1 \cong \mathbf{A}_1$, and therefore this subdirectly irreducible is not "new". As θ^*/θ is the monolith, we get that $\theta \vee \eta_1 \geq \theta^*$. Similarly we have $\theta \vee \eta_2 \geq \theta^*$,

and thus $(\theta \vee \eta_1) \wedge (\theta \vee \eta_2) \geq \theta^*$. So the congruence lattice of $\mathbf{B}$ cannot be distributive.

We have shown that if we are in a variety $\mathcal{V}$, where all the algebras have distributive congruence lattices, then each subdirectly irreducible factor of a subdirect square is already isomorphic to a factor of one of the components. This observation clearly extends to finite subdirect products. If this variety is generated by some class $\mathcal{K}$ of algebras, then it is easy to see that each finite algebra in $\mathcal{V}$ is actually a homomorphic image of a subalgebra of a finite direct product of the generators. Therefore every finite subdirectly irreducible must be contained in $\mathsf{HS}(\mathcal{K})$.

With slightly more effort, but using the same idea, we can control the infinite subdirectly irreducibles as well. In that case we have to allow forming ultraproducts of the generators besides subalgebras and homomorphic images. The following lemma basically settles the RS problem for congruence distributive varieties.

LEMMA 5.1 (Jónsson's Lemma) *Let $\mathcal{V}$ be a congruence distributive variety generated by a class $\mathcal{K}$ of algebras. Then every subdirectly irreducible algebra in $\mathcal{V}$ is a homomorphic image of a subalgebra of an ultraproduct of some members of $\mathcal{K}$.*

How can we make sure that a variety given by some generators is congruence distributive? There is a simple way which is sufficient in most cases. Suppose that there is a term m such that all the generators satisfy the identities

$$m(x, x, y) = m(x, y, x) = m(y, x, x) = x\,.$$

Such a term m is called a majority term (if two arguments agree, then the result is this value). Since these are identities, if they are satisfied in the generators, then they are satisfied throughout the variety. On the other hand, it is not hard to show that an algebra having a majority term has a distributive congruence lattice. The most important example for a majority term is the term

$$(x \vee y) \wedge (x \vee z) \wedge (y \vee z)$$

in lattices. Thus the variety of lattices is congruence distributive.

Now let us return to discriminator varieties. Suppose that a class $\mathcal{K}$ of algebras is given such that a term t is the discriminator on every member of $\mathcal{K}$. Will the variety $\mathcal{V}$ generated by $\mathcal{K}$ be a discriminator variety? The answer is yes, and we have all the tools to show this.

First note that $t(x, t(x, y, z), z)$ is a majority function on all members of $\mathcal{K}$, so $\mathcal{V}$ is congruence distributive. The discriminator function is defined by a first order formula, so t will be the discriminator on every ultraproduct formed from the members of $\mathcal{K}$. Clearly, t will be the discriminator on every

subalgebra, too. Finally, we have already mentioned that any algebra having a discriminator term is simple. Therefore t is indeed the discriminator on every subdirectly irreducible algebra in our variety.

To finish the topic of congruence distributive varieties, we return to the decidability question. The finiteness assumption in the Burris-McKenzie-Valeriote theorem is that the variety is locally finite, that is, each of its finitely generated algebras are finite. As a main example, it is easy to check that every finitely generated variety is locally finite. The theorem implies that *a locally finite, congruence distributive, decidable variety must be a discriminator variety.* In fact, one of the three components in the general case is a discriminator variety.

It is still unknown which (non locally finite) congruence distributive varieties are decidable. Actually, it is an open question to determine all decidable discriminator varieties. However, as we mentioned earlier, finitely generated discriminator varieties of finite signature are decidable.

6. The modular commutator

The results presented so far are relatively easy to prove, but they do not cover the important classical structures: groups and rings. As we have seen, these structures provide examples of residually large varieties (while a finitely generated congruence distributive variety must be residually small by Jónsson's lemma). A natural class that encompasses all these examples, as well as congruence distributive varieties, is the class of congruence modular varieties. For such varieties, the solution of our two problems was much harder. The breakthrough came in the middle of the seventies, with the development of commutator theory (whose ideas play a significant role in tame congruence theory, too). To show the idea of this theory, let us investigate Jónsson's lemma for groups and rings.

Introduce the notation $[I, J] = IJ + JI$ for any two ideals I and J of a ring R . Here IJ denotes the set of all finite sums of elements of the form ij with $i \in I$ and $j \in J$. Clearly, $[I, J]$ is an ideal contained in $I \cap J$. The equation $[I, I] = 0$ is equivalent to saying that I is a zeroring (a ring with zero multiplication). Finally, $[I + J, K] = [I, K] + [J, K]$ holds for any ideals I, J, and K.

This last identity is very similar to the distributive law, since in the lattice of ideals, join is the same as $+$. So let us employ the argument in the proof of Jónsson's lemma (see the figure above showing the congruence lattice of $\mathbf{R} \times \mathbf{R}$). To keep the same notation, we shall write congruences when we think of ideals, and write join instead of $+$.

From $\theta^* \leq \theta \vee \eta_i$ we obtain that

$$[\theta^*, \theta^*] \leq [\theta \vee \eta_1, \theta \vee \eta_2] = [\theta, \theta] \vee [\theta, \eta_1] \vee [\theta, \eta_2] \vee [\eta_1, \eta_2] \leq \theta \vee (\eta_1 \wedge \eta_2) = \theta \, .$$

Thus, if a "new" subdirectly irreducible occurs, then its monolith must have zero multiplication! So Jónsson's lemma is still valid for all other kinds of subdirectly irreducibles (fields, for example).

Can we use a similar idea for groups? Yes, but in this case we have to define the operation $[N, M]$ for two normal subgroups to be their commutator subgroup, that is, the subgroup generated by all elements of the form $n^{-1}m^{-1}nm$, where $n \in N$ and $m \in M$. It is well-known and obvious that all the properties above hold, and so the proof goes through as well. In the group case, the monolith of any "new" subdirectly irreducible must be an abelian normal subgroup.

So even when new subdirectly irreducibles come up we are in a favorable situation, since zerorings and abelian groups are certainly easier to handle than general rings and groups: in both cases these can rather be considered as modules over a ring. The really surprising fact is that the idea just outlined works not just for rings and groups, but for general structures as well.

To see how general these structures can be, let us observe a property of groups and rings, similar to having a majority term for lattices. A ternary function $d(x, y, z)$ is called a Maltsev function if it satisfies the identities

$$d(x, x, y) = d(y, x, x) = y\,.$$

The existence of a Maltsev term of an algebra has a nice effect on the congruences: it ensures that the join of any two congruences can be computed simply by taking their relation product. Algebras with such congruences are called congruence permutable. The discriminator is a Maltsev function. Since $xy^{-1}z$ is a Maltsev term in every group, groups and rings are congruence permutable (as indeed the product of two normal subgroups of a group is always a normal subgroup).

An easy-to-prove consequence of congruence permutability is that the congruence lattice of the algebra in question is modular. And as distributivity implies modularity, the class of congruence modular varieties contains not just groups and rings, but all congruence distributive varieties as well (even though these, for example lattices, are not congruence permutable in general). Commutator theory works for general modular varieties. To build up this theory is a nontrivial task, involving ingenious proofs. The main facts, however, can be stated relatively simply.

In the above argument generalizing Jónsson's lemma, we used some identities for the commutator, and these are easy to understand in the general case. But what do we mean by saying that "a general algebra behaves like a module over a ring"? We mean that it is polynomially equivalent to a module over an associative ring with identity. What do such algebras look like? Can we recover this module from an affine algebra? The answer is

yes, and in fact questions (like decidability, residual smallness) concerning affine algebras are easily translated to questions about "real" modules. If $\mathbf{M}$ is a left $\mathbf{R}$-module, then its polynomials are the functions

$$f(x_1, \ldots, x_n) = r_1 x_1 + \cdots + r_n x_n + m \,,$$

where $r_i \in R$ and $m \in M$. Among these, of particular importance is the Maltsev function $x - y + z$, which must always be a term in any algebra on M that is polynomially equivalent to $\mathbf{M}$. From this, we can always recover the addition, and we can get the ring elements by looking at unary polynomials.

THEOREM 6.1 (The Fundamental Theorem of the commutator) *Let $\mathcal{V}$ be any modular variety. Then there exists a binary operation $[\ ,\ ]$ on the congruence lattice of every algebra of $\mathcal{V}$ such that for all congruences α, β, β_i we have:*

(1) $[\alpha, \beta] = [\beta, \alpha] \leq \alpha \wedge \beta$.
(2) $[\alpha, \bigvee \beta_i] = \bigvee [\alpha, \beta_i]$.
(3) *If $\varphi : \mathbf{A} \to \mathbf{B}$ is an epimorphism, then $[\varphi(\alpha), \varphi(\beta)] = \varphi[\alpha, \beta]$.*
(4) *If $[1_{\mathbf{A}}, 1_{\mathbf{A}}] = 0_{\mathbf{A}}$, then the algebra $\mathbf{A}$ is affine.*

Once the commutator is defined, we can speak of abelian (solvable, nilpotent) algebras or congruences (as their definitions from group theory carry over without modification). For example a congruence α is abelian if it satisfies $[\alpha, \alpha] = 0$. The structure of abelian congruences can also be described using module theoretic methods. The argument we have shown above can be used to prove that subdirectly irreducible algebras with non-abelian monolith satisfy Jónsson's lemma in any modular variety. This statement can be improved further, and using the structure of abelian congruences, and other ideas, one can provide a characterization of finitely generated residually small modular varieties.

THEOREM 6.2 (R. Freese, R. McKenzie) *The following are equivalent for a finite algebra $\mathbf{A}$ generating a modular variety $\mathcal{V}$.*

(1) *$\mathcal{V}$ is residually small.*
(2) *Every subdirectly irreducible algebra in $\mathcal{V}$ has at most $m + \left(m^{m^{m+3}}\right)!$ elements, where m is the size of A.*
(3) *For any two congruences α and β of any subalgebra of $\mathbf{A}$ we have that*

$$\alpha \wedge [\beta, \beta] = [\alpha \wedge \beta, \beta] \,.$$

The "!" in (2) stands for factorial. This is a huge bound, but a finite bound nevertheless. The real bound is unknown, but it is still exponential. Remember, in the congruence distributive case we always have the bound m

by Jónsson's lemma, and the condition in (3) is satisfied for every $\mathbf{A}$ (since in a congruence distributive variety the commutator is always the intersection). Condition (3) may look complicated, but it gives an algorithm to decide if $\mathbf{A}$ generates a residually small variety. In concrete cases it translates to nice structural properties, for example a finite group satisfies (3) if and only if all of its Sylow subgroups are abelian. The reader is encouraged to find a failure of (3) in the groups $\mathbf{Q}$ and $\mathbf{D}_4$, and in the ring $\mathbf{R} = 2\mathbf{Z}_8$ discussed in the previous sections.

To check the condition in (3) one has to be able to compute the commutator of two congruences in general algebras. The commutator has several definitions. We present the one that allows us to use these concepts beyond modularity. In this approach, we define abelianness and centrality first.

Let us investigate the properties of module polynomials. Suppose that an equality $f(a, \mathbf{c}) = f(a, \mathbf{d})$ holds in a module $\mathbf{M}$, where $f(x, \mathbf{y})$ is a polynomial. We have

$$f(x, y_1, \dots, y_n) = rx + r_1 y_1 + \cdots + r_n y_n + m\,,$$

so this equality implies that $r_1 c_1 + \ldots + r_n c_n + m = r_1 d_1 + \ldots + r_n d_n + m$. Therefore $f(x, \mathbf{c}) = f(x, \mathbf{d})$ holds for every $x \in M$.

We say that an algebra $\mathbf{A}$ is abelian if it satisfies the implication just obtained: for every polynomial $f(x, \mathbf{y})$ we have

$$f(a, \mathbf{c}) = f(a, \mathbf{d}) \quad \Longrightarrow \quad f(x, \mathbf{c}) = f(x, \mathbf{d})$$

for any elements a, x and vectors $\mathbf{c}$, $\mathbf{d}$ of A. To understand this concept the reader is encouraged to show that a group is abelian in this sense if and only if it is commutative, and a ring is abelian if and only if it is a zeroring.

If we localize this concept to congruences, we obtain centrality. We say that the congruence α centralizes the congruence β, if the implication above holds for all cases when f is arbitrary, but we have that a and x are α-related, and $\mathbf{c}$ and $\mathbf{d}$ are β-related componentwise. Again, it is easy to check that the congruences corresponding to two normal subgroups of a group centralize each other in this sense if and only if the normal subgroups commute with each other elementwise. In rings, centrality corresponds to mutual annihilation of ideals.

Finally, the commutator of α and β is defined to be the smallest congruence γ satisfying that α/γ and β/γ centralize each other in the factor $\mathbf{A}/\gamma$. Again, the reader can easily check that this is indeed the concept we spoke about in groups and rings.

These definitions of abelianness, centrality, and the commutator make sense in any algebra whatsoever. However, to prove the properties mentioned in the fundamental theorem above is a very nontrivial task, where congruence modularity is used heavily.

The question of decidability can also be settled for the locally finite modular case with these (and other) methods. Recall the structure of finitely boolean representable varieties. We said that they decompose to two varieties, one is a discriminator variety, and one is a variety which we called abelian. Now we can state precisely what this latter concept means: it is simply a (modular) variety consisting of abelian, that is, affine algebras. But what do we mean by the phrase "can be decomposed"?

The "join" $\mathcal{V}$ of two varieties $\mathcal{V}_1$ and $\mathcal{V}_2$ can be defined as the smallest variety containing both $\mathcal{V}_1$ and $\mathcal{V}_2$. Its members are factors of subalgebras of algebras of the form $\mathbf{A}_1 \times \mathbf{A}_2$, where $\mathbf{A}_1 \in \mathcal{V}_1$ and $\mathbf{A}_2 \in \mathcal{V}_2$. These kinds of algebras can be quite complicated, so it is a valuable situation, when the algebras in the join look nicer. If it is the case that every algebra of $\mathcal{V}$ is the direct product of an algebra from $\mathcal{V}_1$ and an algebra from $\mathcal{V}_2$, then we shall say that $\mathcal{V}$ is a varietal product of $\mathcal{V}_1$ and $\mathcal{V}_2$, and write $\mathcal{V} = \mathcal{V}_1 \otimes \mathcal{V}_2$.

This simpler case applies in the case of finitely boolean representable varieties: *every such variety is a product of an affine and a discriminator variety.* We have a similar result for decidability: we now know two components of the three!

THEOREM 6.3 (S. Burris, R. McKenzie) *Every locally finite, decidable, modular variety is a product of an affine and a discriminator variety.*

This hard result allows one to reduce the decidability of locally finite modular varieties to the question of characterizing all finite rings $\mathbf{R}$ such that the variety of all $\mathbf{R}$-modules is decidable (and to the decidability question of discriminator varieties as discussed earlier). The problem of determining such finite rings $\mathbf{R}$ is unsolved, and is being investigated by ring-theorists. As an easy consequence we see that a locally finite variety of groups is decidable if and only if it consists of abelian groups, and it is easy to check that a locally finite variety of rings is decidable if and only if it is generated by finitely many finite fields and a zeroring. These results have been proved earlier by A. P. Zamjatin.

7. The type labeling

Having considered the modular case, we now return to tame congruence theory, which deals with general finite algebras. In Section 2 we have listed the possible induced algebras on minimal neighborhoods of finite simple algebras. Let us first state the theorem that allows us to classify these induced algebras into five types.

Thus, let $\mathbf{A}$ be a finite simple algebra and f a non-constant unary polynomial of $\mathbf{A}$ whose range $N = f(A)$ is minimal (under inclusion). Sets N obtained this way are called the *minimal sets* of $\mathbf{A}$. A surprising fact is that every minimal set must be a neighborhood, that is, the range of an

idempotent polynomial. Thus the minimal sets and minimal neighborhoods of any finite simple algebra are the same. From this observation it follows easily that the induced algebra $\mathbf{A}|_N$ must be a *minimal algebra*, that is, each of its unary polynomials is either a permutation of N, or a constant mapping.

There is a structure theorem for minimal algebras, due to P. P. Pálfy, which has been proved originally to solve an entirely different problem.

THEOREM 7.1 (Pálfy) *A finite algebra is minimal if and only if it is polynomially equivalent to one of the following algebras.*

(1) *A permutation group.*
(2) *A finite vector space (over a finite field).*
(3) *The two-element boolean algebra.*
(4) *The two-element lattice.*
(5) *The two-element semilattice.*

Since N is a neighborhood, it is easy to show that all congruences of $\mathbf{A}|_N$ are restrictions of congruences of $\mathbf{A}$. Thus, $\mathbf{A}|_N$ must be simple, and we get the classification stated in Section 2.

Before proceeding to the non-simple case, let us see some concrete examples and their minimal sets. If $\mathbf{A}$ is a discriminator algebra, then, as we have seen, every function is a polynomial. Therefore the minimal sets are just all two-element subsets of A, and the type is **3**. The same holds for every nonabelian finite simple group.

If $\mathbf{A}$ is an abelian simple group (of prime order), then it is a minimal algebra, so the only minimal set is A itself, and the type is **2**. Now let $\mathbf{F}$ be a finite field, and $M = F^n$, considered as a module over the full matrix ring $\mathbf{F}^{n\times n}$ in the usual way. This is still a simple abelian algebra of type **2**, and its minimal sets are exactly the "lines", that is, the cosets modulo the one dimensional subspaces of the vector space $\mathbf{F}^n$. Now keep only those terms that map to a minimal set, as the basic operations of a new algebra. This new algebra is still simple of type **2**, and its minimal sets are the same, but it is not any more Maltsev. It can be shown that every finite, simple algebra of type **2** can be obtained in a similar way, that is by throwing away some terms and elements of a module.

The lattice $\mathbf{M}_3$ (five elements, three atoms) is a simple algebra of type **4**. Its minimal sets are exactly the pairs that cover each other in the lattice, and the pair $\{0, 1\}$ (so there are seven minimal sets).

Every finite simple semigroup $\mathbf{S}$ with a zero element 0 such that $\mathbf{S}^2 \neq 0$ has type **5**. To see another example, consider a finite directed graph $\mathbf{G}$, add a new symbol ∞ to G, and define a binary operation by letting $ab = b$ if there is an edge $a \to b$, and let all other products be ∞. For most finite graphs the resulting graph-algebra is simple of type **5**. Graph-algebras have

provided important examples when investigating Tarski's famous problem on finite equational bases of varieties. Finally, a class of type **1** simple algebras can be obtained from permutation groups by taking matrix-powers, as explained in the next section.

How can we use the above ideas to investigate arbitrary finite algebras? Instead of requiring that our algebra be simple, we simply pick a covering pair (in other words, a prime quotient) of congruences in its congruence lattice, and do our investigations with respect to this quotient.

So let $\alpha \prec \beta$ be a covering pair of congruences of a finite algebra $\mathbf{A}$. We look for minimal sets that are sensitive to the structure of this prime quotient. Consider all unary polynomials f of $\mathbf{A}$ that don't collapse β to α, that is, there exists a β-related pair (a, b) such that $(f(a), f(b))$ is not α-related. Among the ranges of these polynomials, there are minimal ones, under inclusion. These ranges are called the $\langle\alpha, \beta\rangle$-minimal sets of $\mathbf{A}$, and their set is denoted by $\mathrm{M}_{\mathbf{A}}(\alpha, \beta)$. One can prove that every such minimal set is a neighborhood, which will ensure the good behavior of its induced operations.

Next we have to reveal the structure of the induced algebras on these minimal sets. Let U be an $\langle\alpha, \beta\rangle$-minimal set, $\mathbf{C} = \mathbf{A}|_U$, and δ and θ the restrictions of the congruences α and β to U, respectively. Then $\mathbf{C}$ will be a $\langle\delta, \theta\rangle$*-minimal algebra,* that is, every unary polynomial of $\mathbf{C}$ either collapses θ to δ, or is a permutation of C.

This condition is weaker than the minimality condition in Pálfy's theorem. Nevertheless, such algebras can still be completely classified. One should picture $\mathbf{C}$ as a beast. There are two kinds of θ-classes of $\mathbf{C}$. The interesting ones from the point of view of the structure of the congruence quotient $\langle\delta, \theta\rangle$ are the ones which consist of more than one δ-classes. These θ-classes are the *traces* of $\mathbf{C}$. The union of the traces is the *body*, and the rest of the algebra is the *tail*.

Usually there isn't much we can say about the tail of a minimal set (in nicer cases, like in congruence modular varieties, the tails are empty). The important information is carried by the traces. If we factor out the induced algebra on a trace by δ, then we get a minimal algebra in the sense of Pálfy's theorem. Thus the structure of traces is known, and each trace can be given a type $\mathbf{1} - \mathbf{5}$. A minimal set can have more than one trace (although this cannot happen in the nonabelian types $\mathbf{3} - \mathbf{5}$), but the induced algebras on these traces are always isomorphic, and therefore the minimal set itself has a single type label: that of its traces. Further investigations reveal more about the structure of $\langle\delta, \theta\rangle$-minimal algebras, a different theorem for each type. For example, in the case of type **2**, the whole body has an induced Maltsev polynomial.

Finally, we try to assemble the structure of the quotient $\langle\alpha,\beta\rangle$ from the structure of the $\langle\alpha,\beta\rangle$-minimal sets. The properties listed in Section 2 hold in this general setting, too, in an appropriate form. We have uniformity: any two $\langle\alpha,\beta\rangle$-minimal sets carry isomorphic induced algebras. Therefore each one has the same type label. This type is called the type of the quotient $\langle\alpha,\beta\rangle$.

The isomorphisms between minimal sets (and also the isomorphisms between traces within a single minimal set) are established by unary polynomials of $\mathbf{A}$. To show how this works suppose that U and V are $\langle\alpha,\beta\rangle$-minimal sets, and f is a unary polynomial of $\mathbf{A}$ that is a bijection from U to V, having an inverse g, which is also a polynomial of $\mathbf{A}$. Then the induced algebras $\mathbf{A}|_U$ and $\mathbf{A}|_V$ are isomorphic. Indeed, any basic operation of $\mathbf{A}|_U$ is a polynomial of $\mathbf{A}$ restricted to U. Composing this polynomial with f and g appropriately, we obtain a corresponding basic operation of $\mathbf{A}|_V$. In this situation we shall say that U and V are *polynomially isomorphic*.

We also have separation and connectedness. Let us call a trace in any $\langle\alpha,\beta\rangle$-minimal set an $\langle\alpha,\beta\rangle$-trace. Then the following hold:

- *Separation.* Any pair of elements in $\beta-\alpha$ can be mapped to any $\langle\alpha,\beta\rangle$-trace by a unary polynomial in such a way that the resulting pair is still not in α.
- *Connectedness.* Any two elements that are β-related can be connected by a chain of $\langle\alpha,\beta\rangle$-traces and α-related pairs.

We can still say that the algebra behaves "around the quotient $\langle\alpha,\beta\rangle$" the same way as the corresponding minimal algebras in Pálfy's theorem. For example, the two-element boolean algebra and the two-element lattice are in congruence distributive varieties, and for quotients of type **3** and **4** a weak version of Jónsson's lemma is satisfied. Permutation groups and finite vector spaces are exactly the abelian minimal algebras (in the sense of the previous section), and indeed it can be proved that a cover $\alpha \prec \beta$ has type **1** or **2** if and only if β/α is an abelian congruence.

The proof of this last fact is a beautiful example of the technique introduced in Section 2. We outline it for the case of a simple algebra $\mathbf{A}$; with some work the reader should be able to fill in the details. Consider a failure of the abelian property, and use separation to map it down to a trace with a unary polynomial. This way we obtain a new failure, which involves a polynomial already mapping into a trace. On the other hand, such polynomials can be understood: in the type **1** case they depend on at most one variable, and in the type **2** case they are sums of unary polynomials. These statements can be shown by analyzing the induced algebras on the traces first, and then applying connectedness to prepare an "atlas" of $\mathbf{A}$.

The shape of the congruence lattice of an algebra has an important influence on the type labeling. "Weird" lattices can force the types **1** and **5**

to occur in certain critical intervals. On the other hand, modular varieties can have only types **2**, **3**, and **4**, among these groups and rings must omit type **4**, while congruence distributive varieties must omit type **2**. In general, types **1** and **5** are badly behaved, but they cannot occur in any variety in which the congruence lattices of the algebras satisfy any nontrivial lattice identity.

We conclude this section with some examples of $\langle \alpha, \beta \rangle$-minimal sets. The groups $\mathbf{Q}$, $\mathbf{D}_4$ (and in fact every finite p-group for any prime p) have the property that they are minimal with respect to every prime quotient of congruences, and have labels **2** throughout. The same holds for the ring $\mathbf{R} = 2\mathbf{Z}_8$.

Now consider the ring $\mathbf{Z}_4$. It is subdirectly irreducible, and has exactly one nontrivial congruence μ (its monolith), which corresponds to the ideal $\{0, 2\}$. The two minimal sets for the quotient $\langle 0, \mu \rangle$ are exactly the blocks of μ, this quotient has type **2**. All the remaining four two-element subsets are minimal sets for the quotient $\langle \mu, 1 \rangle$, which has type **3**.

The reader is encouraged to compute the minimal sets for the quotients of the symmetric group $\mathbf{S}_5$, which also has type set $\{\mathbf{2}, \mathbf{3}\}$, and to show that every finite lattice has only label **4**, every finite semilattice has only label **5**, and every algebra in which every basic operation is unary has only label **1**.

8. Solvable and strongly solvable algebras

The concept of solvability comes from group theory: we call an algebra $\mathbf{A}$ solvable if there is a chain of congruences of $\mathbf{A}$ such that all factors of this chain are abelian. This definition works for general algebras, too, and although in the general case it is not any more true that homomorphic images of abelian algebras are abelian (as is the case in modular varieties), one can still use tame congruence theory to show that homomorphic images of finite abelian algebras are solvable. This follows from the fact that a finite algebra is solvable if and only if the only types occurring in it are **1** and **2**. Every locally finite variety has a largest subvariety whose finite members are solvable.

How can we distinguish between types **1** and **2** from a global point of view? We have seen that the typical examples of modular abelian algebras are the modules over rings, which have labels **2** throughout. Similarly, the typical algebras having labels **1** throughout are the unary algebras.

The property of abelianness can be strengthened to characterize type **1** quotients in the following way. We say that a congruence α of an algebra $\mathbf{A}$ is *strongly abelian* (or *combinatorial*), if for every polynomial $f(x, \mathbf{y})$ of $\mathbf{A}$

we have

$$f(a, \mathbf{c}) = f(b, \mathbf{d}) \quad \Longrightarrow \quad f(x, \mathbf{c}) = f(x, \mathbf{d})$$

for all elements a, b, x and vectors $\mathbf{c}$ and $\mathbf{d}$ of $\mathbf{A}$ such that the elements a, b, x are α-related, and the vectors $\mathbf{c}$ and $\mathbf{d}$ are α-related componentwise. The reader is asked to check that this condition implies that α is abelian, and that every unary algebra is strongly abelian, that is, it satisfies this condition for its biggest congruence. It can also be proved that in any finite algebra a prime quotient is strongly abelian if and only if its type label is **1**.

As abelianness leads to the concept of solvability, strong abelianness leads to strong solvability in the same way, and similar statements hold for this concept. Thus a finite algebra is strongly solvable if and only if **1** is the only label that occurs in its congruence lattice. Every homomorphic image of a strongly solvable algebra is strongly solvable, and every locally finite variety has a largest subvariety whose finite members are strongly solvable.

Solvable varieties are an important special case to study when attacking problems. If we take a finite solvable, or even abelian algebra, then the variety it generates can be quite complicated. Of course all finite members in this variety must be solvable, and all type **2** minimal sets must have empty tails, but lots of type **1** quotients generally occur. In fact, if the finite members of a locally finite variety admit only type **2**, then the variety must be congruence permutable.

One should not think only of unary algebras when discussing strong solvability. We present an important class of strongly abelian algebras that occurs frequently in structure theorems. To construct it, we return to the examples of modules. If $\mathbf{M}$ is a module over a ring $\mathbf{R}$, then M^n (that is, column vectors of height n) form a module over the $n \times n$ matrix ring (via matrix multiplication). These new modules are very similar to the original one, for example, every homomorphism between two modules obtained this way is always given by a homomorphism between the two original modules, acting componentwise.

This "matrix power" construction can be defined for arbitrary algebras, and the resulting matrix powers and the original algebras will have very similar properties in the general case, too. Let $\mathbf{A}$ be any algebra, we want to make the set A^n into an algebra. If we consider the direct product $\mathbf{A}^n$, then each of its k-ary terms is given by a term of the original algebra, acting componentwise:

$$\hat{t}\left(\begin{pmatrix} a_{11} \\ \dots \\ a_{n1} \end{pmatrix}, \dots, \begin{pmatrix} a_{1k} \\ \dots \\ a_{nk} \end{pmatrix}\right) = \begin{pmatrix} t(a_{11}, \dots, a_{1k}) \\ \dots \\ t(a_{n1}, \dots, a_{nk}) \end{pmatrix}.$$

Notice that we have the same term t in every component of the right hand side, and in the i-the component of the result, the elements substituted to t are the i-th components of the arguments of $\hat{t}$.

This is the situation in a direct power. But we can add much more operations to A^n. We can have a different term in every component, and each of these terms may depend on all nk components that the k arguments altogether have. This way, we have constructed a function for each n-tuple of nk-ary terms of $\mathbf{A}$. If we consider all these functions as basic operations, then we obtain the n-th matrix power of $\mathbf{A}$, which is denoted by $\mathbf{A}^{[n]}$. It can be proved that this matrix power is similar to $\mathbf{A}$ in every respect: its subalgebras are just matrix powers of the subalgebras of $\mathbf{A}$, its congruence lattice is isomorphic to that of $\mathbf{A}$, it is strongly abelian if and only if $\mathbf{A}$ is strongly abelian, etc.

In our previous examples, this similarity did not bring in new algebras. If the statement of some structure theorem is that the algebra considered must be affine, then it does not matter if a matrix power of the original algebra also satisfies the conditions of the theorem, because a matrix power of an affine algebra is still affine. Similarly, a matrix power of a primal algebra is still primal. But it is not true that a matrix power of a unary algebra is still unary. Therefore matrix powers of unary algebras often come up when formulating the strongly abelian case of structure theorems. As we shall see, this happens in the case of decidability, too.

9. An application: minimal varieties

As an illustration, we outline the proof of the structure theorem of locally finite, abelian, minimal varieties. A variety is called minimal if it has only the two trivial subvarieties. Thus we obtain the well-known characterization of ω-categorical varieties as well, because it is not difficult to show that ω-categorical varieties are locally finite, abelian, and minimal. The argument showing this can be found in Keith Kearnes [6], where a new, simple proof of the Givant-Palyutin theorem characterizing ω-categorical quasivarieties is given.

THEOREM 9.1 (K. Kearnes, E. W. Kiss, Á. Szendrei, M. Valeriote) *Let $\mathcal{V}$ be a locally finite variety such that all finite members are solvable. Then $\mathcal{V}$ is minimal if and only if one of the following possibilities holds.*

(1) *$\mathcal{V}$ is term equivalent to a matrix power of the variety of sets with no operations, or to the variety of sets with one constant operation. In this case $\mathcal{V}$ is strongly abelian.*
(2) *$\mathcal{V}$ is affine (in particular, it is congruence permutable), and is generated by a finite, simple algebra that is polynomially equivalent to a module over a finite ring, and has a one-element subalgebra.*

Since the varieties listed here are ω-categorical, this theorem characterizes ω-categorical varieties as well. We shall now outline the proof of this theorem, using tame congruence theory.

Thus, let $\mathcal{V}$ be a locally finite minimal variety whose finite members are solvable. Then it contains a nontrivial finite algebra, and so we can pick a nontrivial $\mathbf{S} \in \mathcal{V}$ having minimal size. Clearly, $\mathbf{S}$ is simple, and every proper subalgebra must be a singleton, in other words, $\mathbf{S}$ is strictly simple. By the minimality of $\mathcal{V}$, the algebra $\mathbf{S}$ generates $\mathcal{V}$. The type of $\mathbf{S}$ must be **1** or **2**, since $\mathbf{S}$ is solvable.

The main idea of the proof is to consider "multidimensional traces" in $\mathbf{S}$. Let N be a minimal set of $\mathbf{S}$ (which equals to a trace, since $\mathbf{S}$ is simple). Then the multitraces of $\mathbf{S}$ are subsets of the form

$$T = p(N, N, \ldots, N)\,,$$

where p is a polynomial of $\mathbf{S}$. If the traces are thought of as "lines" in the geometry of $\mathbf{S}$, then the multitraces are the higher dimensional planes of this geometry. The reader may look at the example of the module $\mathbf{F}^n$ over the matrix ring $\mathbf{F}^{n\times n}$ discussed above to justify this remark.

We claim that the multitraces are almost as well behaved as the traces themselves. Namely, the induced algebra on T is term equivalent to a matrix power of $\mathbf{S}|_N$, and T is still a neighborhood of $\mathbf{S}$.

To see why this is so, we try to *coordinatize* the set T. That is, we are looking for a k-ary polynomial f, and k unary polynomials $g_1, \ldots, g_k$ such that $T = f(N, \ldots, N)$, and for each $x_i \in N$ we have

$$g_i f(x_1, \ldots, x_n) = x_i\,.$$

It is not hard to calculate, using the definition of a matrix power, that the existence of these coordinate maps implies the statements in the previous paragraph.

To find these polynomials, consider the type **1** case first. We may assume that p depends on all variables on N, and in this case we show that $f = p$ works. As f depends on x_1, there exist elements $a, b \in N$, and a vector $\mathbf{c} \in N$ such that $f(a, \mathbf{c}) \neq f(b, \mathbf{c})$. By the property of separation, there is a unary polynomial g whose range is N such that $gf(a, \mathbf{c}) \neq gf(b, \mathbf{c})$. Thus, $gf(x_1, \ldots, x_k)$ is a basic operation of $\mathbf{S}|_N$, and it depends on its first variable on N. But $\mathbf{S}|_N$ is a permutation group, so this polynomial must not depend on any other variable, and must be a permutation in its first variable. By forming a suitable composition, we may assume that this permutation is the identity map, and so we get the equation above for an appropriate polynomial g_1. In the type **2** case we proceed similarly, but the calculation is more complicated, because we have to work with linear maps on a vector space.

We mention that multitraces are well-behaved in a more general setting, too. We may allow the type to be **3** as well, and may investigate a minimal congruence of any finite algebra, instead of looking at a simple algebra.

Now we return to the theorem on minimal varieties. As the induced algebras on multitraces are matrix powers, it is sufficient to prove that the whole algebra **S** is a multitrace. This proof is technical, but not too complicated, and this is the only place where the minimality of $\mathcal{V}$ is used. The user is directed to [9] for more details.

10. Conclusion

We now give a quick overview of the results on the RS-question and decidability that have been obtained using tame congruence theory. As we have mentioned earlier, varieties omitting types **1** and **5** are reasonably well-behaved.

THEOREM 10.1 (R. McKenzie, D. Hobby) *Let* **A** *be a finite algebra such that the types* **1** *and* **5** *do not occur in the finite members of the variety* $\mathcal{V}$ *generated by* **A**. *Then* $\mathcal{V}$ *is either residually large, or is congruence modular.*

Thus, based on the Freese-McKenzie theorem about the modular case, we can algorithmically decide if $\mathcal{V}$ is residually small (since there is a way to decide if $\mathcal{V}$ is modular or not). And although generally it is a very hard problem to find the set of types occurring in the variety generated by a finite algebra, it is effectively computable if the conditions of this theorem (on omitting **1** and **5**) are met.

So even at this level of generality we have the phenomenon that if $\mathcal{V}$ is a finitely generated variety, then either it is residually large, or there is a *finite* bound on the size of subdirectly irreducible algebras in $\mathcal{V}$ (or equivalently, there are only finitely many subdirectly irreducible algebras in $\mathcal{V}$). The conjecture that this must be so for every finitely generated variety, called the RS-conjecture, looks innocent enough, but it is actually a really hard problem, which has been open for more than twenty years. It is true for most "normal" varieties (as they are encompassed by the above theorem), and R. McKenzie has established it for semigroups also.

Very recently, R. McKenzie has proved important negative results for the general case. He presented various examples of finite algebras, each generating a residually small variety, such that these varieties nevertheless have infinitely many subdirectly irreducibles, some even infinite or uncountable ones. This refutes the RS-conjecture. R. McKenzie also proved that *there is no algorithm which inputs a finite algebra* **A**, *and computes whether there are only finitely many subdirectly irreducibles in the variety generated by* **A**, *or whether each subdirectly irreducible algebra in this variety is finite.* The

interested reader should look at Ross Willard's paper in this volume to find out more on this topic.

We conclude the discussion of the RS-question by mentioning two open problems (on which lots of work has already been done).

PROBLEM 10.2 Does the RS-conjecture hold for finitely generated varieties omitting type **5**? Is there an algorithm to compute the maximum size of a subdirectly irreducible algebra in such varieties (if the generating algebra is given)?

PROBLEM 10.3 (Quackenbush) If a finite algebra has finitely many basic operations, and there are arbitrarily large finite subdirectly irreducibles in the variety it generates, does it follow that there is an infinite subdirectly irreducible algebra in this variety?

The reader can prove as an exercise that if a locally finite variety has an infinite subdirectly irreducible algebra, then there is no bound on the size of its finite subdirectly irreducibles.

Now let us turn our attention to decidability. By ingenious proofs filling an entire book, R. McKenzie and M. Valeriote managed to lift the congruence modular result to the case of arbitrary locally finite varieties. Every locally finite, decidable variety $\mathcal{V}$ must decompose as a varietal product

$$\mathcal{V} = \mathcal{V}_1 \otimes \mathcal{V}_2 \otimes \mathcal{V}_3 \,,$$

where $\mathcal{V}_1$ is strongly abelian, $\mathcal{V}_2$ is affine, and $\mathcal{V}_3$ is a discriminator variety. The variety $\mathcal{V}$ is decidable if and only if all three components are decidable. We have already discussed the last two components. The first component is decidable if and only if it can be obtained using a matrix power construction from multisorted unary algebras in which the "left divisibility ordering" between nonconstant unary terms is linear. The reader is suggested to go to Matthew Valeriote's paper in this volume to read the precise details of this theorem, and about other related model theoretic applications of tame congruence theory.

11. Recommended reading

We have tried to avoid technicalities, including citations, in the text to improve readability. Therefore we give a mini-guide to the references here for the interested reader, who wants to learn more about these theories. To learn the basics about general algebraic structures the textbooks [2] and [12] are recommended. The main reference for the commutator is [3] (but see [4] for a different approach), while for tame congruence theory it is [5]. This book includes a non-technical overview of the theory. The reader may want to study Chapter 4 of this book parallel with [11]. To learn about

extensions of tame congruence theory, [8] and [9] can be studied. Finally the paper [1] and the book [13] contain the proofs of the decidability results. See also the papers [1] and [10] on boolean representations, [6] on categorical quasivarieties, and the manuscript [7], which contains a more detailed, non-technical outline of recent results in this area (and will probably appear in Algebra Universalis).

The references section below is not intended to be comprehensive. These books and papers list lots of other references, and their introductions refer to historical results that have been mentioned.

References

1. S. Burris and R. McKenzie. *Decidability and Boolean Representations*, volume 246 of *Memoirs of the American Mathematical Society*. American Mathematical Society, 1981.
2. S. Burris and H. P. Sankappanavar. *A Course in Universal Algebra*. Springer-Verlag, 1981.
3. R. Freese and R. McKenzie. *Commutator Theory for Congruence Modular Varieties*, volume 125 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, 1987.
4. H. P. Gumm. *Geometrical methods in congruence modular algebras*, volume 286 of *Memoirs of the American Mathematical Society*. American Mathematical Society, 1983.
5. D. Hobby and R. McKenzie. *The Structure of Finite Algebras*, volume 76 of *Contemporary Mathematics*. American Mathematical Society, 1988.
6. K. A. Kearnes. Categorical quasivarieties via Morita equivalence. Preprint, 1994.
7. K. A. Kearnes. Local methods in universal algebra. Manuscript, 1996.
8. K. A. Kearnes. An order-theoretic property of the commutator. *The International Journal of Algebra and Computation*, 3:491–534, 1993.
9. K. A. Kearnes, E. W. Kiss, and M. Valeriote. Minimal sets and varieties. Accepted by the Trans. Amer. Math. Soc., 1993.
10. E. W. Kiss. Finitely boolean representable varieties. *Proc. Amer. Math. Soc.*, 89:579–582, 1983.
11. E. W. Kiss. An easy way to minimal algebras. Accepted by the International Journal of Algebra and Computation, 1995.
12. R. McKenzie, G. McNulty, and W. Taylor. *Algebras, Lattices, Varieties Volume 1*. Wadsworth and Brooks/Cole, Monterey, California, 1987.
13. R. McKenzie and M. Valeriote. *The Structure of Locally Finite Decidable Varieties*, volume 79 of *Progress in Mathematics*. Birkhäuser Boston, 1989.