FINITE ALGEBRAS OF FINITE COMPLEXITY

KEITH A. KEARNES AND EMIL W. KISS

Abstract. We develop a new centrality concept and apply it to solve certain outstanding problems about finite algebras. In particular, we describe all finite algebras of finite complexity and all finite strongly abelian algebras which generate residually small varieties.

1. Introduction

The symbols for the basic operations of abelian groups are the binary sum $+$, unary negation $-$, and nullary unit 0. Using these symbols, there are many different ways to construct terms whose interpretation in any abelian group is subtraction. The most obvious way to express subtraction is with $x + (-y)$, but the term

$$((-z) + x) + ((z + 0) + ((-0) + (-y)))$$

interprets as the same operation in every abelian group. This term, whose composition tree is depicted in Figure 1, is a more complicated composition of basic operation symbols than $x + (-y)$. This is reflected by the fact that its composition tree has greater depth than the composition tree for $x + (-y)$.

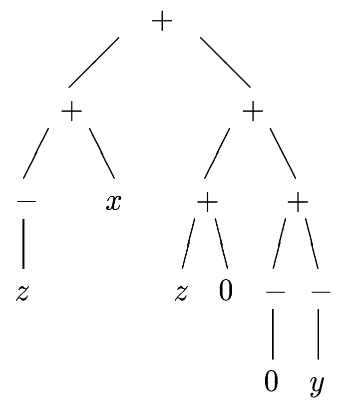

Figure 1: The composition tree for $((-z) + x) + ((z + 0) + ((-0) + (-y)))$.

1991 *Mathematics Subject Classification.* 08A05, 08B26.

Key words and phrases. Nilpotent algebra, tame congruence theory, finite essential arity, residually small variety.

Work supported by the Fields Institute (Toronto, Canada), the Humboldt Foundation, Germany, and the grants OTKA 16432 and FKFP 0877/1997 of Hungary.

We will use the depth of the composition tree as a measure of the *complexity of a term*. So, for example, $x+(-y)$ has complexity 2 while $((-z)+x)+((z+0)+((-0)+(-y)))$ has complexity 4. For a given term operation, like subtraction, there will usually be many different terms which interpret as that operation. We define the *complexity of a term operation* to be the minimum of the complexities of all terms which interpret as that operation. (Subtraction has complexity 2 for most abelian groups. However, subtraction has complexity 1 for any nontrivial abelian group satisfying the equation $y = -y$, since then $x + y$ interprets as subtraction. Subtraction has complexity 0 for the one-element group.) If $\mathbf{A}$ is an algebra in a finite language, then we define the *complexity of* $\mathbf{A}$ to be the ordinal which is the supremum of the complexities of all of its term operations. This ordinal is finite or ω. This paper contains an analysis of the structure of finite algebras which have finite complexity.

Saying that a finite algebra $\mathbf{A}$ in a finite language has finite complexity means that there is a number N such that for any term of $\mathbf{A}$ which has a complicated nesting of subterms (i.e., a deep composition tree) one can always replace the term with an equivalent one whose composition tree has depth less than N. Thus, there exists a bound K, computable from N and the arities of the basic operations, such that each term operation can be obtained as an interpretation of a term having at most K variables. In other words, every term operation may depend on at most K of its variables. It is not too hard to see that this bounds the growth rate of the number of inequivalent term operations of $\mathbf{A}$ as a function of the number of variables. The number of inequivalent n-ary term operations of $\mathbf{A}$ equals the size of the n-generated free algebra in the variety $\mathcal{V}(\mathbf{A})$ generated by $\mathbf{A}$, so if we define the *free spectrum* of $\mathbf{A}$ to be the function $f_{\mathcal{V}(\mathbf{A})}(n) = |F_{\mathcal{V}(\mathbf{A})}(n)|$ then, if $\mathbf{A}$ has a finite language, $\mathbf{A}$ has finite complexity if and only if the free spectrum of $\mathbf{A}$ is bounded from above by a polynomial.

As is witnessed very clearly in the case of groups, the rate of growth of the free spectrum function is intimately related with fundamental structural properties. For example, a finite group whose free spectrum function grows at a rate which is less than doubly exponential must be nilpotent. Furthermore, as is proved in [3] and [12], it is nilpotent of class $\leq k$ if and only its free spectrum function is bounded above by $2^{p(n)}$ for some polynomial $p(n)$ of degree $\leq k$. In this paper the fundamental structural properties associated with a slow-growing free spectrum function are our main concern, not the spectrum function itself; the function interests us only to the extent that bounding its growth rate helps us to discover these structural properties. Just as free spectrum problems and other problems about groups might force one to discover the significance of nilpotence in group theory, and more generally of the group commutator operation, so we have been forced by a circle of problems to isolate a new kind of nilpotence which is associated with a new kind of commutator operation.

By a *commutator operation* we mean a certain kind of binary operation on the congruence lattice of an algebra. For groups the commutator of normal subgroups is a commutator operation; for commutative rings the product of two ideals is a commutator operation. These special examples are generalized in Chapter 3 of [4] to a commutator operation definable for any algebra. This definition is based on a centrality concept we call *normal centrality*. From the normal centralizer relation one defines the normal commutator, and at once has concepts of solvability, nilpotence and abelianness at hand. A new centralizer relation is defined in [5] called the *weak centralizer*, and along with that concept there is a weak commutator and notions of weak solvability, weak nilpotence and weak abelianness. Also appearing in [4] is the notion of a *strongly abelian* algebra or congruence. This notion has never before been associated with a centralizer relation or a commutator operation, but in Section 2 we explain how this can be done. The main focus of this paper is yet a fourth kind of centrality which we call the *rectangular centralizer*. Associated to the rectangular centralizer is a commutator and concepts of rectangular solvability, rectangular nilpotence and rectangular abelianness.

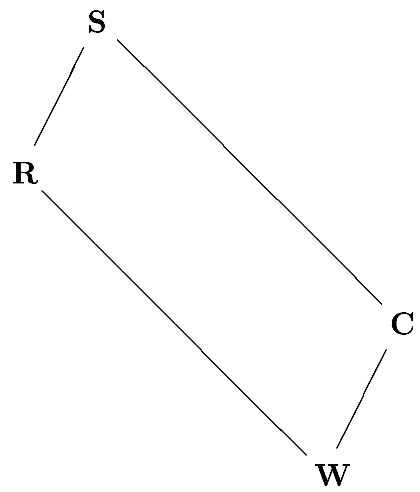

FIGURE 2: **C** = normal, **W** = weak, **R** = rectangular, **S** = strong.

The four centralizer concepts are related to each other as in Figure 2. What the order in Figure 2 is meant to suggest is that the strongest (most restrictive) centralizer condition that one can impose on a pair of congruences, among the centralities mentioned, is the strong centralizer condition, and the weakest is the weak centralizer condition. Normal and weak centrality seem to be closely related, as do strong and rectangular centrality. Parallel edges in Figure 2 are meant to suggest that the relationships seem to be parallel. The relations **R** and **C** seem to be complementary concepts within this interval. Strong centrality is the conjunction of rectangular and normal centrality by definition. Weak centrality seems to behave like the disjunction of rectangular and normal centrality.

The abelianness concepts that arise from the four types of centrality are essentially different from one another. However the concepts of strong and rectangular solvability coincide for finite algebras, and similarly the concepts of normal and weak solvability coincide. The same statements are true if we replace solvability by nilpotence. In this paper we are concerned almost exclusively with rectangular centrality and the associated abelianness and nilpotence concepts. But instead of choosing the obvious phrase "rectangularly abelian" for the abelian concept, we will tend toward more euphonius terminology, and say simply that an algebra is *rectangular* when the rectangular commutator operation is trivial. Furthermore, rather than call an algebra "rectangularly nilpotent" we prefer the sound of "strongly nilpotent", and we will use the latter after proving the equivalence of these two concepts in Lemma 3.4.

Here is a summary of the results of the paper. In Section 2 we give the full definitions of the various centralizers and prove some of their basic properties, with most of our attention devoted to determining when a tolerance rectangularly centralizes a prime quotient. Section 3 further develops the rectangular centralizer and includes a proof of the equivalence of strong and rectangular nilpotence. Lemma 3.4 of this section gives many equivalent formulations of strong nilpotence, including local and equational characterizations. Section 4 uses a Ramsey argument to prove that a finite algebra in a finite language has finite complexity if and only if it is strongly nilpotent. Sections 5, 6 and 7 are about rectangular algebras and their structure. Specifically, Section 5 begins by solving a combinatorial problem about partitioning rectangles, and then it applies the solution to obtain a tight bound on the essential arity of a finite rectangular algebra. Section 6 proves a representation theorem for rectangular algebras. Section 7 gives a Klukovits-type characterization of the clone of a rectangular variety, which we apply to show that any locally finite rectangular variety is finitely generated. Our final section, Section 8, includes the characterization of strongly nilpotent, locally finite varieties which are residually small. The main result states that such a variety is residually small if and only if its algebras are rectangular. This yields an algorithm to decide if a finite strongly nilpotent (in particular, strongly abelian) algebra in a finite language generates a residually small variety. Example 8.7 presents a finite, simple, strongly abelian algebra generating a residually large variety.

The notation and terminology of this paper is standard, and for the most part follows [4]. Algebras are written in bold face, as in $\mathbf{A}, \mathbf{B}, \mathbf{C}, \dots$. The underlying set (or any set without structure) will be written in italics, as in $A, B, C, \dots$. A sequence of elements $(a_1, a_2, \dots)$ of any length will frequently be denoted by a bold face lower case character, as in $\mathbf{a}$, for example. We will not specify the length if it is irrelevant or can be determined from the context. If $\mathbf{A}$ is an algebra and R is a binary relation on $\mathbf{A}$, then we may use the notation $a\ R\ b$, as well as the usual notation $(a,b) \in R$, to denote the fact that a is R-related to b. As an extension of this convention, by the notation $\mathbf{a}\ R\ \mathbf{b}$ we mean that the vectors $\mathbf{a}$ and $\mathbf{b}$ are R-related componentwise, that is, $a_i\ R\ b_i$ for every i. If $R = \alpha$ is a congruence we may also use the notation $a \equiv_\alpha b$ to denote that $(a,b) \in \alpha$. The factor notation $\mathbf{A}/\alpha$ will be extended so that it applies to relations on $\mathbf{A}$. Therefore, if $a \in A$, then a/α is the coset of α containing a, and if R is an n-ary relation, then

$$R/\alpha := \{(a_1/\alpha, \dots, a_n/\alpha) \mid (a_1, \dots, a_n) \in R\}.$$

This will most often be used when the relation in question is a *tolerance*, that is, a compatible, reflexive, symmetric binary relation. If T is a tolerance on $\mathbf{A}$, then T/α is again a tolerance (on $\mathbf{A}/\alpha$), and it equals the tolerance $(\alpha \circ T \circ \alpha)/\alpha$. If β is a congruence (which is nothing more than a transitive tolerance), then β/α is a tolerance. But if β contains α, then β/α is a congruence. If t is a term in the language of $\mathbf{A}$, then we may also use t to denote the term operation of $\mathbf{A}$ represented by t if there is no danger of confusion or need to make a distinction. If it seems wise to make the distinction, then we will write $t^{\mathbf{A}}$ for the interpretation of t in $\mathbf{A}$.

Some of the statements and proofs in Sections 2 and 3 use tame congruence theory. The monograph [4] is the handbook for this theory (see [6] as a companion).

This work was started while the authors were visiting the Technische Hochschule in Darmstadt, Germany, in 1995 and completed while the authors were visiting the Fields Institute in Toronto, Canada in 1996. We gratefully acknowledge the hospitable environment for research and the financial support provided by both institutions.

2. What is Rectangularity?

All the concepts of centrality mentioned in the Introduction can be expressed in the following way. Let L and R be symmetric binary relations of an algebra $\mathbf{A}$. By an L, R-*matrix* we mean a matrix of the form

$$\begin{bmatrix} t(\mathbf{a},\mathbf{c}) & t(\mathbf{a},\mathbf{d}) \\ t(\mathbf{b},\mathbf{c}) & t(\mathbf{b},\mathbf{d}) \end{bmatrix},$$

where t is a polynomial of $\mathbf{A}$, and $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$ are vectors of $\mathbf{A}$ such that $\mathbf{a}\ L\ \mathbf{b}$ and $\mathbf{c}\ R\ \mathbf{d}$. These matrices form a subalgebra in $\mathbf{A}^4$, which is generated by the set

$$\left\{ \begin{bmatrix} a & a \\ b & b \end{bmatrix} \;\middle|\; (a,b) \in L \right\} \cup \left\{ \begin{bmatrix} c & d \\ c & d \end{bmatrix} \;\middle|\; (c,d) \in R \right\}$$

of *trivial* L, R-matrices, and the diagonal of A^4.

Now let δ be a congruence of $\mathbf{A}$. Following [4] we say that L *(normally) centralizes* R *modulo* δ, or that $\mathbf{C}(L,R;\delta)$ *holds*, if for every L,R-matrix

$$\begin{bmatrix} t(\mathbf{a},\mathbf{c}) & t(\mathbf{a},\mathbf{d}) \\ t(\mathbf{b},\mathbf{c}) & t(\mathbf{b},\mathbf{d}) \end{bmatrix} = \begin{bmatrix} u & v \\ w & z \end{bmatrix}$$

we have $u \equiv_\delta v$ if and only if $w \equiv_\delta z$. We now define the three other concepts of centrality mentioned in the Introduction.

Definition 2.1. Let L and R be symmetric binary relations of an algebra $\mathbf{A}$, and δ a congruence of $\mathbf{A}$. We say that

(1) $\mathbf{W}(L,R;\delta)$ *holds* if for every L,R-matrix, if three of its elements are δ-related, then all four elements are δ-related;
(2) $\mathbf{R}(L,R;\delta)$ *holds* if for every L,R-matrix, if the two elements on the main diagonal are δ-related, then all four elements are δ-related;
(3) $\mathbf{S}(L,R;\delta)$ *holds* if and only if $\mathbf{C}(L,R;\delta)$ and $\mathbf{R}(L,R;\delta)$ both hold.

Extending the usual terminology from $\mathbf{C}$ to $\mathbf{W}$ and $\mathbf{S}$, we will express the fact that $\mathbf{W}(L,R;\delta)$ or $\mathbf{S}(L,R;\delta)$ holds by saying that L *weakly* or *strongly centralizes* R *modulo* δ. We will express $\mathbf{R}(L,R;\delta)$ by saying that L *rectangulates* R *modulo* δ. The relation L is called *rectangular* if $\mathbf{R}(L,L;0_{\mathbf{A}})$ holds, and $\mathbf{A}$ is called rectangular if the relation $1_{\mathbf{A}} = A \times A$ is rectangular.

Consider the L,R-matrices of $\mathbf{A}$ of the form

$$\begin{bmatrix} t(\mathbf{a},\mathbf{c}) & t(\mathbf{a},\mathbf{d}) \\ t(\mathbf{b},\mathbf{c}) & t(\mathbf{b},\mathbf{d}) \end{bmatrix} = \begin{bmatrix} u & v \\ w & z \end{bmatrix}$$

where $u = z$. We shall denote by $\mathbf{R}(L,R)$ the set of all pairs (u,v) which occur on the first row of such a matrix. The relation $\mathbf{R}(L,R)$ is a compatible, reflexive relation of $\mathbf{A}$, but is highly asymmetric in general, even when $L = R = A \times A$.

We list some very elementary properties of these concepts. We leave the proof of this result to the reader.

Lemma 2.2. *Let L and R be symmetric binary relations of an algebra $\mathbf{A}$, and δ a congruence of $\mathbf{A}$.*

(1) $\mathbf{R}(L,R;\delta) \Longleftrightarrow \mathbf{R}(R,L;\delta)$.
(2) *We have $\mathbf{R}(L,R;\delta)$ if and only if for every L,R-matrix if the two elements in the main diagonal are δ-related, then the elements in the top row are δ-related.*
(3) $\mathbf{R}(L,R;0_{\mathbf{A}}) \Longleftrightarrow \mathbf{R}(L,R) = 0_{\mathbf{A}}$.
(4) *If δ_i $(i \in I)$ are congruences of $\mathbf{A}$, then $\mathbf{R}(L,R;\delta_i)$ for all $i \in I$ implies $\mathbf{R}(L,R;\bigwedge_{i\in I}\delta_i)$.*
(5) *If γ is a congruence and $\gamma \le \delta$, then we have $\mathbf{R}(L,R;\delta) \Longrightarrow \mathbf{R}(L/\gamma,R/\gamma;\delta/\gamma)$ in $\mathbf{A}/\gamma$.*
(6) *If $L' \subseteq L, R' \subseteq R$ and $\mathbf{R}(L,R;\delta)$ holds, then $\mathbf{R}(L',R';\delta)$ holds.*
(7) *If L and R are compatible, $R = \delta \circ R \circ \delta$ and $L \cap R \subseteq \delta$, then $\mathbf{R}(L,R;\delta)$ holds.*

(8) *If θ is a congruence of* $\mathbf{A}$*, then* $\mathbf{R}(L,\theta;\delta)\Longrightarrow\mathbf{R}(L,\theta;\theta\wedge\delta)$.

(9) $\mathbf{A}$ *is rectangular if and only if for every n-ary polynomial t of* $\mathbf{A}$*, and every element $a\in A$, the set* $\{(a_1,\dots,a_n)\in A^n \mid t(a_1,\dots,a_n)=a\}$ *is a* rectangular subset *of A^n; that is, it can be written in the form $A_1\times\cdots\times A_n$, where the A_i are subsets of A.*

(10) *Each of* $\mathbf{R}(L,R;\delta)$ *and* $\mathbf{C}(L,R;\delta)$ *implies* $\mathbf{W}(L,R;\delta)$.

(11) $\mathbf{S}(\alpha,\alpha;0_{\mathbf{A}})$ *holds for a congruence α of* $\mathbf{A}$ *if and only if α is* strongly abelian *in the sense (studied in* [4]*) that whenever $t(x,\mathbf{y})$ is a polynomial of* $\mathbf{A}$*, $a\equiv_\alpha b\equiv_\alpha c$ and $\mathbf{u}\equiv_\alpha\mathbf{v}$, then* $t(a,\mathbf{u})=t(b,\mathbf{v})\Longrightarrow t(c,\mathbf{u})=t(c,\mathbf{v})$.

We especially encourage the reader to verify (9), which explains the name of rectangular centrality. The statement in (4) allows us to define the *rectangular commutator of L and R* to be the smallest δ for which $\mathbf{R}(L,R;\delta)$ holds. It is denoted by $[L,R]_{\mathbf{R}}$. Thus we can speak about rectangular solvability, and rectangular nilpotence as well. One can similarly define the weak and the strong commutator $[L,R]_{\mathbf{W}}$ and $[L,R]_{\mathbf{S}}$.

If R is a symmetric binary relation on $\mathbf{A}$, then two polynomials $f(\mathbf{x})$ and $g(\mathbf{x})$ are called *R-twins* if there exists a polynomial $h(\mathbf{x},\mathbf{y})$ and vectors $\mathbf{u}\ R\ \mathbf{v}$ such that $f(\mathbf{x})=h(\mathbf{x},\mathbf{u})$ and $g(\mathbf{x})=h(\mathbf{x},\mathbf{v})$ holds for every $\mathbf{x}$. Consider the set of all permutations of A that are R-twins of the identity map. It is easy to see that this set is closed under composition. When $\mathbf{A}$ is finite, this is a group of permutations called the *R-twin group of* $\mathbf{A}$, and it is denoted by $\mathrm{Tw}(\mathbf{A},R)$. The reader can verify that $\mathrm{Tw}(\mathbf{A},R)$ is a normal subgroup of the group consisting of all permutations of $\mathbf{A}$ that are unary polynomials. We extend twin group terminology to E-traces as follows. If N is an E-trace (i.e., the intersection of a congruence class and the range of an idempotent polynomial), then the R-twin group on N is the group of restrictions to N of polynomials which map N bijectively onto N and which are R-twins of the identity on N. This group is denoted by $\mathrm{Tw}(\mathbf{A}|_N,R)$.

Recall from [4] that if $\mathbf{A}$ is an algebra, then a **1**-*snag of* $\mathbf{A}$ is a pair $(a,b)\in A^2$ such that $a\neq b$ and for some binary polynomial $p(x,y)$ of $\mathbf{A}$ we have

$$\begin{bmatrix} p(a,a) & p(a,b) \\ p(b,a) & p(b,b) \end{bmatrix} = \begin{bmatrix} * & a \\ a & b \end{bmatrix}.$$

If we let $f(x)=p(b,x)$, $g(x)=p(a,x)$ and $R=\{(a,b),(b,a)\}$, then f and g are R-twins and $f(a)=a$, $f(b)=b$ and $g(b)=a$. This property for more general twins is a useful concept, so for any symmetric binary relation R we say that (a,b) is a $\mathbf{1},R$-*snag of* $\mathbf{A}$ if $a\neq b$ and there exist R-twin unary polynomials f and g of $\mathbf{A}$ such that $f(a)=a$, $f(b)=b$ and $g(b)=a$. In this paper we shall use this concept frequently in the situation when R is a tolerance.

Our first aim is to understand what it means for $\mathbf{R}(T,\theta;\delta)$ to hold when T is a tolerance and $\langle\delta,\theta\rangle$ is a "prime quotient" (meaning that $\delta\prec\theta$ is a covering pair of congruences). For the following series of lemmas we fix the following notation. Let

$\mathbf{A}$ be a finite algebra, $\delta \prec \theta$ a fixed covering pair of congruences of $\mathbf{A}$, and N a $\langle \delta, \theta \rangle$-trace. (A $\langle \delta, \theta \rangle$*-trace* is a minimal E-trace for θ which is not an E-trace for δ. For an alternate definition, see Definition 2.25 of [4]. Chapters 2–5 of that book describe several properties of traces, although we will not need to know very many of these properties here.) Let T be a tolerance of $\mathbf{A}$. First we show that if $\mathbf{R}(T, \theta; \delta)$ holds and T "intersects" $\theta - \delta$, meaning that $T \cap \theta \not\subseteq \delta$, then the type must be $\mathbf{1}$.

Lemma 2.3. *If $\mathbf{R}(T, \theta; \delta)$ holds, and T intersects $\theta - \delta$, then the type of $\langle \delta, \theta \rangle$ is $\mathbf{1}$. On the other hand, if T does not intersect $\theta - \delta$, then $\mathbf{R}(T, \theta; \delta)$ holds.*

Proof. If there is a pair of elements in $(a, b) \in (T \cap \theta) - \delta$, then according to Theorem 2.8 (4) of [4] we can map this pair by a unary polynomial to a new pair of elements $(p(a), p(b)) \in (T \cap \theta) - \delta$ where both $p(a)$ and $p(b)$ belong to a $\langle \delta, \theta \rangle$-minimal set. Since $(p(a), p(b)) \in \theta - \delta$, both $p(a)$ and $p(b)$ lie in a $\langle \delta, \theta \rangle$-trace. By Corollary 5.2 (2) of [4], all $\langle \delta, \theta \rangle$-traces are polynomially isomorphic, so there is no loss of generality in assuming that $p(a)$ and $p(b)$ belong to N. Changing notation back, if there is a pair $(a, b) \in (T \cap \theta) - \delta$, then we may select such a pair with $a, b \in N$. In all types other than $\mathbf{1}$, any two elements of a $\langle \delta, \theta \rangle$-trace that are not δ-related form a $\mathbf{1}$-snag. But if N has a $\mathbf{1}$-snag $(a, b) \in T$, then the matrix

$$\begin{bmatrix} p(b,a) & p(b,b) \\ p(a,a) & p(a,b) \end{bmatrix} = \begin{bmatrix} a & b \\ * & a \end{bmatrix}$$

(obtained by interchanging rows in the definition of a $\mathbf{1}$-snag), together with $a \not\equiv_\delta b$ clearly contradict $\mathbf{R}(T, \theta; \delta)$. The other statement follows from Lemma 2.2 (7). ☐

Lemma 2.4. *The following statements are equivalent.*

(1) $\mathbf{R}(T, \theta; \delta)$ *holds.*
(2) $\mathbf{R}(T, N^2; \delta)$ *holds.*
(3) *There is a congruence $\gamma \le \delta$ satisfying $\mathbf{R}(T, N^2; \gamma)$.*
(4) *There is no $\mathbf{1}, T$-snag in $N^2 - \delta$.*
(5) *Every T-twin of the identity map which maps N into N equals the identity map modulo δ on N.*
(6) *The T-twin group on N/δ is trivial, and $\mathbf{W}(T, N^2; \delta)$ holds.*
(7) *Every T-twin of a permutation f of N which maps N into N equals f modulo δ on N.*
(8) *For any two T-twin unary polynomials mapping a θ-class C to N, either they are equal modulo δ on C, or both collapse C into δ.*
(9) *For any two T-twin polynomials mapping a product $C = C_1 \times \cdots \times C_k$ of θ-classes to N, either they are equal modulo δ on C, or both collapse C into δ.*
(10) *If $\begin{bmatrix} u & v \\ w & z \end{bmatrix}$ is a T, θ-matrix with all entries in N, then this matrix is trivial modulo δ, that is, either $u \equiv_\delta v$ and $w \equiv_\delta z$, or else $u \equiv_\delta w$ and $v \equiv_\delta z$.*

Proof. (1) $\Longrightarrow$(2). By Lemma 2.2 (6).

(2) $\Longrightarrow$(3). Take $\gamma = \delta$.

(3) $\Longrightarrow$(4). (The idea of the proof of Lemma 2.3 works here.) Let (a, b) be a $\mathbf{1}, T$-snag in $N^2 - \delta$ with respect to the T-twin polynomials f and g. Then the T, N^2-matrix

$$\begin{bmatrix} f(a) & f(b) \\ g(a) & g(b) \end{bmatrix} = \begin{bmatrix} a & b \\ * & a \end{bmatrix}$$

is a failure of $\mathbf{R}(T, N^2; \gamma)$ for any $\gamma \leq \delta$.

(4) $\Longrightarrow$(5). Suppose that $\mathbf{u}$ and $\mathbf{v}$ are T-related, $h(x, \mathbf{u}) = x$ for every $x \in N$, and $h(N, \mathbf{v}) \subseteq N$. Let $b \in N$ and $a = h(b, \mathbf{v}) \in N$. Then (a, b) is a $\mathbf{1}, T$-snag, contradicting (4) unless $a \; \delta \; b$. Thus $h(x, \mathbf{v})$ is also the identity map modulo δ.

(5) $\Longrightarrow$(7). Let $h(x, \mathbf{u})$ and $h(x, \mathbf{v})$ be T-twins, and assume that $f(x) = h(x, \mathbf{u})$ is a permutation of N. Then its inverse can be obtained in the form f^k for some k. Clearly, $f^k \circ f$ and $f^k \circ h(x, \mathbf{v})$ are still T-twins, and the first one is the identity map. So these are δ-related by (5). Since f^k is a permutation of N, we get $h(x, \mathbf{u}) \; \delta \; h(x, \mathbf{v})$.

(7) $\Longrightarrow$(8). Let f and g be T-twin unary polynomials of $\mathbf{A}$ mapping C into N. Let $M \subseteq C$ be any $\langle \delta, \theta \rangle$-trace. By composing f and g with a polynomial isomorphism mapping N to M we get from (7) that either $f|_M$ and $g|_M$ are equal modulo δ, or they both map M into a δ-class.

Now suppose that f does not collapse C to a δ-class, that is, $f(a)$ and $f(b)$ are not δ-related for some $a, b \in C$. We want to show that $f(x) \; \delta \; g(x)$ for every $x \in C$. Connect a and b by traces. There is a trace N' in this chain that is not collapsed to δ by f. Hence our remark applied to $M = N'$ implies that $f|_{N'}$ and $g|_{N'}$ are equal modulo δ. Now let N'' be any trace that overlaps with N' modulo δ. Then $f|_{N''}$ and $g|_{N''}$ are equal modulo δ, even if both of these functions collapse N'' into δ, because of the overlap. So connecting N' to x via traces we see that $f(x) \; \delta \; g(x)$ indeed.

(8) $\Longrightarrow$(9). Let f and g be T-twin polynomials of $\mathbf{A}$ mapping C into N. Clearly, $f(\mathbf{x})$ and $g(\mathbf{x})$ are T-related for every $\mathbf{x} \in C$. Thus if $T \cap N^2 \subseteq \delta$, then we are done. If this is not the case, then by Lemma 2.3 we see that the type of this quotient is $\mathbf{1}$. Hence, both f and g may depend on at most one variable on C modulo δ. But (8) shows that if f does not depend on a variable, then g does not depend on it either. Hence they must depend on the same variable (or on no variables at all), and we are done by (8).

(9) $\Longrightarrow$(10). Suppose that there exists a modulo δ nontrivial T, θ-matrix in N. By switching rows and columns if necessary, we may assume that this matrix has the form

$$\begin{bmatrix} u & v \\ w & z \end{bmatrix} = \begin{bmatrix} t(\mathbf{a}, \mathbf{c}) & t(\mathbf{a}, \mathbf{d}) \\ t(\mathbf{b}, \mathbf{c}) & t(\mathbf{b}, \mathbf{d}) \end{bmatrix},$$

where $u, v, w, z \in N$, $\mathbf{a}\, T\, \mathbf{b}$, $\mathbf{c}\ \theta\ \mathbf{d}$, but $u \not\equiv_\delta v \not\equiv_\delta z$. Look at the polynomials $f(\mathbf{x}) = t(\mathbf{a}, \mathbf{x})$ and $g(\mathbf{x}) = t(\mathbf{b}, \mathbf{x})$. Both map $C = C_1 \times \cdots \times C_n$ to N, where $C_i = c_i/\theta = d_i/\theta$. But f does not collapse C to δ because of the first row, and it does not equal g on C modulo δ because of the second column. This contradiction with (9) proves (10).

(10) $\Longrightarrow$(1). Suppose that there exists a T, θ-matrix where the diagonal is δ-related, but the top row is not. Both the top row and the bottom row are θ-related, and so the entire matrix lies in a single θ-class C. Theorem 2.8 (4) of [4] guarantees that we can choose a unary polynomial p that maps C to N, but does not map the pair in the top row to δ. This way, we get a new failure of $\mathbf{R}(T, \theta; \delta)$ where the corresponding matrix lies entirely in N, but the diagonal is in δ while the top row is not. Therefore this matrix is nontrivial modulo δ, contradicting (10).

We have now established that all statements are equivalent, with the exception of (6). From (5) we deduce the part of (6) which refers to the twin group. From (2) and Lemma 2.2 (10) we deduce the part of (6) which refers to weak centrality. To finish the proof it is sufficient to show (6) $\Longrightarrow$(5). Suppose that $\mathbf{u}$ and $\mathbf{v}$ are T-related, $f(x) = h(x, \mathbf{u}) = x$ for every $x \in N$, and $g(x) = h(x, \mathbf{v})$. Assume that $g(N) \subseteq N$. If g is a permutation on N we are done, since then g is a twin of the identity and (6) asserts that the twin group is trivial. If g collapses N to a $\delta|_N$-class $D \subseteq N$, then let $b \in N - D$ be outside this class and let $a = g(b) \in D$. The T, N^2-matrix

$$\begin{bmatrix} f(a) & f(b) \\ g(a) & g(b) \end{bmatrix} = \begin{bmatrix} a & b \\ a' & a \end{bmatrix},$$

where $b \not\equiv_\delta a \equiv_\delta a'$, is clearly a failure of $\mathbf{W}(T, N^2; \delta)$. □

Here are some easy consequences of this lemma.

Lemma 2.5. *The following statements hold.*

(1) $\mathbf{R}(T, \theta; \delta) \Longrightarrow \mathbf{C}(\theta, T; \delta)$.
(2) *A congruence of a finite algebra is rectangularly solvable if and only if it is strongly solvable.*

Proof. To see (1) suppose that there exists a θ, T-matrix M such that

$$M = \begin{bmatrix} u & v \\ w & z \end{bmatrix},$$

and $u \equiv_\delta v$ while $w \not\equiv_\delta z$. Since both columns are θ-related, we can map the whole matrix M into a trace N with a unary polynomial which keeps the bottom row in $\theta - \delta$ although the top row will still be in δ. The transpose of this matrix is a modulo δ nontrivial T, θ-matrix in N, which contradicts Lemma 2.4 (10). Therefore the first statement of this lemma is proved.

To show the second statement, first note that if a congruence α of a finite algebra $\mathbf{A}$ is strongly solvable, then it is rectangularly solvable, too, since strong centrality implies rectangular centrality. Conversely, if α is rectangularly solvable, then there is a chain $0_{\mathbf{A}} = \alpha_0 \leq \alpha_1 \leq \cdots \leq \alpha_n = \alpha$ of congruences of $\mathbf{A}$ such that we have $\mathbf{R}(\alpha_{i+1}, \alpha_{i+1}; \alpha_i)$ for every i. Let $\alpha_i \leq \delta \prec \theta \leq \alpha_{i+1}$ be any prime quotient. Then we have $\mathbf{R}(\theta, \theta; \alpha_i)$, so by Lemma 2.4 (3) we get $\mathbf{R}(\theta, \theta; \delta)$. Thus, by Lemma 2.3 we see that the type of $\langle \delta, \theta \rangle$ is **1**. Therefore we can build a chain of type **1** quotients from $0_{\mathbf{A}}$ to α showing that α is strongly solvable. □

We remark that, while $\mathbf{R}(L, R; \delta)$ is symmetric in its first two variables, the relation $\mathbf{C}(L, R; \delta)$ is not. The order of the relations in Lemma 2.5 (1) is therefore critical, since it is not true that $\mathbf{R}(T, \theta; \delta)$ implies $\mathbf{C}(T, \theta; \delta)$. To see this, let S be the two-element left zero semigroup and $\mathbf{A} = S^0$ the semigroup obtained from S by adding 0 as an absorbing element. Then for the only nontrivial congruence θ of $\mathbf{A}$ (which has blocks S and $\{0\}$) we have $0_{\mathbf{A}} \prec \theta$ and $\mathbf{R}(1_{\mathbf{A}}, \theta; 0_{\mathbf{A}})$, but we do not have $\mathbf{C}(1_{\mathbf{A}}, \theta; 0_{\mathbf{A}})$.

3. Rectangular and Strong Nilpotence

We shall now turn our attention from centrality to nilpotence. We shall find that all possible concepts of nilpotence with respect to strong centrality coincide with rectangular nilpotence. We shall characterize this concept in several ways, and use these characterizations throughout the paper.

Since the normal commutator is not symmetric, one can speak about left and right nilpotence in general. One can even have a mixed expression like

$$[[\alpha, [\alpha, [[[\alpha, \alpha], \alpha], \alpha]]], \alpha] = 0_{\mathbf{A}} ,$$

which is still another possible type of nilpotence for α. We can reformulate this by saying that $\mathbf{A}$ has a chain $0_{\mathbf{A}} = \alpha_0 \leq \alpha_1 \leq \cdots \leq \alpha_n = \alpha$ of congruences such that each quotient α_{i+1}/α_i in this chain is centralized either from the left or from the right by α. Here "centralize" may mean any of the centrality concepts, like normal, weak, strong, or rectangular centrality, with respect to which we are investigating nilpotence.

Now suppose that α is a strongly nilpotent congruence of $\mathbf{A}$ in the weakest possible sense described above, that is, for a suitable chain of congruences we have either $\mathbf{S}(\alpha, \alpha_{i+1}; \alpha_i)$ or $\mathbf{S}(\alpha_{i+1}, \alpha; \alpha_i)$ for every i. In any case we have $\mathbf{R}(\alpha, \alpha_{i+1}; \alpha_i)$, so α is rectangularly nilpotent. Since rectangularity is symmetric, we have only one concept of rectangular nilpotence. We shall show that conversely, if α is rectangularly nilpotent, then it *strongly* centralizes *every* prime quotient of $\mathbf{A}$, *both from left and right.*

For applications to the study of residually small varieties we need to state some of our results for tolerances. The following lemma suggests an appropriate definition of nilpotence for tolerances.

Lemma 3.1. *Let* $\mathbf{A}$ *be any algebra, and* α *a congruence of* $\mathbf{A}$. *Then* α *is rectangularly nilpotent if and only if there is a chain* $0_{\mathbf{A}} = \alpha_0 \leq \alpha_1 \leq \cdots \leq \alpha_k = 1_{\mathbf{A}}$ *of congruences of* $\mathbf{A}$ *such that each quotient* $\langle \alpha_{i-1}, \alpha_i \rangle$ *is rectangulated by* α*; that* *is,* $\mathbf{R}(\alpha, \alpha_i; \alpha_{i-1})$ *holds for every* i*. The analogous statement holds for left and right nilpotence with regard to other types of centrality as well.*

Proof. The fact that α is rectangularly nilpotent implies that there is a chain of congruences connecting $0_{\mathbf{A}}$ to α such that each quotient is rectangulated by α. Add $1_{\mathbf{A}}$ to the end of this chain. As $\mathbf{R}(\alpha, 1_{\mathbf{A}}; \alpha)$ obviously holds, we get a chain like the one in the statement. Conversely, if some chain α_i connects $0_{\mathbf{A}}$ to $1_{\mathbf{A}}$ such that every quotient is rectangulated by α, then the chain $\alpha \wedge \alpha_i$ connects $0_{\mathbf{A}}$ to α, and still every quotient is rectangulated by α. The same argument works for all other types of centrality. □

Now we explain what we will mean for a tolerance T to be left or right nilpotent in one of our senses. For example, if we are referring to strong centrality we will say that T is *strongly left nilpotent* if there is a chain $0_{\mathbf{A}} = \alpha_0 \leq \alpha_1 \leq \cdots \leq \alpha_k = 1_{\mathbf{A}}$ of congruences of $\mathbf{A}$ such that $\mathbf{S}(T, \alpha_{i+1}; \alpha_i)$ holds for all i; we say T is *strongly right nilpotent* if there is a chain $0_{\mathbf{A}} = \beta_0 \leq \beta_1 \leq \cdots \leq \beta_\ell = 1_{\mathbf{A}}$ of congruences such that $\mathbf{S}(\beta_{i+1}, T; \beta_i)$ holds for all i. We allow ourselves to replace strong centrality with any other form of centrality in this definition. Since the rectangular and weak centralizer relations are symmetric with respect to the first two variables, it is apparent that left and right nilpotence agree for these two centralities. We will learn that for finite algebras strong left and right nilpotence agree as well. Ordinary left and right nilpotence do not agree. (See [5]).

For our later discussion of nilpotent radicals we shall need a refinement of the snag concept. For the appropriate definition, recall from [1] that a $\langle \delta, \theta \rangle$*-subtrace* of $\mathbf{A}$ is any pair $\{a, b\}$ of elements of $\mathbf{A}$ such that a, b are contained in some $\langle \delta, \theta \rangle$-trace, but a and b are not δ-related. A *subtrace* is a $\langle \delta, \theta \rangle$-subtrace for some prime quotient $\langle \delta, \theta \rangle$.

Definition 3.2. Let $\mathbf{A}$ be an algebra and T a tolerance of $\mathbf{A}$. A *strong* $\mathbf{1}, T$*-snag* of $\mathbf{A}$ is any subtrace (a, b) of $\mathbf{A}$ such that there exists a binary polynomial p of $\mathbf{A}$ and $(c, d) \in T$ such that, for $f(x) := p(x, c)$ and $g(x) := p(x, d)$, we have $f(a) = a$, $f(b) = b$ and $g(b) = a$.

Thus, a strong snag is a snag that is a subtrace and for which the twins f and g differ in a single parameter. To complete the picture, we shall call the pairs in $\mathbf{R}(T, T) - 0_{\mathbf{A}}$ *weak* $\mathbf{1}, T$*-snags.* Clearly, $\mathbf{A}$ has no weak $\mathbf{1}, T$-snags if and only if T is rectangular. We will need to be able to push snags forward into a factor algebra and to pull them back. The next lemma proves that weak snags can be pushed forward, strong snags can be pulled back, and ordinary snags go both ways.

Lemma 3.3. *Let* $\mathbf{A}$ *be a finite algebra, and* α *a congruence of* $\mathbf{A}$*. If* (a, b) *is a (weak)* $\mathbf{1}, T$*-snag of* $\mathbf{A}$*, and* $(a, b) \notin \alpha$*, then* $(a/\alpha, b/\alpha)$ *is a (weak)* $\mathbf{1}, T/\alpha$*-snag of* $\mathbf{A}/\alpha$*. Conversely, if* $(a/\alpha, b/\alpha)$ *is a (strong)* $\mathbf{1}, T/\alpha$*-snag of* $\mathbf{A}/\alpha$*, then there exist* $a' \ \alpha \ a$ *and* $b' \ \alpha \ b$ *such that* (a', b') *is a (strong)* $\mathbf{1}, T$*-snag of* $\mathbf{A}$*.*

Proof. The polynomials that witness that (a, b) is a (weak) $\mathbf{1}, T$-snag are also polynomials which, modulo α, witness that $(a/\alpha, b/\alpha)$ is a (weak) $\mathbf{1}, T$-snag. Therefore, we only need to show how to pull back a (strong) snag from a factor algebra.

Suppose that $(a/\alpha, b/\alpha)$ is a $\mathbf{1}, T/\alpha$-snag in $\mathbf{A}/\alpha$. Then there exist T-twin unary polynomials f and g of $\mathbf{A}$ such that $f(a)\ \alpha\ a$, $f(b)\ \alpha\ b$ and $g(b)\ \alpha\ a$. Let k be such that f^k is idempotent, let $b' = f^k(b)$ and $a' = f^{2k-1}g(b')$. It is straightforward to check that these elements satisfy the conditions with respect to the T-twin polynomials f^{2k} and $f^{2k-1}g$. This shows how to pull back an ordinary $\mathbf{1}, T$-snag.

If $(a/\alpha, b/\alpha)$ is a strong $\mathbf{1}, T$-snag, then the same argument works to pull back the snag structure, and pulling back this way preserves the fact that the T-twins differ in a single parameter. But we have to pull it back in such a way that we get a subtrace. From the fact that $(a/\alpha, b/\alpha)$ is a subtrace, we get that there exist congruences $\alpha \leq \delta \prec \theta$ and a $\langle \delta/\alpha, \theta/\alpha \rangle$-trace N/α such that $a, b \in N$ and $(a, b) \in \theta - \delta$. Let U/α be a $\langle \delta/\alpha, \theta/\alpha \rangle$-minimal set containing N/α, and e a unary polynomial of $\mathbf{A}$ such that e/α is idempotent, and its range is U/α. Then e does not collapse (a, b) to δ, so there exists a $\langle \delta, \theta \rangle$-minimal set U' of $\mathbf{A}$ such that $U' \subseteq e(A)$. Let $e'(A) = U'$ for an idempotent polynomial e' of $\mathbf{A}$. Clearly, U'/α is a $\langle \delta/\alpha, \theta/\alpha \rangle$-minimal set of $\mathbf{A}/\alpha$ contained in U/α, and hence $U/\alpha = U'/\alpha$. Therefore $a'' = e'(a)\ \alpha\ a$ and $b'' = e'(b)\ \alpha\ b$. Clearly, $(a'', b'') \in \theta - \delta$, so they are contained in a trace $N' \subseteq U'$. Replacing f and g with $e'f$ and $e'g$, respectively, the conditions in $\mathbf{A}/\alpha$ do not change, but the new f is now a permutation of N'. Therefore the pair (a', b') obtained using the new f and g with the method in the previous paragraph (starting from a'' and b'') is in $(N')^2 \cap (\theta - \delta)$, so it is indeed a subtrace. □

In the next lemma, $\mathbf{A}$ is a finite algebra, T is a tolerance of $\mathbf{A}$, $\langle \delta, \theta \rangle$ denotes a typical, unspecified prime quotient, U is an arbitrary $\langle \delta, \theta \rangle$-minimal set, and e_U is any idempotent unary polynomial for which $e_U(A) = U$.

Lemma 3.4. *The following are equivalent.*

(1) *T is rectangularly nilpotent.*
(2) *For every pair (e, f) of T-twin unary polynomials of $\mathbf{A}$, if $e^2 = e$ then $efe = e$.*
(3) *For every triple (e, f, g) of simultaneous T-twin unary polynomials of $\mathbf{A}$, if $e^2 = e$ and $g^2 = g$, then $efg = eg$.*
(4) *For every pair (h, k) of T-twin unary polynomials of $\mathbf{A}$ there exists a positive integer n such that $h^{2n-1}kh^n = h^n$.*
(5) *There is no $\mathbf{1}, T$-snag in $\mathbf{A}$.*
(6) *There is no strong $\mathbf{1}, T$-snag in $\mathbf{A}$.*
(7) *For any prime quotient $\langle \delta, \theta \rangle$, if $p(x, y)$ is a binary polynomial and $(c, d) \in T$, then $e_U p(x, c)$ is the identity map modulo δ on U if and only if $e_U p(x, d)$ is the identity map modulo δ on U.*

(8) *For any prime quotient* $\langle\delta,\theta\rangle$*, any two* T*-twin unary polynomials of* $\mathbf{A}$ *mapping* U *into* U *have the property that if one is the identity map on* U *modulo* δ*, then the other one is also the identity map on* U *modulo* δ*.*
(9) *For any prime quotient* $\langle\delta,\theta\rangle$*, any two* T*-twin polynomials of* $\mathbf{A}$ *mapping any product* $C = C_1 \times \cdots \times C_k$ *of* θ*-classes into* U *have the property that either they are equal modulo* δ *on* C*, or both collapse* C *into a* δ*-class.*
(10) *For any prime quotient* $\langle\delta,\theta\rangle$*, if* $\begin{bmatrix} u & v \\ w & z \end{bmatrix}$ *is a* T,θ*-matrix with entries in* U*, then it is trivial modulo* δ*, that is, either* $u \equiv_\delta v$ *and* $w \equiv_\delta z$ *or else* $u \equiv_\delta w$ *and* $v \equiv_\delta z$*.*
(11) T *strongly centralizes every prime quotient of* $\mathbf{A}$ *from the left and from the right.*
(12) T *rectangulates every prime quotient of* $\mathbf{A}$*.*
(13) *For every prime quotient* $\langle\delta,\theta\rangle$ *and each* $\langle\delta,\theta\rangle$*-trace* N*, the tolerance* T *weakly centralizes* N^2 *modulo* δ *and the* T*-twin group on* N/δ *is trivial.*

Proof. We will prove this statement by induction on the size of $\mathbf{A}$, so we can assume that all the statements are equivalent for all proper factors of $\mathbf{A}$.

(1)$\Longrightarrow$(2). If (1) holds, then there exists a chain $0_\mathbf{A} = \alpha_0 \le \alpha_1 \le \cdots \le \alpha_n = 1_\mathbf{A}$ such that $\mathbf{R}(T, \alpha_{i+1}; \alpha_i)$ holds for every $0 \le i < n$. If (2) fails, then $efe(a) \ne e(a)$ for some $a \in A$. Let $b = e(a)$, $c = ef(b)$, so $b \ne c$. Choose i so that $(b, c) \in \alpha_{i+1} - \alpha_i$. Since e and f are T-twins, so are ee and ef. Therefore

$$\begin{bmatrix} ee(c) & ee(b) \\ ef(c) & ef(b) \end{bmatrix} = \begin{bmatrix} c & b \\ * & c \end{bmatrix}$$

is a T, α_{i+1}-matrix showing that $\mathbf{R}(T, \alpha_{i+1}; \alpha_i)$ fails.

(2)$\Longleftrightarrow$(3). Obviously (2) is the special case of (3) where $e = g$. For the converse, applying (2) we get that $g = geg$, and by (2) again, $efge = eeee = e$, since fg is a T-twin of $ee = e$. Hence

$$efg = ef(geg) = (efge)g = eg\,.$$

(2)$\Longrightarrow$(4). Let $e = h^n$ be an idempotent power of h. Then (2) applied to $g = h^{n-1}k$ gives that

$$h^{2n-1}kh^n = h^n(h^{n-1}k)h^n = ege = e = h^n\,.$$

(4)$\Longrightarrow$(5). Suppose that (a, b) is a $\mathbf{1}, T$-snag, and so there exist T-twin polynomials h and k satisfying $h(a) = a$, $h(b) = b$, $k(b) = a$. Then we have $h^n(b) = b$, but $h^{2n-1}kh^n(b) = a$, contradicting (4).

(5)$\Longrightarrow$(6). A strong $\mathbf{1}, T$-snag is an ordinary $\mathbf{1}, T$-snag.

(6)$\Longrightarrow$(7). Suppose that $f(x) = e_U p(x, c)$ is the identity map of U modulo δ. We have to show that $g(x) = e_U p(x, d)$ is also the identity map on U modulo δ. Since f is the identity on U modulo δ, we have $f(\theta|_U) \not\subseteq \delta$. From Theorem 2.8 (3) of [4] it follows that f is a permutation of U. By composing both f and g with a polynomial inverse of f we may assume that f is the identity map on all of U. Now consider the algebra $\mathbf{A}/\theta$. By (6) and Lemma 3.3, this algebra does not contain strong $\mathbf{1}, T/\theta$ snags. Therefore, by the induction hypothesis, this algebra satisfies (2). That means that we have $fgf(x)\ \theta\ f(x)$ for every $x \in A$. In other words, $a = g(b)\ \theta\ b$ for every $b \in U$. Now if a and b are not δ-related, then from $f(a) = a$ and $f(b) = b$ we see that (a, b) is a strong $\mathbf{1}, T$-snag, which is impossible by (6). Therefore we have $g(b)\ \delta\ b$ for every $b \in U$.

(7) $\Longrightarrow$(8). The difference in statements (7) and (8) is that (7) is concerned only with T-twins built from binary polynomials and (8) deals with arbitrary twins. To prove the statement for arbitrary twins $f(x) = e_U t(x, \mathbf{c})$ and $g(x) = e_U t(x, \mathbf{d})$, consider first the binary polynomial $p(x, y) = e_U t(x, y, c_2, \ldots, c_k)$. From the fact that $f(x) = e_U p(x, c_1)$ is the identity map on U modulo δ, and $c_1\ T\ d_1$, we get that $e_U t(x, d_1, c_2, \ldots, c_k)$ is also the identity map of U modulo δ. Now using the binary polynomial $e_U t(x, d_1, y, c_3, \ldots, c_k)$ we see that $e_U t(x, d_1, d_2, c_3, \ldots, c_k)$ is again the identity map on U modulo δ. Continuing this process of switching c_i's to d_i's we finally get that g is the identity map on U modulo δ as well. Thus (8) is proved.

(8) $\Longrightarrow$(9). This implication is similar to (7) $\Longrightarrow$(9) of Lemma 2.4. First we prove the unary version. Let f and g be any two T-twin unary polynomials of $\mathbf{A}$ mapping A into U. We show that if f is a permutation on U, then f and g are equal on U modulo δ. Indeed, prefixing f and g by a polynomial inverse of f on U we can assume that f is the identity map on U. Now, from (8), we get that g is the identity on U modulo δ.

This observation shows that our assumption is stronger than that of Lemma 2.4 (7). Therefore, if f and g map the θ-class C into the same trace of U, then we are done by Lemma 2.4 (7) $\Longrightarrow$(8). We are also done if they map C into the same θ-class in the tail of U, since then both of them collapses C to δ. Therefore we may assume that $f(C)$ and $g(C)$ are contained in different θ-classes within U, and we have to show that both sets are contained in some δ-class.

Suppose that $f(a)$ and $f(b)$ are not δ-related for some $a, b \in C$. Connect a and b by traces. There is a trace N' in this chain that is not collapsed to δ by f. Let $U' \supseteq N'$ be a $\langle \delta, \theta \rangle$-minimal set. Then $f|_{U'}$ is a bijection between U' and U, let h be its polynomial inverse. Thus $f \circ h$ is the identity map on U, hence it must be equal to its T-twin $g \circ h$ modulo δ by (8). But this is impossible, since $f(N') \subseteq f(C)$ and $g(N') \subseteq g(C)$ are not even θ-related. This contradiction proves the unary case of (9).

For the general case we can also assume (this time using Lemma 2.4 (7) $\Longrightarrow$(9)) that the k-ary T-twin polynomials f and g map C to different θ-classes within U. Fixing any $k-1$ variables of f and g arbitrarily, the resulting unary functions cannot be equal modulo δ, and so by the unary version that we have already proved we see that they both must collapse the corresponding C_i to a δ-class. Therefore f and g both collapse C to δ.

(9) $\Longrightarrow$(10). This is the same proof as that of Lemma 2.4 (9) $\Longrightarrow$(10), with N replaced by U.

(10) $\Longrightarrow$(11). Let $\langle \delta, \theta \rangle$ be a prime quotient of $\mathbf{A}$. We show $\mathbf{C}(T, \theta; \delta)$ first. Suppose that there exists a T, θ-matrix where the elements in the top row are δ-related, but elements in the bottom row are not. We can map this matrix into any $\langle \delta, \theta \rangle$-minimal set U with a unary polynomial while keeping the bottom row in $\theta - \delta$. This yields a modulo δ nontrivial T, θ-matrix in U, contradicting (10). We have shown that (10) $\Longrightarrow \mathbf{C}(T, \theta; \delta)$. The same argument can be used to show that (10) implies both $\mathbf{C}(\theta, T; \delta)$ and $\mathbf{R}(T, \theta; \delta)$. Therefore $\mathbf{S}(T, \theta; \delta)$ and $\mathbf{S}(\theta, T; \delta)$ are both implied by (10).

(11) $\Longrightarrow$(12). Is tautologous.

(12) $\Longleftrightarrow$(13). This follows from the equivalence of (1) and (6) in Lemma 2.4.

(12) $\Longrightarrow$(1). If (12) holds and $0_{\mathbf{A}} = \alpha_0 \leq \alpha_1 \leq \cdots \leq \alpha_n = 1_{\mathbf{A}}$ is any maximal chain of congruences, then $\mathbf{R}(T, \alpha_{i+1}; \alpha_i)$ will hold for every $0 \leq i < n$. □

It is a consequence of the equivalence between conditions (11) and (13) of this lemma that if all T-twin groups on sets of the form N/δ are trivial, then all concepts of nilpotence coincide for T. In particular, from (12) $\Longleftrightarrow$(11) it follows that rectangularly nilpotent tolerances are the same as strongly nilpotent tolerances. We shall prefer this second name. An algebra is called *strongly nilpotent* if the tolerance $A \times A$ is strongly nilpotent. Here are some easy consequences of Lemma 3.4.

Lemma 3.5. *Let* $\mathbf{A}$ *be a finite algebra.*

(1) *If* $\mathbf{A}$ *is strongly nilpotent, then every finite algebra in the variety generated by* $\mathbf{A}$ *is strongly nilpotent.*
(2) *Every rectangular congruence of* $\mathbf{A}$ *is strongly nilpotent.*
(3) *If* T *is a strongly nilpotent tolerance of* $\mathbf{A}$*, then the ranges of any two idempotent* T*-twin polynomials are polynomially isomorphic.*
(4) *If* T *is a strongly nilpotent tolerance of* $\mathbf{A}$*, and* α *is a congruence of* $\mathbf{A}$*, then* T/α *is a strongly nilpotent tolerance in* $\mathbf{A}/\alpha$*.*

Proof. The statement in (1) follows from the fact that a finite algebra is strongly nilpotent if and only if it satisfies the property in Lemma 3.4 (4), which is an equationally expressible property. For (2), if α is a rectangular congruence, then $0_{\mathbf{A}} \leq \alpha \leq 1_{\mathbf{A}}$ is a chain which witnesses that α is rectangularly nilpotent. Therefore α is strongly nilpotent. To see that (3) holds, let e and f be idempotent T-twins, and apply Lemma 3.4 (2) to get $efe = e$ and $fef = f$. It follows that e and f are polynomial isomorphisms between their ranges. Finally, (4) is an easy consequence of Lemma 2.2 (5). □

After hearing about the results proved in the next section, Joel Berman asked whether one can check if a finite algebra $\mathbf{A}$ is strongly nilpotent by looking at subalgebras of small powers of $\mathbf{A}$. The next lemma answers this question affirmatively, and indicates a polynomial time algorithm to determine if a finite algebra in a finite language is strongly nilpotent.

Lemma 3.6. *A finite algebra* $\mathbf{A}$ *is strongly nilpotent if and only if for any two different elements* $a, b \in A$*, the subalgebra of* $\mathbf{A}^3$ *generated by* (a, b, b) *and the set* $\{(x, x, y) \mid x, y \in A\}$ *does not contain* (a, b, a)*.*

Proof. The elements of this subalgebra are triples $(f(a), f(b), g(b))$, where f and g are $(A \times A)$-twin unary polynomials of $\mathbf{A}$. Therefore the condition stated in this lemma is the condition that $\mathbf{A}$ has no $\mathbf{1}, (A \times A)$-snags. By Lemma 3.4 (5)$\Longleftrightarrow$(11) this is equivalent to strong nilpotence. □

In the final part of this section we shall define and investigate the strongly nilpotent radical and coradical of an algebra.

Definition 3.7. Let $\mathbf{A}$ be an algebra, e an idempotent unary polynomial of $\mathbf{A}$, and $E = e(A)$. Define the congruence ρ_E of $\mathbf{A}$ by $(c, d) \in \rho_E$ if and only if for every binary polynomial $p(x, y)$ of $\mathbf{A}$ we have that $ep(x, c)$ is the identity map of E if and only if $ep(x, d)$ is the identity map of E.

It is obvious that ρ_E is indeed a congruence, and it depends only on E, and not on e. It is also clear that if $\mathbf{A}$ is finite, $ep(x, c)$ is any permutation of E, and $(c, d) \in \rho_E$, then $ep(x, c) = ep(x, d)$ for every $x \in E$, because one can compose ep with a polynomial inverse of $ep(x, c)$ on E. If E is polynomially isomorphic to the range F of some other idempotent polynomial, then clearly $\rho_E = \rho_F$. Therefore, if $\langle \delta, \theta \rangle$ is a prime quotient of $\mathbf{A}$, then ρ_U is the same for every $\langle \delta, \theta \rangle$-minimal set U. Denote by $\rho^{\mathbf{A}}$ the intersection of the congruences ρ_U for every such $\langle \delta, \theta \rangle$. This congruence is called the *strongly nilpotent radical* of $\mathbf{A}$.

Lemma 3.8. *If* $\mathbf{A}$ *is a finite algebra, then* $\rho^{\mathbf{A}}$ *is a strongly nilpotent congruence which contains every strongly nilpotent tolerance of* $\mathbf{A}$*.*

Proof. We have already pointed out that $\rho^{\mathbf{A}}$ is a congruence. By the definition of $\rho^{\mathbf{A}}$, the statement of the lemma follows if we can show that the following modified form of Lemma 3.4 (7) is also equivalent to rectangular (that is, strong) nilpotence:

(7′) For any prime quotient $\langle \delta, \theta \rangle$, if $p(x, y)$ is a binary polynomial and $(c, d) \in T$, then $e_U p(x, c)$ is the identity map on U if and only if $e_U p(x, d)$ is the identity map on U.

(In this modified form the phrase "modulo δ" is missing.)

To show that (7′)$\Longrightarrow$(7) one has to prefix $f(x) = e_U p(x, c)$ and $g(x) = e_U p(x, d)$ by a polynomial inverse of f. To show that (7)$\Longrightarrow$(7′) we use that (7) implies (2) of Lemma 3.4, and as f is idempotent we get that $fgf = f$, which proves that g is indeed the identity map on U. □

Now we turn our attention to coradicals.

Definition 3.9. Let $\mathbf{A}$ be a finite algebra, and T a tolerance of $\mathbf{A}$. The congruence generated by all strong $\mathbf{1}, T$-snags is called the *strongly nilpotent coradical of* T, and is denoted by $\sigma^{\mathbf{A}}(T)$.

Lemma 3.10. *If* $\mathbf{A}$ *is a finite algebra, and* T *is a tolerance of* $\mathbf{A}$*, then the following hold.*

(1) *The congruence* $\sigma^{\mathbf{A}}(T)$ *is the smallest congruence* σ *of* $\mathbf{A}$ *such that* T/σ *is a strongly nilpotent tolerance of* $\mathbf{A}/\sigma$.
(2) *The tolerance* T *is strongly nilpotent if and only if* $\sigma^{\mathbf{A}}(T) = 0_{\mathbf{A}}$.
(3) *The congruence* $\sigma^{\mathbf{A}}(T)$ *equals the congruence generated by all (ordinary)* $\mathbf{1}, T$*-snags of* $\mathbf{A}$.

These statements are easy consequences of Lemma 3.3 and Lemma 3.4. Finally we show that the coradical-radical pair determines a polarity of the congruence lattice of any finite algebra. (Recall that a *polarity* on a lattice $\mathbf{L}$ is a pair $\langle \sigma, \rho \rangle$ where σ is a decreasing join endomorphism of $\mathbf{L}$, ρ is an increasing meet endomorphism of $\mathbf{L}$, and $\sigma\rho(x) \leq x \leq \rho\sigma(x)$ holds for all $x \in L$.) For a finite algebra $\mathbf{A}$ and a congruence α on $\mathbf{A}$, let $\rho^{\mathbf{A}}(\alpha)$ be the unique congruence above α for which $\rho^{\mathbf{A}}(\alpha)/\alpha = \rho^{\mathbf{A}/\alpha}$. Clearly, $\rho^{\mathbf{A}} = \rho^{\mathbf{A}}(0_{\mathbf{A}})$.

Lemma 3.11. *If* $\mathbf{A}$ *is a finite algebra, then* $\langle \sigma^{\mathbf{A}}, \rho^{\mathbf{A}} \rangle$ *is a polarity of the congruence lattice of* $\mathbf{A}$. *The congruence of* $\mathbf{Con}(\mathbf{A})$ *generated by the tolerance corresponding to this polarity is the strong solvability congruence.*

Proof. By Lemma 1.2 (1) of [4], it is sufficient to prove that $\sigma^{\mathbf{A}}$ is decreasing and for any two congruences α and β of $\mathbf{A}$ that $\sigma^{\mathbf{A}}(\alpha) \leq \beta$ if and only if $\alpha \leq \rho^{\mathbf{A}}(\beta)$. From the statements proved so far the reader can check that this is indeed the case. (Use the fact that both $\sigma^{\mathbf{A}}(\alpha) \leq \beta$ and $\alpha \leq \rho^{\mathbf{A}}(\beta)$ are equivalent to saying that β contains all $\mathbf{1}, \alpha$-snags.) The final statement follows from the fact that for each prime quotient $\langle \delta, \theta \rangle$ we have that θ/δ is strongly abelian if and only if it is strongly nilpotent. □

4. Finite Algebras of Finite Complexity

The purpose of the preceding sections was to build up machinery concerning rectangular and strong centrality. In this section we come to our first main application, which is the determination of which finite algebras have finite complexity.

Let $\mathbf{A}$ be a finite algebra in a *finite language.* In the Introduction we defined the complexity of a term as the depth of its composition tree, and the complexity of a term operation of $\mathbf{A}$ to be the minimum of the complexities of all terms which interpret as that operation. The complexity of $\mathbf{A}$ was defined to be the ordinal which is the supremum of the complexities of all of its term operations. We have seen that the complexity of $\mathbf{A}$ is finite if and only if there exists an integer K such that each term operation of $\mathbf{A}$ can be obtained as an interpretation of a term of at most K variables.

For an operation $f(\mathbf{x}, \mathbf{z})$ on a set A we say that f is *independent of* $\mathbf{x}$ if

$$f(\mathbf{x}, \mathbf{z}) = f(\mathbf{y}, \mathbf{z})$$

holds for every $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ in A. Otherwise $f(\mathbf{x}, \mathbf{z})$ *depends on* $\mathbf{x}$. We can define dependence/independence for each individual variable, and then we say that the *essential arity* of the operation f is the number of individual variables on which f depends. We say that an algebra $\mathbf{A}$ is of *finite essential arity* if there is an integer K which bounds the essential arity of all term operations of $\mathbf{A}$. By the remarks above, an algebra in a finite language has finite complexity if and only if it is of finite essential arity.

As we shall prove, an algebra in a finite language is strongly nilpotent if and only if it has finite essential arity. This statement does not hold for infinite languages, as the following example shows.

Example 4.1. Let $\mathbf{A}$ denote the algebra whose universe is $\mathbb{Z}_4$ (integers modulo 4), and whose basic operation symbols are all operations

$$\oplus_n(x_1, \dots, x_n) := 2x_1 + \cdots + 2x_n.$$

This algebra is strongly nilpotent, but it has no bound on the essential arity of its term (or basic) operations. However, any reduct of $\mathbf{A}$ to operations of arity less than some finite number K is an algebra whose essential arity is no more than K.

The previous example motivates the following definition.

Definition 4.2. Say that an algebra $\mathbf{A}$ is of *locally finite essential arity* if for each finite set F of basic operation symbols, the reduct $\mathbf{A}_F$ of $\mathbf{A}$ to the operations in F is of finite essential arity.

We shall prove that this property is equivalent to strong nilpotence for every finite algebra, even in infinite languages.

The above definition of finite complexity is also problematic if we allow infinite languages. A contrary demon could trivialize this concept in the following way. Given any algebra $\mathbf{A}$, the demon could construct a new algebra $\mathbf{A}'$ of small complexity which is term equivalent to $\mathbf{A}$ simply by adding a new basic operation symbol to the language of $\mathbf{A}'$ for each term operation of $\mathbf{A}$. Then $\mathbf{A}'$ would have the property that all of its term operations have complexity 1, and yet $\mathbf{A}'$ has the same term operations as the arbitrarily chosen algebra $\mathbf{A}$. However, the arguments in this section do say something nontrivial about the complexity of term operations for finite algebras in infinite languages. For this reason we choose to define a slightly different measure of complexity in this section. It is more complicated to state, but it better explains what we prove. For algebras in finite languages we will show that the two notions of 'finite complexity' coincide.

To define 'essential complexity' we need the concept of 'inessential reduction'. Assume that an algebra $\mathbf{A}$ has a term of the form $r(s(t(\mathbf{x}), \mathbf{x}), \mathbf{x})$ where r, s and t are terms. If

$$\mathbf{A} \models r(s(t(\mathbf{x}), \mathbf{x}), \mathbf{x}) = r(t(\mathbf{x}), \mathbf{x}),$$

then an *inessential reduction* of the term on the left hand side of the equality is the act of replacing it by the term on the right hand side of the equality. Thus an inessential reduction of a term $R = r \circ s \circ t$ is a modification of R by pruning off a subterm $S = s \circ t$ and then replacing S with a subterm t of S. This is a way of replacing a term by an equivalent one of smaller complexity which does not introduce new operation symbols.

If $\mathbf{A}$ is an algebra and t is a term of $\mathbf{A}$, then we define the *essential complexity of t* to be the minimum complexity among terms which can be obtained from t by a sequence of inessential reductions. We define the *essential complexity of* $\mathbf{A}$ to be the ordinal which is the supremum of the essential complexities of the terms of $\mathbf{A}$. Having 'finite essential complexity' is a more restrictive notion than having 'finite complexity'. It is a more complicated notion too, but it is less sensitive to the difference between finite and infinite languages.

Example 4.3. Let $\mathbf{A}$ be a finite algebra whose basic operation symbols are the unary symbols $\sigma_1, \sigma_2, \ldots$, etc. Then any term of $\mathbf{A}$ has the form

$$\sigma_{i_N} \cdots \sigma_{i_2} \sigma_{i_1}(x).$$

Fix such an expression. Now, if one considers the terms

$$\begin{aligned} f_1(x) &:= \sigma_{i_1}(x) \\ f_2(x) &:= \sigma_{i_2}\sigma_{i_1}(x) \\ &\vdots \\ f_N(x) &:= \sigma_{i_N} \cdots \sigma_{i_2}\sigma_{i_1}(x), \end{aligned}$$

then the f_i cannot all represent different term operations unless $N \leq |A|^{|A|}$, since there are only $|A|^{|A|}$ functions from A to A. Therefore, if $N > |A|^{|A|}$, then there is some $j < k$ such that $\mathbf{A} \models f_j(x) = f_k(x)$. Letting $t(x) = f_j(x) = \sigma_{i_j} \cdots \circ \sigma_{i_2}\sigma_{i_1}(x)$, $s(x) = \sigma_{i_k} \cdots \sigma_{i_{j+1}}(x)$ and $r(x) = \sigma_{i_N} \cdots \sigma_{i_{k+1}}(x)$, we have that $t(x) = f_j(x)$ and $s \circ t(x) = f_k(x)$. Therefore we have

$$\mathbf{A} \models r \circ s \circ t(x) = r \circ t(x).$$

This implies that it is always possible to perform an inessential reduction on a term whose composition tree has depth $> |A|^{|A|}$. The essential complexity of any term (hence the essential complexity of $\mathbf{A}$) is at most $|A|^{|A|}$.

Our goal in this section is to prove the following theorem.

Theorem 4.4. *Let $\mathbf{A}$ be a finite algebra. The following are equivalent.*

(1) **A** *has finite essential complexity.*
(2) **A** *is of locally finite essential arity.*
(3) **A** *is strongly nilpotent.*

The part of the proof of Theorem 4.4 that is difficult to prove is that (3) $\Longrightarrow$(1), so we postpone that part of the argument. We explain the easy parts of the proof now.

Proof. (1) $\Longrightarrow$(2). Note that if the algebra **A** from (1) has essential complexity less than some positive number K, then so does any reduct of **A** to finitely many of its basic operation symbols. Therefore it suffices to prove that when **A** has only finitely many basic operation symbols, if **A** has finite essential complexity, then **A** has finite essential arity. Fixing K as a finite essential complexity bound for **A**, let N be a finite bound on the arities of the basic operation symbols of **A**. Any composition tree of depth $\leq K$ built from $\leq N$-ary operations has at most N^K leaves. Any term operation of **A** is equivalent to a term with such a composition tree, so it cannot depend on more than N^K of its variables.

(2) $\Longrightarrow$(3). An algebra satisfies (2) or (3) if and only if every reduct to finitely many basic operation symbols satisfies the same condition. Therefore, we may assume that **A** is defined with finitely many basic operations symbols. We only need to show that if **A** is not strongly nilpotent, then it is not of finite essential arity. By Lemma 3.4 (2), if **A** is not strongly nilpotent, then there are twin unary polynomials $e^2 = e$ and f and an $a \in A$ such that $efe(a) \neq e(a)$. As e and f are twins, there exists a term t and tuples $\mathbf{u}$ and $\mathbf{v}$ of **A** such that $e(x) = t(x, \mathbf{u})$ and $f(x) = t(x, \mathbf{v})$. Let

$$p(x, \mathbf{y}^1, \ldots, \mathbf{y}^K) = t(t(\ldots(t(t(x, \mathbf{y}^1), \mathbf{y}^2), \ldots), \mathbf{y}^{K-1}), \mathbf{y}^K)\,.$$

We show that p depends on $\mathbf{y}^i$ for every $1 < i < K$ (so p depends on at least $K-2$ variables). Indeed, let $x = a$ and $\mathbf{y}^j = \mathbf{u}$ for every $j \neq i$. If $\mathbf{y}^i$ is set to $\mathbf{u}$, then p evaluates to $e^K(a) = e(a)$, since e is idempotent. On the other hand, if $\mathbf{y}^i$ is set to $\mathbf{v}$, then p evaluates to $e^{K-i}fe^{i-1}(a) = efe(a) \neq e(a)$. Thus p indeed depends on $\mathbf{y}^i$. This shows that (2) $\Longrightarrow$(3). □

We now prove the difficult implication in Theorem 4.4. From here until the end of the proof of Lemma 4.6 we will assume that **A** is a finite algebra which does not have finite essential complexity. We want to prove that **A** is not strongly nilpotent. We shall take a term of **A** having large essential complexity, modify its composition tree suitably, and then employ a Ramsey argument. In this modification process it would be convenient to label nodes with terms rather than basic operation symbols. Instead of extending the definition of a composition tree in this way, we simply enrich the language of **A** with new basic operation symbols for all terms. This cannot affect the strong nilpotence of **A**. It also cannot create new opportunities to make inessential reductions in the terms that previously existed, since inessential reductions in pre-existing terms cannot take advantage of the new symbols. Therefore these modifications cannot change $\mathbf{A}$ from an algebra of infinite essential complexity to an algebra of finite essential complexity.

Lemma 4.5. *For arbitrarily large integers $K \geq 1$ one can find K nonunary basic operation symbols $q_1, \ldots, q_K$ such that the composition*

$$p(x, \mathbf{y}^1, \ldots, \mathbf{y}^K) := q_K(\cdots q_2(q_1(x, \mathbf{y}^1), \mathbf{y}^2) \cdots, \mathbf{y}^K)$$

depends on every $\mathbf{y}^i$ in $\mathbf{A}$.

Proof. Let $f(\mathbf{z})$ by a term whose essential complexity is a very large number N — how large we need N to be will become clear during the course of the argument. The composition tree for f is deep and we assume, as we may, that there are no inessential reductions possible. Fixing a node n in the composition tree for f, which is labelled with a variable z_j or a basic operation symbol b, we can express f as $s(z_j, \mathbf{z})$ or $s(b(t_1(\mathbf{z}), \ldots, t_k(\mathbf{z})), \mathbf{z})$ for some terms $s, t_1, \ldots, t_k$. We will call the node n a *fertile* node if the term operation represented by $s(x, \mathbf{z})$ depends on the (new) variable x in $\mathbf{A}$. Otherwise it will be called *sterile.* Only fertile nodes can have children, for if n is an internal sterile node then the transformation

$$s(b(t_1(\mathbf{z}), \ldots, t_k(\mathbf{z})), \mathbf{z}) \mapsto s(z_i, \mathbf{z})$$

is an inessential reduction, provided that z_i is one of the variables in $\mathbf{z}$ which labels a descendant of n. Therefore, all internal nodes (nodes that are not labelled by variables) of the composition tree for f must be fertile. We shall modify the composition tree to that of an operation where all nodes are fertile.

We have seen that each sterile node is a leaf. By identification of variables in the operations which label the parent node of a sterile node, we can produce a new composition tree for an equivalent term, of essential complexity N, where there are no sterile nodes except possibly some leaves which are 'only children' of a fertile parent. Pruning away all such leaves, and relabelling their parents by new variables, produces a new composition tree for an equivalent term whose essential complexity is at least $N - 1$, and which has no sterile nodes. By making all variables distinct, introducing new variables if necessary, we get a term which depends on all of its variables.

By permuting variables in operations, we may assume that the leftmost descending path L is the longest descending path in the composition tree. This path has length at least $N - 1$. The nodes which are children of a node in L (including the bottom-most node in L) will be called *the children of L.* Now, successively remove all subtrees rooted at one of the children of L, and then replace these subtrees with a single node. We label the replaced nodes with new variables. We end up with a composition tree for a term which looks like the one in Figure 3. One can check that this pruning process does not introduce sterile nodes or produce any new instances where an inessential reduction might be employed.

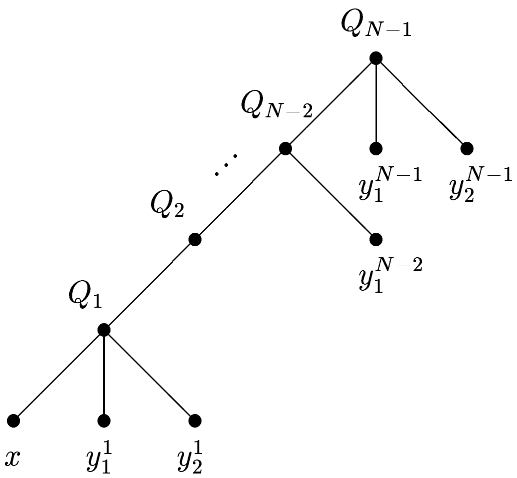

FIGURE 3: All nodes are fertile, no inessential reductions.

We have arranged the names of the new variables to suggest how the proof will be finished. The composition tree in Figure 3 is almost that of the kind claimed in the lemma, since the term represented has large essential complexity and it depends on all variables. However, many of the Q_i's could be unary basic operation symbols, as Q_2 is in Figure 3, and so the possibility remains that there are only few blocks of variables $\mathbf{y}^i = (y_1^i, \ldots, y_m^i)$ on which this term depends. To finish the proof, we will show that there must be a large number of blocks $\mathbf{y}^i$ on which the term depends.

Starting at x and traveling up L toward the root node, we cannot pass by any sequence of more than $|A|^{|A|}$ consecutive nodes which are labelled with unary term operations. The reason for this is explained by the argument in Example 4.3: any composition of more than $|A|^{|A|}$ unary terms admits an inessential reduction. This implies that at least $[(N-1)/|A|^{|A|}] - 1$ nodes are labelled with nonunary basic operation symbols of $\mathbf{A}$, so there are at least this many blocks of $\mathbf{y}^i$ on which the term depends.

We perform one last modification to our composition tree. If there are consecutive internal nodes (necessarily on L) labelled Q_i and Q_{i+1}, and at least one of them is labelled with a unary basic operation symbol, then we delete the node labelled by the unary symbol and relabel the other node by an operation symbol corresponding to the composite term $Q_{i+1}Q_i$. (Such an operation symbol exists, since we have enriched the language of $\mathbf{A}$ so that all term operations are represented by basic operation symbols). For example, in Figure 3 we might be considering Q_1 and Q_2; we delete the node labelled by Q_2 and relabel the node which had label Q_1 with a basic operation symbol representing the composite Q_2Q_1. There may be multiple ways to do this, but if one does this as many times as possible, then one obtains a term whose composition tree looks like that in Figure 3, except that there are no occurrences of unary basic operation symbols among the labels. If the labels that occur on internal nodes are now $q_1, \dots, q_K$, then the resulting term is of the form

$$p(x, \mathbf{y}^1, \dots, \mathbf{y}^K) := q_K(\cdots q_2(q_1(x, \mathbf{y}^1), \mathbf{y}^2) \cdots, \mathbf{y}^K).$$

By construction, this term depends on all of its variables. The depth K of the composition tree for this term is at least $[(N-1)/|A|^{|A|}] - 1$. Since we can choose N as large as we like, we can guarantee that K is arbitrarily large. □

The proof that (3)$\Longrightarrow$(1) in Theorem 4.4 is completed with the following lemma.

Lemma 4.6. *Let* $\mathbf{A}$ *be a finite algebra. Suppose that there are arbitrarily large integers* $K \geq 1$ *for which one can find* K *nonunary basic operation symbols* $q_1, \dots, q_K$ *such that the composition*

$$p(x, \mathbf{y}^1, \dots, \mathbf{y}^K) := q_K(\cdots q_2(q_1(x, \mathbf{y}^1), \mathbf{y}^2) \cdots, \mathbf{y}^K)$$

depends on every $\mathbf{y}^i$ *in* $\mathbf{A}$. *Then* $\mathbf{A}$ *is not strongly nilpotent.*

Proof. If $t(x, \mathbf{y})$ is a term, then let $U(t)$ denote the set of all unary polynomials $t(x, \mathbf{u}) : A \to A$, where $\mathbf{u}$ runs over all tuples of $\mathbf{A}$, that is, $U(t)$ is the set of all unary twins of t acting in its first variable. Call two terms $t(x, \mathbf{y})$ and $s(x, \mathbf{z})$ equivalent, if $U(t) = U(s)$. Clearly, this is an equivalence relation on the set of all terms of $\mathbf{A}$. The number of equivalence classes is at most $2^{|A|^{|A|}}$, since each $U(t)$ is a set of functions from A to A.

Now use the assumption of the lemma for a large integer K, which will be specified in the course of the proof, to get a term $p(x, \mathbf{y}^1, \dots, \mathbf{y}^K)$. For $0 \leq i < j \leq K$ denote by $p_{i,j}$ the "subcomposition"

$$p_{i,j}(x, \mathbf{y}^{i+1}, \dots, \mathbf{y}^j) = q_j(q_{j-1}(\dots(q_{i+2}(q_{i+1}(x, \mathbf{y}^{i+1}), \mathbf{y}^{i+2}), \dots), \mathbf{y}^{j-1}), \mathbf{y}^j).$$

Color each two-element set $\{i, j\} \subseteq \{1, 2, \dots, K\}$ by $U(p_{i,j})$. There is a bounded number of colors, so if K is sufficiently large, then, by Ramsey's Theorem, there exists an arbitrarily large homogeneous subset $I \subseteq \{1, 2, \dots, K\}$. Fix $m \in I$ "in the middle" of I, that is, in such a way that the set of elements of I smaller than m and the set of elements of I bigger than m are both as large as we need to make the rest of the argument work.

Since p depends on the tuple of variables $\mathbf{y}^m$, there exist tuples $\mathbf{u}^i$ for $1 \leq i \leq K$, and a tuple $\mathbf{v}$ such that for some $a \in A$ we have

$$p(a, \mathbf{u}^1, \dots, \mathbf{u}^m, \dots, \mathbf{u}^K) \neq p(a, \mathbf{u}^1, \dots, \mathbf{v}, \dots, \mathbf{u}^K)$$

(the arguments differ only at the m-th tuple). Consider the elements

$$a_i = p_{0,i}(a, \mathbf{u}^1, \dots, \mathbf{u}^i),$$

and let b_i be defined analogously, with the m-th tuple being $\mathbf{v}$ in place of $\mathbf{u}^m$. The above inequality says that $a_K \neq b_K$. For every $i \geq m$ we have that $a_{i+1} = q_i(a_i, \mathbf{u}^i)$

and $b_{i+1} = q_i(b_i, \mathbf{u}^i)$. Therefore $a_K \neq b_K$ implies that $a_i \neq b_i$ for every $i \geq m$. Obviously, $a_i = b_i$ for $i < m$.

Since the number of elements of I smaller than m can be made arbitrarily large, we can assume that it is bigger than $|A|$. So, there exist elements $i_1 < i_2 < m$ of I such that $a_{i_1} = a_{i_2}$. Similarly, as the number of elements of I bigger than m can be made arbitrarily large, we can assume that it is bigger than $|A|^2$. Then there exist elements $m < i_3 < i_4$ of I such that $a_{i_3} = a_{i_4}$ and $b_{i_3} = b_{i_4}$. Now set

$$\begin{aligned} g(x) &= p_{i_1,i_2}(x, \mathbf{u}^{i_1+1}, \dots, \mathbf{u}^{i_2}) \\ f(x) &= p_{i_2,i_3}(x, \mathbf{u}^{i_2+1}, \dots, \mathbf{u}^{m}, \dots, \mathbf{u}^{i_3}) \\ f'(x) &= p_{i_2,i_3}(x, \mathbf{u}^{i_2+1}, \dots, \mathbf{v}, \dots, \mathbf{u}^{i_3}) \\ e(x) &= p_{i_3,i_4}(x, \mathbf{u}^{i_3+1}, \dots, \mathbf{u}^{i_4})\,. \end{aligned}$$

So, letting $b = a_{i_1} = a_{i_2}$, $c = a_{i_3} = a_{i_4}$, $d = b_{i_3} = b_{i_4}$, we have that

$$(*) \qquad c \neq d,\ g(b) = b,\ f(b) = c,\ f'(b) = d,\ e(c) = c,\ e(d) = d\,.$$

We show that the four unary functions e, f, f', g are simultaneous $1_{\mathbf{A}}$-twins with respect to the term p_{i_2,i_3}. This is clear for f and f'. We know that $i_1, i_2, i_3, i_4 \in I$, and as I is homogeneous, we see that $\{i_1, i_2\}$ and $\{i_2, i_3\}$ have the same color. The function $g(x)$ is an element of $U(p_{i_1,i_2}) = U(p_{i_2,i_3})$, and therefore $g(x)$ can be obtained by substituting appropriate parameters into p_{i_2,i_3}. A similar argument shows that e is also a member of this family of twins.

If $h : A \to A$ is any function, then an easy and well known argument shows that for $k := (|A|)!$ we have $h^k(h^k(x)) = h^k(x)$. Replace g by g^k, e by e^k, f by $e^{k-1}f$, and f' by $e^{k-1}f'$. The new e, f, f', g are still simultaneous twins, and they still satisfy the set $(*)$ of equations above, but now we have in addition that e and g are idempotent.

If $\mathbf{A}$ is strongly nilpotent, then from Lemma 3.4 (3) we have that $efg = eg$ and $ef'g = eg$. This forces $efg = ef'g$, which is not the case, because applied to the element b, by $(*)$, we get $efg(b) = ef(b) = e(c) = c$, which is not the same as $ef'g(b) = d$. This completes the proof of the lemma, and also the proof of Theorem 4.4. □

The next corollary shows that the complexity measure defined in the Introduction agrees with that of this section for finite algebras in a finite language.

Corollary 4.7. *If* $\mathbf{A}$ *is a finite algebra in a finite language, then the following are equivalent.*

(1) $\mathbf{A}$ *has finite essential complexity.*
(2) $\mathbf{A}$ *has finite complexity.*
(3) $\mathbf{A}$ *has finite essential arity.*
(4) $\mathbf{A}$ *is strongly nilpotent.*

Proof. (1) $\iff$ (3) $\iff$ (4) follows from Theorem 4.4. Since the definition of 'finite essential complexity' is a restriction of that of 'finite complexity', we have (1) $\implies$ (2). The proof that (2) $\iff$ (3) is given in the opening paragraphs of this section. □

A slight rephrasing of part of this corollary, which is worth writing down, is the characterization of finite algebras of finite essential arity.

Corollary 4.8. *A finite algebra* **A** *has finite essential arity if and only if the following two conditions hold.*

(1) **A** *has a finite bound on the essential arity of its basic operations; and*
(2) **A** *is strongly nilpotent.*

It has long been known that a finite strongly abelian algebra has finite essential arity. Consequently, any algebra representable as a homomorphic image of a strongly abelian algebra has finite essential arity. We do not know whether, conversely, every algebra of finite essential arity has such a representation. We pose this as a problem.

Problem 4.9. Is it true that every finite algebra of finite essential arity is a homomorphic image of a finite strongly abelian algebra?

From the results of the next section, we will see that it may be possible to solve this problem by solving two possibly easier subproblems: one might first prove that every finite algebra of finite essential arity is a homomorphic image of a finite rectangular algebra, and then prove that every finite rectangular algebra is a homomorphic image of a finite strongly abelian algebra.

5. The Essential Arity of a Rectangular Algebra

Since the abelian concept associated with rectangular centrality is rectangularity, the algebras satisfying this condition deserve the closest scrutiny. Of course, any rectangular algebra is strongly nilpotent, and so from the previous section we know that any finite rectangular algebra has locally finite essential arity. However, we prove more about the essential arity of a rectangular algebra in this section. We prove that a finite rectangular algebra has finite essential arity *whether or not* its language is finite. Moreover, the bound we obtain on the essential arity is sharp: the essential arity of any k-element rectangular algebra is no more than $k-1$, and equality occurs for some k-element rectangular algebra. To prove this statement, we shall investigate a combinatorial problem about partitioning rectangles into rectangular subsets.

Definition 5.1. A *rectangle* is a set of the form $\mathbf{A} = A_1 \times \cdots \times A_K$ where $A_1, \ldots, A_K$ are nonempty sets. A *rectangular subset* of $\mathbf{A}$ is a nonempty subset of the form $\mathbf{B} = B_1 \times \cdots \times B_K$ with $B_i \subseteq A_i$ for each i. We say that a rectangular subset $\mathbf{B}$ has *full extent* in direction i if $B_i = A_i$.

Lemma 5.2. *Let $A_1, \dots, A_K$ be nonempty sets. If the rectangle $\mathbf{A} = A_1 \times \cdots \times A_K$ is partitioned into at most K rectangular subsets, then there exists an $1 \le i \le K$ such that each of the rectangular subsets has full extent in direction i.*

Proof. Let $\mathcal{B}$ be a partition of $\mathbf{A}$ into exactly ℓ rectangular subsets, where $\ell \le K$. For $\mathbf{B} \in \mathcal{B}$ we shall denote the components of $\mathbf{B}$ by $B_1, \dots, B_K$, which means that $\mathbf{B} = B_1 \times \cdots \times B_K$.

The lemma is obvious for $K = 1$. We proceed by induction on K. Suppose that $x \in A_j$ is such that there is a rectangular subset $\mathbf{B} \in \mathcal{B}$ with $x \notin B_j$. Then consider

$$\mathbf{A}' = A_1 \times \cdots \times A_{j-1} \times \{x\} \times A_{j+1} \times \cdots \times A_K \,.$$

This can be identified with the product of $K-1$ sets which one obtains by deleting the j-th coordinate. If we intersect $\mathbf{A}'$ with any member of $\mathcal{B}$, then we get a rectangular subset of $\mathbf{A}'$, or the empty set. In the case of $\mathbf{B}$ we get the empty set, because $x \notin B_j$. This produces a partition of $\mathbf{A}'$ into at most $K-1$ rectangular subsets. By the induction assumption, there exists an $i \ne j$ such that each rectangular subset in $\mathcal{B}$ which intersects $\mathbf{A}'$ has full extent in direction i. Draw an arrow of color x from j to every such i. Thus

$$j \xrightarrow{x} i \quad \Longleftrightarrow \quad \begin{cases} i \ne j,\, x \in A_j, \\ (\exists \mathbf{B} \in \mathcal{B})(x \notin B_j), \\ (\forall \mathbf{B}' \in \mathcal{B})(x \in B'_j \Longrightarrow B'_i = A_i) \,. \end{cases}$$

If direction j is not a direction of full extent for all rectangular subsets in $\mathcal{B}$, then there exists a $\mathbf{B} \in \mathcal{B}$, and an $x \in A_j$ such that $x \notin B_j$, which means that there is an arrow of color x starting out from j. Thus if the statement of the lemma is false, then there is a colored arrow starting out from every $1 \le j \le K$.

Suppose, to get a contradiction, that this is the case. Then there is a directed cycle formed by arrows. We may assume, by rearranging the coordinates, that this cycle is

$$1 \xrightarrow{x_1} 2 \xrightarrow{x_2} \dots \xrightarrow{x_{m-1}} m \xrightarrow{x_m} 1 \,.$$

The first arrow of the cycle shows that there is a rectangular subset $\mathbf{B} \in \mathcal{B}$ such that $x_1 \notin B_1$. Choose an arbitrary element $(y_1, \dots, y_K) \in \mathbf{B}$, and let $\mathbf{B}'$ be the rectangular subset containing $(x_1, y_2, \dots, y_K)$. Then by using the arrows above we have

$$\begin{aligned} x_1 \in B'_1 \implies B'_2 = A_2 \implies x_2 \in B'_2 \implies B'_3 = A'_3 \implies \dots \\ \dots \implies x_m \in B'_m \implies B'_1 = A_1 \,. \end{aligned}$$

But then $(y_1, \dots, y_K) \in \mathbf{B} \cap \mathbf{B}'$. This is a contradiction, since $\mathcal{B}$ is a partition and yet $\mathbf{B}'$ contains a point in $\mathbf{B}$ and a point not in $\mathbf{B}$. ☐

Theorem 5.3. *Let $\mathbf{A}$ be a rectangular algebra. Then each term operation of $\mathbf{A}$ depends on fewer than $|A|$ variables.*

Proof. Suppose instead that $t(x_1, \dots, x_K)$ is a term operation of $\mathbf{A}$ depending on all variables and that $K \geq |A|$. Let R denote the range of the function $t : A^K \to A$, and set $\ell = |R| \leq |A| \leq K$. For each $r \in R$ define $\mathbf{B}^r := t^{-1}(r)$. By Lemma 2.2 (9), each $\mathbf{B}^r$ is a rectangular subset of A^K. Hence, $\mathcal{B} = \{\mathbf{B}^r \mid r \in R\}$ is a partition of A^K into $\ell \leq K$ rectangular subsets. Lemma 5.2 proves that there is an i which is a direction of full extent for all rectangular subsets in $\mathcal{B}$, that is, $B^r_i = A$ for every $r \in R$. We shall get a contradiction by showing that t does not depend on x_i. Indeed, choose any $a_1, \dots, a_K \in A$ and let $r = t(a_1, \dots, a_K)$. Then $a_j \in B^r_j$ for all j and, since $B^r_i = A$, we get that

$$t(a_1, \dots, a_{i-1}, b, a_{i+1}, \dots, a_K) = r = t(a_1, \dots, a_{i-1}, c, a_{i+1}, \dots, a_K)$$

for every $b, c \in A$. This proves that t does not depend on x_i. □

Example 5.4. The bound in Theorem 5.3 is sharp, as we now show. Define an algebra $\mathbf{A}$ on the set $\{0, 1, \dots, n\}$ to have exactly one n-ary basic operation symbol f, whose interpretation in $\mathbf{A}$ is

$$f^{\mathbf{A}}(a_1, \dots, a_n) = \begin{cases} i & \text{if } a_i = n, \text{ but } a_j \neq n \text{ for } j > i; \\ 0 & \text{if } a_i \neq n \text{ for every } i. \end{cases}$$

Clearly, $f^{\mathbf{A}}$ depends on all n of its variables. Identities of the form

$$f(x_1, \dots, x_{i-1}, f(y_1, \dots, y_n), x_{i+1}, \dots, x_n) = f(x_1, \dots, x_{i-1}, y_n, x_{i+1}, \dots, x_n)$$

hold in $\mathbf{A}$, and can be used to prove that each term operation of this algebra can be obtained from f by permuting and identifying variables. Therefore, to prove that $\mathbf{A}$ is rectangular it suffices to show only that $f^{\mathbf{A}}$ itself has a "rectangular" operation table. That is, it suffices to observe that for each $i \leq n$ the set $(f^{\mathbf{A}})^{-1}(i)$ is a rectangular subset of A^n. From the definition of $f^{\mathbf{A}}$ we get that $(f^{\mathbf{A}})^{-1}(i)$ equals

$$\{0, \dots, n\} \times \cdots \times \{0, \dots, n\} \times \{n\} \times \{0, \dots, n-1\} \times \cdots \times \{0, \dots, n-1\}.$$

Therefore, the algebra $\mathbf{A}$ is rectangular, and it has essential arity $|A| - 1$.

There is a combinatorial problem suggested by Lemma 5.2 which appears to be completely unrelated to any of the algebraic questions we are considering, yet which seems worth including here.

Problem 5.5. Let $A_1, \dots, A_K$ be nonempty sets. If the rectangle $\mathbf{A} = A_1 \times \cdots \times A_K$ is partitioned into less than 2^K rectangular subsets, does it follow that there exists a rectangular subset in this partition which has full extent in some direction i?

6. A Representation Theorem for Rectangular Algebras

Assume that $\mathbf{B}$ is an algebra and that $\langle B; \vee \rangle$ is a semilattice on the same universe. We say that $\vee$ is a *compatible semilattice operation of* $\mathbf{B}$ if $\vee : \mathbf{B}^2 \to \mathbf{B}$ is a homomorphism. Our purpose in this section is to characterize rectangular algebras as those algebras $\mathbf{A}$ for which there is an algebra $\mathbf{B}$ in the same language which has a compatible semilattice operation $\vee$ such that $\mathbf{A}$ is representable as a subalgebra of $\mathbf{B}$ which is an antichain in the $\vee$ order. The proof we give can easily be localized to rectangular tolerances, and we explain how to do this after proving the theorem for rectangular algebras.

Theorem 6.1. *An algebra* $\mathbf{A}$ *is rectangular if and only if there is an algebra* $\mathbf{B}$ *in the same language and a compatible join semilattice operation* $\vee$ *of* $\mathbf{B}$ *such that* $\mathbf{A}$ *is isomorphic to the subalgebra of minimal elements of* $\mathbf{B}$ *under the* $\vee$ *order.*

Proof. We first prove the easy direction, which is that if $\mathbf{A}$ is the subalgebra of minimal elements of $\mathbf{B}$ then $\mathbf{A}$ is rectangular. Assume that

$$\begin{bmatrix} t(\mathbf{a},\mathbf{c}) & t(\mathbf{a},\mathbf{d}) \\ t(\mathbf{b},\mathbf{c}) & t(\mathbf{b},\mathbf{d}) \end{bmatrix} = \begin{bmatrix} u & v \\ w & z \end{bmatrix}$$

is a $1_{\mathbf{A}}, 1_{\mathbf{A}}$-matrix of elements from $\mathbf{A}$ where $u = z$. Then, as $\vee$ is a compatible operation of $\mathbf{B}$, we calculate in $\mathbf{B}$ to find that

$$u = z = u \vee z = t(\mathbf{a},\mathbf{c}) \vee t(\mathbf{b},\mathbf{d}) = t(\mathbf{a} \vee \mathbf{b}, \mathbf{c} \vee \mathbf{d}) \geq t(\mathbf{a},\mathbf{d}) = v,\ t(\mathbf{b},\mathbf{c}) = w.$$

But if $u = z \geq v$, w, and all are minimal elements, then $u = z = v = w$. Thus $\mathbf{A}$ is rectangular.

For the other direction, let $\mathbf{C}$ be the algebra of nonempty subsets of $\mathbf{A}$ under the complex operations of $\mathbf{A}$. By *complex operations* we mean that if $f(x_1, \dots, x_n)$ is a basic operation of $\mathbf{A}$ and $S_1, \dots, S_n$ are nonempty subsets of $\mathbf{A}$, then in $\mathbf{C}$ we have

$$f(S_1, \dots, S_n) := \{f(s_1, \dots, s_n) \mid s_i \in S_i\}.$$

Observe first that $\mathbf{C}$ is an algebra in the same language as $\mathbf{A}$ which can be equipped with a semilattice operation — the union operation $\cup$ on subsets of $\mathbf{A}$. Observe next that $\varphi : \mathbf{A} \to \mathbf{C}$ defined by $a \mapsto \{a\}$ is an isomorphism of $\mathbf{A}$ onto the subalgebra of singleton subsets, which is the subalgebra of minimal elements of $\mathbf{C}$. The algebra $\mathbf{C}$ may or may not be the one we seek; the only remaining difficulty is that we don't know if $\cup$ is a compatible semilattice operation of $\mathbf{C}$.

The idea to complete the proof will be to define a congruence θ on $\mathbf{C}$ which is compatible with $\cup$ so that the composite homomorphism

$$\mathbf{A} \overset{\varphi}{\hookrightarrow} \mathbf{C} \to \mathbf{C}/\theta =: \mathbf{B}$$

is injective. Moreover, we want to define θ so that $\cup$ is a compatible semilattice operation of $\mathbf{B}$.

Let $\mathbf{C}^* = \langle \mathbf{C}; \cup \rangle$. Since all we lack at present is to have $\cup$ commute with the operations of $\mathbf{C}$, we define θ to be the congruence on $\mathbf{C}^*$ generated by all failures of commutativity between $\cup$ and operations of $\mathbf{C}$. Specifically, we let θ be generated by all pairs of the form

$$\langle\, f(S_1, \dots, S_n) \cup f(T_1, \dots, T_n),\ f(S_1 \cup T_1, \dots, S_n \cup T_n) \,\rangle$$

where f is a basic operation of $\mathbf{A}$. Since θ is defined to be a congruence of $\mathbf{C}^*$, it follows that θ is a congruence of $\mathbf{C}$ which is compatible with $\cup$. Therefore, $\mathbf{B} := \mathbf{C}/\theta$ has a semilattice operation, and from the definition of θ it must be a compatible semilattice operation. We need only to check now that the induced homomorphism from $\mathbf{A}$ into $\mathbf{B}$ is an embedding. This is established by the following claim.

Claim 6.2. *If $\langle U, V\rangle \in \theta$ and $U = \{a\}$, then $V = \{a\}$.*

Proof. To see that this is so, it suffices to check the claim only in the case where $\langle U, V\rangle$ is a polynomial image of a generating pair for θ. Here, by a polynomial, we mean a unary polynomial of $\mathbf{C}^*$, however it suffices to consider only unary polynomials p of the forms

(g) $p(x) = g(x, W_1, \dots, W_k)$ where g is the complex operation associated with some term of $\mathbf{A}$ and each $W_i \subseteq A$, and also

($\cup$) $p(x) = x \cup Z$ where $Z \subseteq A$.

Since the generating pairs for θ are comparable, the operations of $\mathbf{C}^*$ are monotone and $U = \{a\}$ is minimal, we can assume that

$$\begin{aligned}\{a\} \;=\; U \;&=\; p(f(S_1, \dots, S_n) \cup f(T_1, \dots, T_n)) \\ &\leq\; p(f(S_1 \cup T_1, \dots, S_n \cup T_n)) \;=\; V.\end{aligned}$$

Assume that there is some $b \in V - U$. Then, in Case (g), after reordering the variables of f, we have that there are elements $s_i \in S_i$ $(1 \leq i \leq j)$, $t_i \in T_i$ $(j+1 \leq i \leq n)$, $w_i \in W_i$ $(1 \leq i \leq k)$ such that

$$b = g(f(s_1, \dots, s_j, t_{j+1}, \dots, t_n), w_1, \dots, w_k) = g(f(\mathbf{s}, \mathbf{t}), \mathbf{w})\,.$$

However, the fact that

$$\{a\} = U = p(f(S_1, \dots, S_n) \cup f(T_1, \dots, T_n))$$

implies that there are tuples $\mathbf{s}'$ and $\mathbf{t}'$ such that with the previous $\mathbf{s}, \mathbf{t}$ and $\mathbf{w}$ we have

$$\begin{bmatrix} g(f(\mathbf{s}, \mathbf{s}'), \mathbf{w}) & g(f(\mathbf{s}, \mathbf{t}), \mathbf{w}) \\ g(f(\mathbf{t}', \mathbf{s}'), \mathbf{w}) & g(f(\mathbf{t}', \mathbf{t}), \mathbf{w}) \end{bmatrix} = \begin{bmatrix} a & b \\ * & a \end{bmatrix}.$$

This is a failure of rectangularity. Therefore $V = U = \{a\}$ in Case (g). The argument for Case ($\cup$) is essentially the same, but a trifle easier, so we omit it. □

Now that the claim has been proven, we see that θ restricts trivially to the minimal elements of $\mathbf{C}$, and therefore the induced homomorphism of $\mathbf{A}$ into $\mathbf{B} = \mathbf{C}/\theta$ is an embedding. This completes the argument. □

One can localize the above proof to describe the structure on a rectangular tolerance, although it is not as easy to state the final result. What one can show (with the same argument as above) is that T is a rectangular tolerance of $\mathbf{A}$ if and only if there is an algebra $\mathbf{B}$ in the same language as $\mathbf{A}$ such that

(1) $\mathbf{A}$ is a subalgebra of $\mathbf{B}$;
(2) $\mathbf{B}$ has a rectangular tolerance S for which $S|_A = T$;
(3) There is a homomorphism $\vee : \mathbf{S} \to \mathbf{B}$ from the subalgebra $\mathbf{S} \leq \mathbf{B}^2$ to $\mathbf{B}$ which satisfies $x \vee x = x$, $x \vee y = y \vee x$ and $x \vee (y \vee z) = (x \vee y) \vee z$; and
(4) any block of T is an antichain with respect to the $\vee$-ordering.

For the proof of this local version one begins by defining $\mathbf{C}$ to be the algebra of those nonempty subsets $U \subseteq A$ for which $(U \times U) \subseteq T$ under the complex operations. Let S' be the tolerance on $\mathbf{C}$ defined by $U\ S'\ V$ if and only if $(U \times V) \subseteq T$. If U and V are S'-related, then $(U \cup V) \in C$ and $(U \cup V)$ is S'-related to U and V. Thus any S'-block is closed under $\cup$. One next defines θ as we did in the proof of the theorem: it is the least congruence on $\mathbf{C}$ compatible with the partial operation $\cup$ modulo which $\cup$ commutes with the operations of $\mathbf{C}$ on any product of S'-blocks. We define $\mathbf{B} = \mathbf{C}/\theta$, $S = S'/\theta$ and $\vee = \cup/\theta$. Verification of items (1) – (4) can be accomplished using the same arguments as in the theorem.

7. The Clone of a Rectangular Variety

The four centralizer concepts described in the Introduction each have their own model of what an "abelian" algebra is (or more specifically, what a "self-centralizing" algebra is). In all instances, whether we are considering the abelian, strongly abelian, weakly abelian or rectangular property, the property is equivalent to the satisfaction of a family of universal Horn formulas, but not equivalent to a set of equations. This implies that there are algebras which centralize themselves in one of the senses, but which generate varieties containing algebras which are not self-centralizing in the same sense. For each type of centrality this is a situation which demands investigation.

We believe that an adequate solution to the question of which finite algebras generate strongly abelian varieties or (normally) abelian varieties is provided by the papers [7] and [8] respectively. These papers include Klukovits-type characterizations of the clones of locally finite strongly abelian and abelian varieties. In this section we prove an analogous characterization theorem for locally finite rectangular varieties. We expect that a similar result holds for locally finite weakly abelian varieties, but we do not know of one.

To explain what we intend to do here, it is best to begin by describing the results in the literature which directly precede the results of this section. We begin with the following definition.

Definition 7.1. Let $t(\mathbf{y}, \mathbf{z})$ be an $(m+n)$-ary term of a variety $\mathcal{V}$. We say that $\mathbf{y}$, considered as an unordered set, is a *klukovits subset* of variables for t if there is a $(2m+1)$-ary term $k(\mathbf{x}, \mathbf{y}, u)$ such that

$$\mathcal{V} \models k(\mathbf{x}, \mathbf{y}, t(\mathbf{y}, \mathbf{z})) = t(\mathbf{x}, \mathbf{z})\,.$$

If this happens then k is called a *klukovits term for* $t(\mathbf{y},\mathbf{z})$ *in the variables* $\mathbf{y}$. We call $\mathbf{y}$ a *strong klukovits subset* if there is an $(m+1)$-ary term $k(\mathbf{x},u)$ such that

$$\mathcal{V} \models k(\mathbf{x},t(\mathbf{y},\mathbf{z})) = t(\mathbf{x},\mathbf{z})\,,$$

and if this happens then k is called a *strong klukovits term for* t *in* $\mathbf{x}$.

The importance of klukovits terms is that they give one a way to change the values substituted in a term operation $t(\mathbf{a},\mathbf{b})$ — values which appear "on the inside" — by operating on the result "from the outside". In particular, if a term $t(\mathbf{x},\mathbf{y})$ has a klukovits term in $\mathbf{x}$, then it must satisfy the "term condition" (described in Chapter 3 of [4]) with respect to the variables $\mathbf{x}$. The *term condition* asserts that for all terms $t(\mathbf{x},\mathbf{y})$ the following implication holds:

$$t(\mathbf{x},\mathbf{y}) = t(\mathbf{x},\mathbf{y}') \Longrightarrow t(\mathbf{x}',\mathbf{y}) = t(\mathbf{x}',\mathbf{y}')\,.$$

To see how a klukovits term in $\mathbf{x}$ helps to establish the term condition in the variables $\mathbf{x}$, assume that

$$t(\mathbf{a},\mathbf{c}) = t(\mathbf{a},\mathbf{d})$$

and that k is a klukovits term for $t(\mathbf{x},\mathbf{y})$ in $\mathbf{x}$. Then

$$t(\mathbf{b},\mathbf{c}) = k(\mathbf{b},\mathbf{a},t(\mathbf{a},\mathbf{c})) = k(\mathbf{b},\mathbf{a},t(\mathbf{a},\mathbf{d})) = t(\mathbf{b},\mathbf{d})\,.$$

Thus, if every subset of variables of every term of $\mathcal{V}$ is a klukovits subset, then every algebra in $\mathcal{V}$ satisfies the term condition, and so is abelian. It is rather straightforward to show, in the same way, that if every subset of variables of every term of $\mathcal{V}$ is a strong klukovits subset, then every algebra in $\mathcal{V}$ is strongly abelian.

The usefulness of klukovits terms was first noticed in Klukovits' paper [9], which contains the observation that a variety $\mathcal{V}$ has the hamiltonian property (that for every $\mathbf{A} \in \mathcal{V}$ all subalgebras of $\mathbf{A}$ are congruence blocks) if and only if every subset of variables of every term of $\mathcal{V}$ is what we are calling a klukovits subset. In particular, hamiltonian varieties are abelian. For locally finite varieties the converse statement is true, and is proved in [8]. The most important consequence of this is the following characterization of locally finite abelian varieties:

> *A locally finite variety $\mathcal{V}$ is abelian if and only if every subset of variables of every term of $\mathcal{V}$ is a klukovits subset.*

The corresponding result for locally finite strongly abelian varieties appears in [7]:

> *A locally finite variety $\mathcal{V}$ is strongly abelian if and only if every subset of variables of every term of $\mathcal{V}$ is a strong klukovits subset.*

Our aim in this section is to prove a similar statement, which we roughly state now as:

> *A locally finite variety $\mathcal{V}$ is rectangular if and only if there are* **enough** *terms of $\mathcal{V}$ which have* **enough** *strong klukovits subsets of variables.*

To make sense of this we need to explain the emphasized words. If $\mathcal{V}$ is a locally finite rectangular variety, we will define a set of terms of $\mathcal{V}$ called the "maximal terms" which will have the property that every term is equivalent to one derived from a maximal term by identification of variables. Where we say above that "... *there are* **enough** *terms of* $\mathcal{V}$ *which have* ..." we mean that "... *the* **maximal** *terms of* $\mathcal{V}$ *have* ...". We will learn that in a rectangular variety, if $m(\mathbf{x})$ is a maximal term, then it is possible to reorder the variables of m so that $\mathbf{x} = (x_1, \dots, x_n)$ and each of the subsets

$$\emptyset \subseteq \{x_1\} \subseteq \{x_1, x_2\} \subseteq \cdots \subseteq \{x_1, \dots, x_{n-1}\} \subseteq \{x_1, x_2, \dots, x_n\}$$

is a strong klukovits subset of m. This is what we mean by having "... **enough** *strong klukovits subsets of variables*".

In fact, the subset system determined by the collection of strong klukovits subsets of variables of a maximal term is a fascinating combinatorial object, and we shall completely describe the structure of such subset systems in this section.

An intriguing question which we leave the reader to ponder is whether or not the pattern of ideas described in the previous paragraphs of this section can be made complete by proving that locally finite weakly abelian varieties are precisely those locally finite varieties where enough terms have enough (ordinary) klukovits subsets.

Problem 7.2. Give a Klukovits-type characterization of locally finite weakly abelian varieties.

We begin by introducing an easy combinatorial concept which we shall refer to frequently in this section.

Definition 7.3. Let X be a finite set and let K be a collection of subsets of X. We say that K is a *system* of subsets of X if K has the following properties.

(i) $\emptyset, X \in \mathsf{K}$; and

(ii) if $U, V \in \mathsf{K}$, then $U \cup V \in \mathsf{K}$.

We say that a system K is a *separating* system of subsets if it has the following separation property:

(iii) if $V \in \mathsf{K}$ and $i, j \in V$ are distinct, then there is a $W \in \mathsf{K}$ such that $W \subseteq V$ and W *separates* i and j, meaning that W contains exactly one of i and j.

Strictly speaking, the following concepts from lattice theory are not needed to understand the material of this section. However we think they help to understand separating subset systems.

Definition 7.4. A lattice $\mathbf{L}$ is *locally distributive* if whenever $x \in L$ and y is the join of the covers of x, then the interval $[x, y]$ is distributive. A lattice is *meet semidistributive* if it satisfies the implication

$$x \wedge y = x \wedge z \Longrightarrow x \wedge y = x \wedge (y \vee z)\,.$$

A lattice is *semimodular* if it satisfies the implication

$$x \wedge y \prec x \Longrightarrow y \prec x \vee y\,.$$

Lemma 7.5. *Let X be a finite set and let K be a collection subsets of X. Let $\mathbf{K}$ denote the poset which is K under the inclusion order. The following are equivalent.*

(1) *K is a separating subset system on X.*
(2) *The poset $\mathbf{K}$ is a locally distributive lattice of height $|X|$.*
(3) *The poset $\mathbf{K}$ is a meet semidistributive, semimodular lattice of height $|X|$.*
(3′) *The poset $\mathbf{K}$ is a semimodular lattice of height $|X|$.*

If these conditions hold, then there is a sequence of subsets from K of the form

$$\emptyset \subseteq \{x_1\} \subseteq \{x_1, x_2\} \subseteq \cdots \subseteq \{x_1, \ldots, x_{n-1}\} \subseteq \{x_1, x_2, \ldots, x_n\} = X.$$

Proof. The proof of this lemma is not difficult, nor is the result of the lemma central to the material in this section. Therefore we choose to sketch the proof only, by listing a sequence of easily verified claims which taken together justify the lemma.

(I) For (1)$\Longrightarrow$(2).
- If (1) holds, then $\mathbf{K}$ is a lattice.
- If (1) holds and $U, V \in \mathsf{K}$ with $U \prec V$, then $|V - U| = 1$. (Otherwise, if i and j are distinct elements in $V - U$, then choose a $W \in \mathsf{K}$ such that $W \subseteq V$ and W separates i and j. Then $U \cup W \in \mathsf{K}$ and $U \subset U \cup W \subset V$, contradicting $U \prec V$.) In particular, the height of $\mathbf{K}$ is $|X|$.
- If (1) holds, $U \in \mathsf{K}$, and V is the join (= union) of the upper covers of U, then the interval $[U, V]$ in $\mathbf{K}$ is the Boolean lattice of all W such that $U \subseteq W \subseteq V$. Thus (2) holds.

(II) For (2)$\Longleftrightarrow$(3).
- A finite lattice is locally distributive if and only if it is meet semidistributive and semimodular. (Help for this claim can be found in Chapter 7 of [2].)

(III) For (3′)$\Longrightarrow$(1).
- If $\mathbf{K}$ is a finite semimodular lattice, and $U, V \in \mathsf{K}$ with $U \subseteq V$, then all maximal chains in $\mathbf{K}$ from U to V have the same length.
- If (3′) holds, then for $U, V \in \mathsf{K}$ with $U \subseteq V$ all maximal chains from U to V are of length $|V - U|$. In particular, the natural rank function of the lattice, given by the height, evaluates as $\mathrm{rank}(U) = |U|$.
- Consequently, if (3′) holds, then $\emptyset, X \in \mathsf{K}$ and K is closed under union.
- If (3′) holds, then K is separating. (To see this, assume that $V \in \mathsf{K}$ contains distinct i and j. To find a separating W, let $V' \in \mathsf{K}$ be minimal among subsets of V in K which contain i. If V' doesn't contain j let $W = V'$. Otherwise, if V' contains both i and j, let W be any subset which is a lower cover of V' in $\mathbf{K}$. Then $|V'| = |W| + 1$ and W does not contain i; therefore W contains j.)

The last statement of the lemma follows from the claims in (III). □

The previous lemma shows that if K is a separating system of subsets of a finite set X, then under inclusion the sets in K form a locally distributive lattice of height $|X|$. Conversely, given any locally distributive lattice $\mathbf{L}$ of finite height k, then $\mathbf{L}$ can be realized as the lattice of sets in a separating system of subsets of a k-element set. The idea for how to do this is to let X be the set of meet irreducibles of $\mathbf{L}$ (different from $1_{\mathbf{L}}$) and define K to be the system of subsets of X which have the following form: $U \in \mathsf{K}$ if and only if there is an $x \in L$ such that U is the set of meet irreducibles not above x. Using local distributivity it is easy to prove that K is a separating system of subsets of $\mathbf{L}$, and that if $\mathbf{K}$ is the lattice of sets in K under inclusion then $\mathbf{K} \cong \mathbf{L}$ in a natural way.

Turning from combinatorics back to algebra, let $\mathbf{A}$ be an algebra. If $t(x_1, \ldots, x_n)$ is a term in the language of $\mathbf{A}$, let $X(t) = \{x_1, \ldots, x_n\}$ and let $\mathsf{K}(t)$ be the subsets of $X(t)$ which are strong klukovits subsets of t. If, after possibly reordering the variables, $\mathsf{K}(t)$ contains a sequence of subsets of the form

$$\emptyset \subseteq \{x_1\} \subseteq \{x_1, x_2\} \subseteq \cdots \subseteq \{x_1, \ldots, x_{n-1}\} \subseteq \{x_1, x_2, \ldots, x_n\} = X(t),$$

then we will say that t *has enough strong klukovits subsets.* Even when $\mathbf{A}$ is a finite algebra in a rectangular variety it can happen that not all terms of $\mathbf{A}$ have enough strong klukovits subsets, as we show in Example 7.9. However, a certain class of terms do. We proceed with the definition of these terms.

Definition 7.6. A term $m(\mathbf{x})$ of $\mathbf{A}$ is a *maximal term* if

(i) $m^{\mathbf{A}}(\mathbf{x})$ depends on all of its variables; and
(ii) there is no way to obtain m from a term t by identification of variables if $t^{\mathbf{A}}$ has larger essential arity than $m^{\mathbf{A}}$.

If $\mathbf{A}$ does not have finite essential arity, then it may have no maximal terms. But if $\mathbf{A}$ has finite essential arity, then clearly every term operation of $\mathbf{A}$ which depends on all of its variables is the interpretation of some term obtained from a maximal term by identification of variables.

The principal result of this section is the following theorem. (A slight generalization of this result appears as Corollary 7.11.)

Theorem 7.7. *Let $\mathbf{A}$ be a finite algebra of finite essential arity. The following are equivalent.*

(1) *Every maximal term for $\mathbf{A}$ has enough strong klukovits subsets.*
(2) *If $m(\mathbf{x}, \mathbf{u}, \mathbf{z})$ is a maximal term for $\mathbf{A}$, then*

$$\mathcal{V}(\mathbf{A}) \models (m(\mathbf{x}, \mathbf{u}, \mathbf{z}) = m(\mathbf{y}, \mathbf{v}, \mathbf{z}) \Longrightarrow m(\mathbf{x}, \mathbf{v}, \mathbf{z}) = m(\mathbf{x}, \mathbf{u}, \mathbf{z}) = m(\mathbf{y}, \mathbf{u}, \mathbf{z}))\ .$$

(3) *$\mathcal{V}(\mathbf{A})$ is rectangular.*

(4) *For any maximal term m, the collection $\mathsf{K}(m)$ of all strong klukovits subsets of $X(m)$ is a separating subset system.*

Proof. (1)$\Longrightarrow$(2). Assume instead that (1) holds and that (2) fails. Then, because of the symmetry of the implication in (2) with respect to the first two sequences of variables of m, there is a maximal term $m(\mathbf{x},\mathbf{y},\mathbf{u})$ for $\mathbf{A}$ and an algebra $\mathbf{B}\in\mathcal{V}(\mathbf{A})$ which has tuples $\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d},\mathbf{e}$ such that

$$m(\mathbf{a},\mathbf{c},\mathbf{e}) \;=\; m(\mathbf{b},\mathbf{d},\mathbf{e}) \;\neq\; m(\mathbf{b},\mathbf{c},\mathbf{e})\,.$$

To any such failure we associate the positive integer ℓ which is the sum of the lengths of the sequences $\mathbf{x}$ and $\mathbf{u}$ in $m(\mathbf{x},\mathbf{u},\mathbf{z})$. We assume that the partition $(\mathbf{x}\mid\mathbf{u}\mid\mathbf{z})$ has been chosen so that we have a failure of (2) in which ℓ is minimal for this term m. (Of course, to have a failure, both $\mathbf{x}$ and $\mathbf{u}$ must be nonempty, so we must have $\ell\geq 2$.) To obtain a contradiction, we shall produce another failure where ℓ is smaller.

A consequence of (1) is that $m(\mathbf{x},\mathbf{u},\mathbf{z})$ has a strong klukovits subset which contains exactly one variable from the sequence $\mathbf{xu}$. We choose such a subset, and assume that the variable it has in common with $\mathbf{xu}$ is u_n, the last variable in the sequence $\mathbf{u}$. (The argument in all other cases is similar.) This strong klukovits subset may contain many or no variables from $\mathbf{z}$, but we can split $\mathbf{z}$ into $\mathbf{z}_1$ and $\mathbf{z}_2$, with $\mathbf{z}_1$ possibly empty, such that

- $\mathbf{z}=\mathbf{z}_1\mathbf{z}_2$;
- $\mathbf{u}=\mathbf{u}_1u_n$; and
- $u_n\mathbf{z}_1$ is a strong klukovits subset of m.

If $k(u_n,\mathbf{z}_1,u)$ is the corresponding strong klukovits term, then from our assumption that

$$m(\mathbf{a},\mathbf{c}_1,c_n,\mathbf{e}_1,\mathbf{e}_2)=m(\mathbf{b},\mathbf{d}_1,d_n,\mathbf{e}_1,\mathbf{e}_2)\neq m(\mathbf{b},\mathbf{c}_1,c_n,\mathbf{e}_1,\mathbf{e}_2)$$

we have

$$\begin{aligned}\underline{m(\mathbf{a},\mathbf{c}_1,c_n\mathbf{e})} \;=\; m(\mathbf{a},\mathbf{c}_1,c_n,\mathbf{e}_1,\mathbf{e}_2) \;&= k(c_n,\mathbf{e}_1,m(\mathbf{a},\mathbf{c}_1,c_n,\mathbf{e}_1,\mathbf{e}_2))\\ &= k(c_n,\mathbf{e}_1,m(\mathbf{b},\mathbf{d}_1,d_n,\mathbf{e}_1,\mathbf{e}_2))\\ &= m(\mathbf{b},\mathbf{d}_1,c_n,\mathbf{e}_1,\mathbf{e}_2) \;=\; \underline{m(\mathbf{b},\mathbf{d}_1,c_n\mathbf{e})}\\ &\neq m(\mathbf{b},\mathbf{c}_1,c_n,\mathbf{e}_1,\mathbf{e}_2) \;=\; \underline{m(\mathbf{b},\mathbf{c}_1,c_n\mathbf{e})}.\end{aligned}$$

Hence, if we repartition the variables of m as $(\mathbf{x}\mid\mathbf{u}_1\mid u_n\mathbf{z})$ rather than $(\mathbf{x}\mid\mathbf{u}_1u_n\mid\mathbf{z})$, then we have constructed a new failure of condition (2) in $\mathbf{B}$ which has a smaller value of ℓ associated to it.

(2)$\Longrightarrow$(3). Condition (2) can be restated as follows: If $\mathbf{B}$ is any algebra in $\mathcal{V}(\mathbf{A})$ which has tuples $\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d}$ and $\mathbf{e}$, and if $t(\mathbf{x},\mathbf{y})$ is a polynomial of $\mathbf{B}$ of the form $m(\mathbf{x},\mathbf{y},\mathbf{e})$ for some maximal term m, then for the $1_{\mathbf{B}},1_{\mathbf{B}}$-matrix

$$\begin{bmatrix} u & v\\ w & z\end{bmatrix} := \begin{bmatrix} t(\mathbf{a},\mathbf{c}) & t(\mathbf{a},\mathbf{d})\\ t(\mathbf{b},\mathbf{c}) & t(\mathbf{b},\mathbf{d})\end{bmatrix}$$

we have $(u = z) \Longrightarrow (u = v = w = z)$. That is, we have the rectangularity implication for certain of the $1_{\mathbf{B}}, 1_{\mathbf{B}}$-matrices related to maximal terms.

If we have the rectangularity implication for matrices related to maximal terms, then it is not too hard to see that we also have it for all terms obtainable from maximal terms by identification of variables. Since every term operation is the interpretation of such a term (by the definition of "maximal term") we get the rectangularity implication for all $1_{\mathbf{B}}, 1_{\mathbf{B}}$-matrices. Thus $\mathbf{R}(1_{\mathbf{B}}, 1_{\mathbf{B}}; 0)$ holds for every $\mathbf{B}$ in $\mathcal{V}(\mathbf{A})$.

$(3) \Longrightarrow (4)$. First we explain why it is that if t is any term of any algebra, then $\mathsf{K}(t)$ is a subset system. To see that $\emptyset \in \mathsf{K}(t)$ we have to produce a strong klukovits term for the empty set, i. e., a term $k(u)$ such that

$$k(t(\mathbf{y})) = t(\mathbf{y})\,.$$

Just take $k(u) = u$. To see that $X(t) \in \mathsf{K}(t)$ we have to produce a term $k(\mathbf{x}, u)$ such that

$$k(\mathbf{x}, t(\mathbf{y})) = t(\mathbf{x})\,.$$

Just take $k(\mathbf{x}, u) = t(\mathbf{x})$. To see that $\mathsf{K}(t)$ is closed under union, we must show that if $t(\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z})$ has a strong klukovits term $k(\mathbf{w}, \mathbf{x}, u)$ for t in the variables $\mathbf{wx}$, and a strong klukovits term $k'(\mathbf{x}, \mathbf{y}, u)$ for t in the variables $\mathbf{xy}$, then there is a term $k''(\mathbf{w}, \mathbf{x}, \mathbf{y}, u)$ which is a strong klukovits term for t in $\mathbf{wxy}$. Such a k'' is

$$k''(\mathbf{w}, \mathbf{x}, \mathbf{y}, u) := k(\mathbf{w}, \mathbf{x}, k'(\mathbf{x}, \mathbf{y}, u))\,.$$

What remains to show is that if $\mathbf{A}$ is a finite algebra which generates a rectangular variety and m is a maximal term for $\mathbf{A}$, then the subset system $\mathsf{K}(m)$ separates the points of $X(m)$. This is the hard part of the proof of this theorem. Our arguments are inspired by the proof of Lemma 3.1 of [7].

Fix a maximal term m and arrange the variables in the order $m(w, x, \mathbf{y}, \mathbf{z})$ where $wx\mathbf{y}$ is a strong klukovits subset of the variables and we wish to separate the variables w and x. If $k(w, x, \mathbf{y}, u)$ is the strong klukovits term for m in $wx\mathbf{y}$, then we have the klukovits identity

$$k(w', x', \mathbf{y}', m(w, x, \mathbf{y}, \mathbf{z})) = m(w', x', \mathbf{y}', \mathbf{z})\,.$$

Let $\mathbf{F}$ be the free algebra in $\mathcal{V}$ generated by the set of variables $\{a, b, c, d, \mathbf{y}, u\}$. Let γ be the congruence of $\mathbf{F}$ generated by the pair $\langle g, h\rangle = \langle k(a, c, \mathbf{y}, u), k(b, d, \mathbf{y}, u)\rangle$. We have $k(a, d, \mathbf{y}, u)\ \gamma\ k(b, d, \mathbf{y}, u)$, since $\mathbf{F}/\gamma$ is rectangular, so there is a Mal'tsev chain in $\mathbf{F}$ connecting these elements by polynomial images of $\langle g, h\rangle$. By putting in trivial links if necessary we may assume that this chain has the form shown in Figure 4.

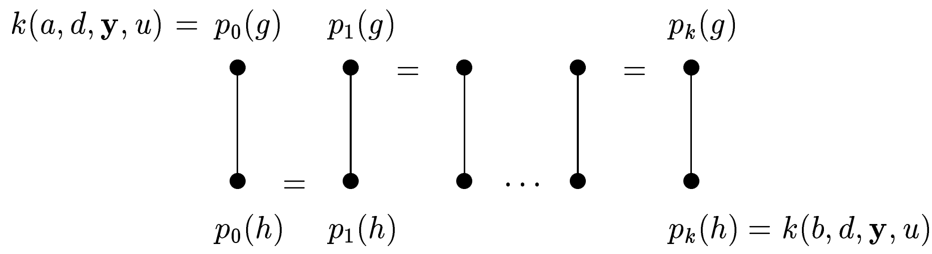

FIGURE 4: The Mal'tsev chain.

This Mal'tsev chain gives us a sequence of identities that hold in $\mathcal{V}(\mathbf{A})$: each polynomial $p_i(x)$ is of the form $s_i(x, a, b, c, d, \mathbf{y}, u)$, where s_i is a term, and the chain forces

$$\mathcal{V}(\mathbf{A}) \models s_i(k(a, c, \mathbf{y}, u), a, b, c, d, \mathbf{y}, u) = s_{i+1}(k(a, c, \mathbf{y}, u), a, b, c, d, \mathbf{y}, u)$$

when i is odd, and a similar identity with the first a and c replaced by b and d when i is even. Now we substitute $m(a', d', \mathbf{y}', \mathbf{z})$ for every occurrence of u in each of the identities determined by this chain. At the beginning of the chain this substitution gives us

$$k(a, d, \mathbf{y}, m(a', d', \mathbf{y}', \mathbf{z})) = m(a, d, \mathbf{y}, \mathbf{z})\,.$$

In the middle of the chain, at a typical point, we get

$$s_i(k(a, c, \mathbf{y}, m(a', d', \mathbf{y}', \mathbf{z})), a, b, c, d, \mathbf{y}, m(a', d', \mathbf{y}', \mathbf{z}))$$

(or the same expression, with b, d instead of a, c at the beginning) which simplifies to

$$s_i(m(a, c, \mathbf{y}, \mathbf{z}), a, b, c, d, \mathbf{y}, m(a', d', \mathbf{y}', \mathbf{z}))$$

using the klukovits identity. At the end of the chain we get

$$k(b, d, \mathbf{y}, m(a', d', \mathbf{y}', \mathbf{z})) = m(b, d, \mathbf{y}, \mathbf{z})\,.$$

Since m depends on all variables, $\mathbf{A} \not\models m(a, d, \mathbf{y}, \mathbf{z}) = m(b, d, \mathbf{y}, \mathbf{z})$, so the beginning and end of the chain are different. There is a first place in the chain where the substitution of $m(a', d', \mathbf{y}', \mathbf{z})$ into u does not yield $m(a, d, \mathbf{y}, \mathbf{z})$, and this implies the existence of some $s = s_i$ such that

$$\begin{aligned} m(a, d, \mathbf{y}, \mathbf{z}) &= s(m(a, c, \mathbf{y}, \mathbf{z}), a, b, c, d, \mathbf{y}, m(a', d', \mathbf{y}', \mathbf{z})) \\ &\neq s(m(b, d, \mathbf{y}, \mathbf{z}), a, b, c, d, \mathbf{y}, m(a', d', \mathbf{y}', \mathbf{z})), \end{aligned} \tag{7.1}$$

or else the same conclusion with the long expressions on the right interchanged. The argument forks now depending on which case we are in. Let's assume first that we are in the case displayed, and not in the case where we interchange the expressions. The equality in (7.1) tells us that the maximal term $m(a, d, \mathbf{y}, \mathbf{z})$ can be obtained by identification of the double primed variables with their unprimed counterparts in the term

$$s(m(a,c,\mathbf{y},\mathbf{z}),a'',b'',c'',d'',\mathbf{y}'',m(a',d',\mathbf{y}',\mathbf{z}'')). \tag{7.2}$$

By the maximality of m, this term cannot depend on more variables than m; but since we obtain $m(a,d,\mathbf{y},\mathbf{z})$ when we replace each double primed element by its unprimed mate, it follows that the term (7.2) depends on at least one of $\{a,a''\}$, on d'', on at least one variable from each set $\{y_i,y_i''\}$, and at least one of each $\{z_i,z_i''\}$. But this already accounts for the total number of variables on which this term can depend. Therefore the term (7.2) does not depend on $c,c'',b'',a',d',\mathbf{y}'$, and in every instance where we said "at least one" in the last sentence we have "exactly one". Moreover, since the term (7.2) does not depend on c or c'', one can use the inequality from (7.1) to prove that $s(m(x,c,\mathbf{y},\mathbf{z}),a,b,c,d,\mathbf{y},m(a',d',\mathbf{y}',\mathbf{z}))$ depends on x. Therefore, when choosing between a and a'', we see that the term (7.2) depends on a and not on a''.

Since the term (7.2) depends on exactly one of $\{y_i,y_i''\}$ for each i, it is possible to partition both $\mathbf{y}$ and $\mathbf{y}''$ in the same way so that $\mathbf{y}=\mathbf{y}_1\mathbf{y}_2,\mathbf{y}''=\mathbf{y}_1''\mathbf{y}_2''$ and the term (7.2) depends on $\mathbf{y}_2$ and $\mathbf{y}_1''$ but not on $\mathbf{y}_1$ and $\mathbf{y}_2''$. The same kind of partitioning can be done for $\mathbf{z}$ and $\mathbf{z}''$. (To maintain uniformity of notation, we also partition $\mathbf{y}'$ in (7.2) into $\mathbf{y}_1'\mathbf{y}_2'$ in the same way as $\mathbf{y}$ and $\mathbf{y}''$ even though (7.2) does not depend on $\mathbf{y}'$). This allows us to write the term in (7.2) as

$$s(m(\underline{a},c,\mathbf{y}_1\underline{\mathbf{y}_2},\mathbf{z}_1\underline{\mathbf{z}_2}),a'',b'',c'',\underline{d''},\underline{\mathbf{y}_1''}\mathbf{y}_2'',m(a',d',\mathbf{y}_1'\mathbf{y}_2',\underline{\mathbf{z}_1''}\mathbf{z}_2'')) \tag{7.3}$$

where this term depends only on the underlined variables. Let $\widehat{x}=(x,x,\dots,x)$ be a sequence of copies of x of the same length as $\mathbf{y}_2$, and define

$$K(x,\mathbf{y}_1,u):=s(u,x,x,x,x,\mathbf{y}_1\widehat{x},u)\,.$$

With this definition, and using our understanding of which variables the term (7.3) depends on, we calculate that $K(x',\mathbf{y}_1',m(w,x,\mathbf{y}_1\mathbf{y}_2,\mathbf{z}_1\mathbf{z}_2))$ is equal to

$$\begin{aligned} & s(m(\underline{w},x,\mathbf{y}_1\underline{\mathbf{y}_2},\mathbf{z}_1\underline{\mathbf{z}_2}),x',x',x',\underline{x'},\ \underline{\mathbf{y}_1'}\,\widehat{x'},\ m(w,\ x,\ \mathbf{y}_1\mathbf{y}_2,\underline{\mathbf{z}_1}\mathbf{z}_2)) \\ = \ & s(m(\underline{w},c,\mathbf{y}_1'\underline{\mathbf{y}_2},\mathbf{z}_1\underline{\mathbf{z}_2}),w\ ,\ b\ ,\ c\ ,\ \underline{x'},\ \underline{\mathbf{y}_1'}\,\mathbf{y}_2,m(a',d',\mathbf{y}_1'\mathbf{y}_2',\underline{\mathbf{z}_1}\mathbf{z}_2)). \end{aligned}$$

Using the equation in (7.1), we see that the second of these expressions simplifies to $m(w,x',\mathbf{y}_1'\mathbf{y}_2,\mathbf{z}_1\mathbf{z}_2)$. Therefore $K(x,\mathbf{y}_1,u)$ is a strong klukovits term for $m(w,x,\mathbf{y}_1\mathbf{y}_2,\mathbf{z}_1\mathbf{z}_2)$ in the variables $x\mathbf{y}_1$. The strong klukovits subset $x\mathbf{y}_1$ is contained in $wx\mathbf{y}_1\mathbf{y}_2=wx\mathbf{y}$ and it separates w and x.

Now, regarding the fork in the argument that took place earlier, if we had pursued the alternate fork with the same arguments as above, then we would have obtained a strong klukovits subset contained in $wx\mathbf{y}$ which contained w and not x instead of one that contained x and not w.

(4) $\Longrightarrow$(1). This follows from the final remark in the statement of Lemma 7.5. ☐

Is Theorem 7.7 more complicated than it has to be? The next example shows that it isn't, and in fact shows that the theorem gives the best information possible about the strong klukovits structure of terms in a rectangular variety.

Example 7.8. Theorem 7.7 shows that if $\mathbf{A}$ generates a rectangular variety, then the strong klukovits subsets $\mathsf{K}(m)$ of a maximal term m is a separating system of subsets of $X(m)$. We now show that if K is any separating system of subsets of a finite set X, then there is a finite algebra $\mathbf{A}$ which generates a rectangular variety and has a maximal term $m(x_1, \dots, x_n)$ such that $\mathsf{K}(m) = \mathsf{K}$.

First we construct a semigroup $\mathbf{S}$ on the disjoint union $\mathsf{K} \cup X \cup \{0\}$. All products in this semigroup are defined to be zero, with the following exceptions: For $U \in \mathsf{K}$ and $v \notin U$ define the product Uv to be v. For $U, V \in \mathsf{K}$ define the product $UV = U \cup V$. It is easy to check that $\mathbf{S}$ is a semigroup with zero. Let $\mathbf{R}$ denote the semigroup ring over the two-element field obtained from $\mathbf{S}$. Thus the elements of $\mathbf{R}$ are the formal sums of nonzero elements of S, with the empty sum corresponding to the zero element of $\mathbf{R}$ which we identify with the element $0 \in \mathbf{S}$.

Let $\mathbf{M}$ be the free $\mathbf{R}$-module on $|X| + 1$ generators. $\mathbf{M}$ is finite. We define an algebra $\mathbf{A}$, which is a reduct of $\mathbf{M}$, and which generates a rectangular variety. The universe of $\mathbf{A}$ is M and the basic operations of $\mathbf{A}$ will consist of the linear operations $s_0x_0 + \cdots + s_kx_k$, where $0 \le k \le |X|$, which have the properties that

(i) at most one coefficient is from K, the rest are distinct elements from X; and
(ii) if $s_i = U \in \mathsf{K}$, then all s_j, for $j \neq i$, are elements of U.

With these provisions it is easy to show that all nonzero term operations of $\mathbf{A}$ are represented by terms obtained from basic operations by identifying variables. The argument is as follows: call a term a *linear* term if the variables which label the leaves of its composition tree are distinct. Every term is derivable from a linear term by identification of variables, clearly. Therefore it suffices to show that every linear term is a basic operation. Equations that can be used to prove this are equations which reduce expressions like

(1) $Ux_0 + u_1x_1 + \cdots + u_{i-1}x_{i-1} + u_i(s_0y_0 + \cdots + s_\ell y_\ell) + u_{i+1}x_{i+1} + \cdots + u_kx_k$;
(2) $U(Vy_0 + v_1y_1 + \cdots + v_\ell y_\ell) + u_1x_1 + \cdots + u_kx_k$; and
(3) $U(v_1y_1 + \cdots + v_\ell y_\ell) + u_1x_1 + \cdots + u_kx_k$

to

(1)′ $Ux_0 + u_1x_1 + \cdots + u_{i-1}x_{i-1} + u_{i+1}x_{i+1} + \cdots + u_kx_k$;
(2)′ $(U \cup V)y_0 + v_{i_1}y_{i_1} + \cdots + v_{i_r}y_{i_r} + u_1x_1 + \cdots + u_kx_k$; and
(3)′ $v_{i_1}y_{i_1} + \cdots + v_{i_r}y_{i_r} + u_1x_1 + \cdots + u_kx_k$,

respectively, where in the latter two reductions $(v_{i_1}, \dots, v_{i_r})$ is the subsequence of $(v_1, \dots, v_\ell)$ of elements not in U.

The claims made above imply that $|X|+1$ is a bound on the essential arity of $\mathbf{A}$. In particular, every term operation which depends on all of its variables is the interpretation of a term obtained from a maximal term by identification of variables, and the maximal terms represent basic operations. Now it is not very hard to show that every maximal term has enough strong klukovits subsets. For example, consider a maximal term of the form

$$Ux_0 + u_1x_1 + \cdots + u_kx_k\,,$$

with $U' := \{u_1, \ldots, u_k\} \subseteq U$. Since K is a separating system of subsets, there exist

$$V_1 \subseteq V_2 \subseteq \cdots \subseteq V_k \subseteq U$$

with $V_i \in \mathsf{K}$ such that, after reordering the u_i's, we have $U' \cap V_i = \{u_i, \ldots, u_k\}$. Then the following subsets of $\{x_0, \ldots, x_k\}$ are strong klukovits subsets:

$$\emptyset \subseteq \{x_k\} \subseteq \{x_{k-1}, x_k\} \subseteq \cdots \subseteq \{x_0, x_1, \ldots, x_k\}\,.$$

A strong klukovits term in the last $k-i$ variables, for $0 \leq i \leq k$, is

$$k(x_i, \ldots, x_k, u) = V_iu + u_ix_i + \cdots + u_kx_k\,.$$

The case where the maximal term has no coefficient in K is handled similarly.

We now know that $\mathbf{A}$ generates a rectangular variety. If $X = \{m_1, m_2, \ldots, m_n\}$, then the term $m(\mathbf{x}) := m_1x_1 + m_2x_2 + \cdots m_nx_n$ is a maximal term which the reader can check has $\{x_{i_1}, \ldots, x_{i_j}\}$ as a strong klukovits subset if and only if $\{m_{i_1}, \ldots, m_{i_j}\} \in \mathsf{K}$. Therefore, the set $\mathsf{K}(m)$ can be identified naturally with K.

Example 7.9. We have seen that the maximal terms in a rectangular variety have enough strong klukovits subsets. The purpose of this example is to show that non-maximal terms may not have enough strong klukovits subsets. This example is a special case of the previous one. We take $X = \{a, b, c\}$ and $\mathsf{K} = \{\emptyset, \{a\}, \{a, b\}, \{a, b, c\}\}$. Then $m(x, y, z) := ax + by + cz$ is a maximal term of the algebra $\mathbf{A}$ constructed above. What we want to point out here is that the term $m(x, y, x) = (a + c)x + by$ has no strong klukovits subset of size one, so it is an example of a term of a rectangular variety which does not have enough strong klukovits subsets.

If $(a + c)x + by$ did have a one-element strong klukovits subset, then it could not be $\{x\}$, because the klukovits identity for the corresponding strong klukovits term $k(x', u) = rx' + su$ would require $s(a + c) = 0$ and $sb = b$. These equalities force $s \in \mathsf{K}$, $a, c \in s$ and $b \notin s$. However K has no such set s. If $\{y\}$ was a strong klukovits subset and $k(y', u) = ry' + su$ was the corresponding strong klukovits term, then we would have to have $sb = 0$ and $s(a + c) = a + c$. This forces $s \in \mathsf{K}$ and $b \in s$ while $a, c \notin s$. There is no element of this type in K, either.

The following theorem will be used to complete the proof of our Klukovits-type characterization theorem of locally finite rectangular varieties. But, at the same time, it is a nontrivial application of the strong klukovits terms which a rectangular variety possesses!

Theorem 7.10. *Let $\mathcal{V}$ be a locally finite rectangular variety. No algebra in $\mathcal{V}$ has a term operation of essential arity $\geq |F_{\mathcal{V}}(2)|$. Therefore $\mathcal{V}$ is generated by the finite algebra $\mathbf{F}_{\mathcal{V}}(|F_{\mathcal{V}}(2)|-1)$.*

Proof. Suppose, to get a contradiction, that $n \geq |F_{\mathcal{V}}(2)|$ and that $t(x_1,\ldots,x_n)$ is a term such that $t^{\mathbf{B}}$ depends on all variables in some algebra $\mathbf{B} \in \mathcal{V}$. The variety $\mathcal{V}(\mathbf{B})$ is rectangular, and finitely generated, and so we can apply Theorem 7.7 to it. The term t can be obtained from a maximal term for $\mathbf{B}$ by identifying variables. This maximal term still depends on at least n variables, and so by changing notation we may assume that t is actually a maximal term for $\mathbf{B}$.

Since $\mathsf{K}(t)$ is separating, we can rearrange the variables of t so that

$$\emptyset \subseteq \{x_1\} \subseteq \{x_1,x_2\} \subseteq \cdots \subseteq \{x_1,\ldots,x_{n-1}\} \subseteq \{x_1,x_2,\ldots,x_n\}$$

are strong klukovits subsets of t. Thus, for every $0 \leq i \leq n$ we have a term k_i such that $\mathbf{B}$ satisfies the identity

$$k_i(x_1,\ldots,x_i,t(y_1,\ldots,y_n)) = t(x_1,\ldots,x_i,y_{i+1},\ldots,y_n)\,.$$

Consider the binary terms $k_i'(x,y) = k_i(x,\ldots,x,y)$. These $n+1$ terms may be identified with elements of $F_{\mathcal{V}(\mathbf{B})}(2)$, which is a homomorphic image of $F_{\mathcal{V}}(2)$. Therefore at most n of these binary terms are distinct. There must exist $0 \leq i < j \leq n$ such that $k_i'(x,y) = k_j'(x,y)$ is an identity of $\mathbf{B}$. We shall get a contradiction by showing that $t(x_1,\ldots,x_n)$ does not depend on the variables $x_{i+1},\ldots,x_j$.

To simplify notation, we shall write the term $t(x_1,x_2,\ldots,x_n)$ as $t(\mathbf{x},\mathbf{y},\mathbf{z})$, where $\mathbf{x} = (x_1,\ldots,x_i)$, $\mathbf{y} = (x_{i+1},\ldots,x_j)$, $\mathbf{z} = (x_{j+1},\ldots,x_n)$. Here $\mathbf{x}$ and $\mathbf{z}$ may be empty sequences. For a variable u let $\widehat{u}$ denote any sequence of the form $(u,\ldots,u)$. With this new notation we have

$$k_i(\mathbf{x}',t(\mathbf{x},\mathbf{y},\mathbf{z})) = t(\mathbf{x}',\mathbf{y},\mathbf{z})\,,$$

and

$$k_j(\mathbf{x}',\mathbf{y}',t(\mathbf{x},\mathbf{y},\mathbf{z})) = t(\mathbf{x}',\mathbf{y}',\mathbf{z})\,.$$

By substituting u in every component of $\mathbf{x}'$ we get that

$$k_i'(u,t(\mathbf{x},\mathbf{y},\mathbf{z})) = t(\widehat{u},\mathbf{y},\mathbf{z})\,,$$

and similarly,

$$k_j'(u,t(\mathbf{x},\mathbf{y},\mathbf{z})) = t(\widehat{u},\widehat{u},\mathbf{z})\,.$$

From $\mathbf{B} \models k_i' = k_j'$, we see that $\mathbf{B} \models t(\widehat{u},\mathbf{y},\mathbf{z}) = t(\widehat{u},\widehat{u},\mathbf{z})$. Now apply k_i again to get

$$\mathbf{B} \models t(\mathbf{x},\mathbf{y},\mathbf{z}) = k_i(\mathbf{x},t(\widehat{u},\mathbf{y},\mathbf{z})) = k_i(\mathbf{x},t(\widehat{u},\widehat{u},\mathbf{z})) = t(\mathbf{x},\widehat{u},\mathbf{z})\,.$$

This equation shows that $t(\mathbf{x}, \mathbf{y}, \mathbf{z})$ does not depend on $\mathbf{y}$, which is a contradiction since t is a maximal term for $\mathbf{B}$.

The last statement of the theorem follows from the first, since the first statement shows that any subvariety of $\mathcal{V}$ which satisfies the same $(|F_{\mathcal{V}}(2)| - 1)$-variable equations as $\mathcal{V}$ must equal $\mathcal{V}$. □

Corollary 7.11. *A locally finite variety $\mathcal{V}$ is rectangular if and only if*

- *no algebra in $\mathcal{V}$ has a term operation of essential arity $\geq |F_{\mathcal{V}}(2)|$; and*
- *every maximal term has enough strong klukovits subsets.*

Proof. Theorem 7.10 shows that locally finite rectangular varieties have the first property and are finitely generated, while Theorem 7.7 proves that finitely generated rectangular varieties have the second property. Conversely, any variety with the first property is generated by the finite algebra $\mathbf{A} := \mathbf{F}_{\mathcal{V}}(|F_{\mathcal{V}}(2)| - 1)$, and Theorem 7.7 applied to this algebra proves that the second property is a necessary and sufficient condition for $\mathcal{V}$ to be rectangular. □

Corollary 7.12. *The class of finite algebras in a fixed finite language which generate a rectangular variety is recursive.*

Proof. If a finite algebra $\mathbf{A}$ in a finite language is given, then one can check if it is rectangular. This involves generating the subalgebra of all $1_{\mathbf{A}}, 1_{\mathbf{A}}$-matrices in $\mathbf{A}^4$. If $\mathbf{A}$ is rectangular, then according to Theorem 5.3 the essential arity of any term operation of $\mathbf{A}$ is at most $|A| - 1$, and all term operations of $\mathbf{A}$ can be calculated. Once this is done it is easy to verify whether the conditions of Corollary 7.11 hold. □

8. Residually Small Strongly Nilpotent Varieties

A variety is said to be *residually large* if it has a proper class of subdirectly irreducible members, and otherwise it is *residually small.* Residual smallness is a highly desirable property for a variety to have, and it seems to be an important requirement for any kind of structure theory.

In [15], Jacob Shapiro proved that every finitely generated strongly abelian variety of algebras has only finitely many subdirectly irreducible algebras, each of which is finite. In the same paper he pointed out that there are finite strongly abelian algebras which generate residually large varieties. The strongly abelian algebras he produced were (and *had to be*, according to his theorem) algebras which did not generate strongly abelian varieties. But one can easily construct strongly abelian algebras which do not generate strongly abelian varieties, yet which generate residually small varieties. The situation concerning the residual character of varieties generated by strongly abelian algebras has remained mysterious and unresolved for a long time. An essential feature of Shapiro's argument is the use of the fact that strongly abelian varieties have a bound on the essential arities of their terms. In fact, one might view his argument as a generalization of the natural proof that essentially unary varieties are residually small. The material from the preceding seven sections of this paper gives us a great deal of insight into the structure of varieties which have a bound on their essential arity, and in this section we intend to present the 'limiting version' of the Shapiro argument: we describe all locally finite, residually small varieties which have a bound on the essential arity of their term operations.

If $\mathcal{V}$ is a locally finite variety which has a bound on its essential arity, then every finite algebra in $\mathcal{V}$ must be strongly nilpotent, according to Theorem 4.8. A variety whose finitely generated members are strongly nilpotent is said to be *locally strongly nilpotent*. We shall prove local results about residual smallness which entail the following theorem.

Theorem 8.1. *Let $\mathcal{V}$ be a locally finite variety which is locally strongly nilpotent. Then $\mathcal{V}$ is residually small if and only if it is rectangular. Moreover, if $\mathcal{V}$ is residually small, then it has finitely many subdirectly irreducible algebras, all of which are finite.*

We point out that this theorem, coupled with Corollary 7.12, yields an algorithm to decide if a finite, strongly nilpotent algebra in a finite language generates a residually small variety. This fact contrasts with the main result of [10], which states that there is no algorithm to decide if an arbitrary finite algebra generates a residually small variety.

To prove this theorem we have two tasks. One task is to show that a locally finite variety which is locally strongly nilpotent but not rectangular is residually large. The other task, which we attend to immediately, is to show that a locally finite rectangular variety is residually small.

Definition 8.2. Let A be a set, and ρ an equivalence relation on A. We shall say that a function $f : A^n \to A$ *does not depend on its first variable on ρ,* if

$$f(a, \mathbf{u}) = f(b, \mathbf{u})$$

holds for every $(a, b) \in \rho, and\ \mathbf{u} \in A^{n-1}$.

By permuting the variables of f we can speak of f depending on any given variable on ρ, and about the essential arity of f on ρ as well.

Lemma 8.3. *Let $\mathbf{S}$ be a subdirectly irreducible algebra, and α a rectangular congruence of $\mathbf{S}$. Suppose that every term operation of $\mathbf{S}$ depends on at most r variables on α. Let $m = |S/\alpha|$, and $M = |F_{\mathcal{V}(\mathbf{S})}(r)|$. Then*

$$|S| \leq m \cdot 2^{M \cdot m^{r-1}}.$$

Proof. Let μ be the monolith of $\mathbf{S}$ and choose $(a, b) \in \mu - 0_\mathbf{S}$. Let C be an arbitrary α-class. For each pair $(c, d) \in C^2 - 0_\mathbf{S}$ there is a Mal'tsev chain witnessing that a and b are congruent modulo the congruence generated by c and d. By looking at the first nontrivial link in this chain starting from a we see that $\mathbf{S}$ has a unary polynomial $p(x, \mathbf{u})$, where p is a term, such that one of $p(c, \mathbf{u})$ and $p(d, \mathbf{u})$ is equal to a and the other is not. For the undirected pair $\{c, d\}$ in C define the set $L\{c, d\}$ to contain, for every such term p and parameter sequence $\mathbf{u}$, the pair $(p, \mathbf{u})$. We may, and do, assume that p depends on all of its variables on α.

Consider the union U of all sets of labels $L\{c, d\}$, where $c \neq d \in C$. Call two elements $(p, \mathbf{u})$ and $(q, \mathbf{v})$ of U equivalent, if $p = q$ is an identity of $\mathbf{S}$, and $\mathbf{u}\ \alpha\ \mathbf{v}$ holds componentwise. Pick a system R of representatives from each equivalence class. We show that $|C| \leq 2^{|R|}$.

Indeed, consider, for each element $c \in C$, the set

$$P_c = \{(p, \mathbf{v}) \in R \mid p(c, \mathbf{v}) = a\}.$$

It is sufficient to show that if $c \neq d$, then $P_c \neq P_d$. Pick an element $(p, \mathbf{u}) \in L\{c, d\}$, and let its representative in R be $(p, \mathbf{u}')$. We show that $(p, \mathbf{u}')$ is contained in exactly one of the sets P_c and P_d. The pair $(p, \mathbf{u}') \in R \subseteq U$, so there exist $c' \neq d' \in C$ such that $(p, \mathbf{u}') \in L\{c', d'\}$. Interchanging c with d and c' with d' if necessary we get that

$$p(d, \mathbf{u}) \neq p(c, \mathbf{u}) = a = p(c', \mathbf{u}') \neq p(d', \mathbf{u}')\,.$$

By the rectangularity of α we get that $p(c, \mathbf{u}') = a$, and so it is sufficient to show that $p(d, \mathbf{u}') \neq a$. Suppose instead that $p(d, \mathbf{u}') = a = p(c, \mathbf{u})$. Then by rectangularity we get that $p(d, \mathbf{u}) = a$, which is false.

It remains to count the elements of R. Each element is determined by a term p, and a parameter sequence $\mathbf{u}$. Since p depends on all of its variables on α, its arity is at most r. The number of such terms is at most $M = |\mathbf{F}_{\mathcal{V}(\mathbf{S})}(r)|$. The length of $\mathbf{u}$ is at most $r - 1$, and each component can take at most $|S/\alpha| = m$ values. Thus, the number of such $\mathbf{u}$ is at most m^{r-1}, showing that $|R| \leq M \cdot m^{r-1}$. Since α has m classes, from this estimate on $|C|$ we get the formula to be proved. □

Corollary 8.4. *A locally finite rectangular variety is residually small.*

Proof. Let $\mathcal{V}$ be a locally finite rectangular variety which contains the subdirectly irreducible algebra $\mathbf{S}$. In the previous lemma we can take $\alpha = 1_{\mathbf{S}}$, in which case $m = 1$. Furthermore, if $N = |F_{\mathcal{V}}(2)|$, then by Theorem 7.10 we can take $r = N - 1$. This yields that $M = |F_{\mathcal{V}}(N - 1)|$. Thus, Lemma 8.3 gives a finite cardinality bound on $\mathbf{S}$ valid for all subdirectly irreducible algebras in $\mathcal{V}$. □

Since the bound on the cardinality of the subdirectly irreducible algebras in $\mathcal{V}$ is a finite bound which is determined by the free spectrum of $\mathcal{V}$, it follows that $\mathcal{V}$ has only finitely many subdirectly irreducible algebras and all are finite.

We prove in Theorem 8.6 that a locally strongly nilpotent variety which is not rectangular is residually large. In fact we shall prove that if a finite algebra has a strongly nilpotent *tolerance* which is not rectangular, then it generates a residually large variety. This result will be derived from the following lemma which describes a method for constructing large subdirectly irreducibles.

For a binary relation T on a set A, and a finite or infinite cardinal κ set

$$T^{[\kappa]} = \{\mathbf{x} \in A^\kappa \mid (x_i, x_j) \in T \text{ for all } i < j < \kappa\}.$$

Lemma 8.5. *Let*

(1) $\mathbf{A}$ *be a finite algebra;*
(2) α *a minimal congruence of* $\mathbf{A}$*;*
(3) N *a* $\langle 0_{\mathbf{A}}, \alpha\rangle$*-trace;*
(4) $u, v, w \in N$*, which are not all equal;*
(5) T *a tolerance of* $\mathbf{A}$*;*
(6) $(a^\ell, b^\ell) \in T$ *for* $1 \le \ell \le m$*.*

Suppose that

(7) $\mathbf{R}(T, N^2; 0_{\mathbf{A}})$ *holds;*
(8) $(u, v, w) \in T^{[3]}$*;*
(9) *The congruence of* $\mathbf{T}^{[3]}$ *generated by collapsing* (b^ℓ, a^ℓ, a^ℓ) *with* (a^ℓ, a^ℓ, b^ℓ) *for every* $1 \le \ell \le m$ *collapses* (v, v, v) *with* (u, v, w)*.*

Then $\mathcal{V}(\mathbf{A})$ *is residually large.*

Proof. Let κ be an infinite cardinal. Let $\mathbf{B}$ be the subalgebra of $\mathbf{A}^\kappa$ whose universe is $B = T^{[\kappa]}$. For each polynomial operation p of $\mathbf{A}$ let $\hat{p}$ be the polynomial of $\mathbf{B}$ which is p in each coordinate; for each $a \in A$ let $\hat{a}$ be the element of B which is a in every coordinate. Let $M = N^\kappa \cap B$. The fact that u, v and w are T-related, in N and not all equal implies that M contains an element $\mathbf{z}$ which is not of the form $\hat{a}$. Also, by conditions (7), (8) and Lemma 2.3 we know that the type of N is $\mathbf{1}$.

Let $(f_i \mid i < \kappa)$ be an induced operation of $\mathbf{B}|_M$. Each f_i is an operation of $\mathbf{A}|_N$, so it is an essentially unary operation. If one of these is nonconstant, then, since the f_i are T-twin polynomials, Lemma 2.4 (7) and our hypothesis (7) guarantee that the f_i are all equal permutations of N. This means that $(f_i \mid i < \kappa)$ is constant or it agrees with some $\hat{\pi}$, where π is a unary polynomial permutation of $\mathbf{A}|_N$. From this one deduces that $\mathbf{B}|_M$ is polynomially equivalent to a G-set which is a diagonal subdirect power of the G-set $\mathbf{A}|_N$. The diagonal is an orbit of the G-set structure, so $\mathbf{B}|_M$ has a congruence which partitions M into two classes: the diagonal and the off-diagonal. Since M is an E-trace, this congruence can be extended to a congruence ψ of $\mathbf{B}$ (see Lemma 2.4 in [4]).

Let ψ_0 be a maximal congruence of $\mathbf{B}$ containing ψ that separates $\hat{v}$ and $\mathbf{z}$. Then $\mathbf{B}/\psi_0$ is a a subdirectly irreducible factor of $\mathbf{B}$, let λ denote its cardinality. To finish the proof it is sufficient to show that $\lambda \ge \kappa$.

Let ${}^i\mathbf{d}^\ell$ be the element of B whose every component is a^ℓ except that the i-th component is b^ℓ. We show that for each $i \ne j < \kappa$ there exists an $1 \le \ell \le m$ such that $({}^i\mathbf{d}^\ell, {}^j\mathbf{d}^\ell) \notin \psi_0$. Indeed, suppose that this fails for some $i \ne j$. Let $\mathbf{C}$ be the subalgebra of $\mathbf{B}$ consisting of those functions that are constant on the set $\kappa - \{i, j\}$. $\mathbf{C}$ is a retract of $\mathbf{B}$ which is isomorphic to $\mathbf{T}^{[3]}$. To see this, choose $k \notin \{i, j\}$. A retraction $\rho : \mathbf{B} \to \mathbf{C}$ is given by defining $\rho(b)_x = b_x$ if $x \in \{i, j\}$ and $\rho(b)_x = b_k$ otherwise. An isomorphism $\varphi : \mathbf{C} \to \mathbf{T}^{[3]}$ is given by $\varphi(c) = (c_i, c_k, c_j)$.

Observe that for all ℓ the elements ${}^i\mathbf{d}^\ell$ and ${}^j\mathbf{d}^\ell$ are in $\mathbf{C}$ and correspond under φ to the elements (b^ℓ, a^ℓ, a^ℓ) and $(a^\ell, a^\ell, b^\ell) \in T^{[3]}$ respectively. Condition (9) of the lemma and the observations of the last paragraph imply that the congruence of $\mathbf{C}$ generated by all pairs $({}^i\mathbf{d}^\ell, {}^j\mathbf{d}^\ell)$ contains the pair $(\hat{v}, \mathbf{z}')$ where $\mathbf{z}'$ denotes the element of $\mathbf{C}$ which is u in the i-th coordinate, w in the j-th coordinate and v elsewhere. Since $\mathbf{C} \leq \mathbf{B}$ and all pairs of the form $({}^i\mathbf{d}^\ell, {}^j\mathbf{d}^\ell)$ belong to ψ_0, by assumption, we must have $(\hat{v}, \mathbf{z}') \in \psi_0$. But $\mathbf{z}' \in M$ is outside the diagonal, hence $\mathbf{z}'\ \psi\ \mathbf{z}$. Therefore $\hat{v}\ \psi_0\ \mathbf{z}$, which is contrary to the definition of ψ_0. This proves our claim that for each $i \neq j < \kappa$ there exists an $1 \leq \ell \leq m$ such that $({}^i\mathbf{d}^\ell, {}^j\mathbf{d}^\ell) \notin \psi_0$.

To finish the proof, define a mapping $g : \kappa \to (B/\psi_0)^m$ by $g(i) = ({}^i\mathbf{d}^1, \ldots, {}^i\mathbf{d}^m)$. What we have just proved means exactly that g is injective. Therefore $\kappa \leq \lambda^m = \lambda$, proving the statement of the lemma. □

Theorem 8.6. *If* $\mathbf{A}$ *is a finite algebra in a residually small variety, then every strongly nilpotent tolerance of* $\mathbf{A}$ *is rectangular.*

Proof. Choose $\mathbf{A}$ to be a counterexample of minimal cardinality, and let α be a minimal congruence of $\mathbf{A}$. Our aim is to find a failure of $\mathbf{R}(T, T; 0_\mathbf{A})$ such that the corresponding T, T-matrix is contained in a single $\langle 0_\mathbf{A}, \alpha\rangle$-trace, and then to apply Lemma 8.5 with u, v, w being the entries of this matrix.

Since T is not rectangular, there exists a polynomial t and T-related pairs of vectors $(\mathbf{a}, \mathbf{b})$, and $(\mathbf{c}, \mathbf{d})$ in A such that

$$t(\mathbf{a}, \mathbf{c}) = t(\mathbf{b}, \mathbf{d}) \neq t(\mathbf{a}, \mathbf{d})\,.$$

By Lemma 3.5 (4), the tolerance T/α of $\mathbf{A}/\alpha$ is strongly nilpotent, and so it must be rectangular by the minimality of $\mathbf{A}$. Therefore all four entries of the T, T-matrix

$$\begin{bmatrix} v & u \\ w & v \end{bmatrix} := \begin{bmatrix} t(\mathbf{a}, \mathbf{c}) & t(\mathbf{a}, \mathbf{d}) \\ t(\mathbf{b}, \mathbf{c}) & t(\mathbf{b}, \mathbf{d}) \end{bmatrix}$$

are α-related. Choose a unary polynomial p that maps the α-class containing u, v and w into a $\langle 0_\mathbf{A}, \alpha\rangle$-trace N, and separates u and v. Then $p(w) \in N$ as well. Now replace t by $p \circ t$. This gives a new failure of $\mathbf{R}(T, T; 0_\mathbf{A})$, where the entries are already in N. By changing notation we shall assume that t has already been prefixed by p, and so $u, v, w \in N$.

We now show that the conditions of Lemma 8.5 are satisfied. The elements a^ℓ and b^ℓ will be the components of $\mathbf{a}$ and $\mathbf{b}$, respectively. From the fact that T is strongly nilpotent we get that $\mathbf{R}(T, N^2; 0_\mathbf{A})$ holds. Therefore the only condition to be checked is (9). We define an m-ary polynomial on $\mathbf{T}^{[3]}$ by

$$q((x_1, y_1, z_1), \ldots, (x_m, y_m, z_m)) = (t(\mathbf{x}, \mathbf{d}), t(\mathbf{y}, \mathbf{c}), t(\mathbf{z}, \mathbf{c}))\,.$$

Then

$$q((b^1, a^1, a^1), \ldots, (b^m, a^m, a^m)) = (v, v, v)\,,$$

and

$$q((a^1, a^1, b^1), \ldots, (a^m, a^m, b^m)) = (u, v, w)\,,$$

hence (9) is satisfied. □

This theorem supplies the missing direction of Theorem 8.1.

Example 8.7. We apply the results of the last two sections to demonstrate the existence of a surprising example: we will show that there is a finite, simple, strongly abelian algebra that generates a residually large variety.

To construct the example, let $U = \{0, 1, \ldots, n\}$, where $n \geq 3$. Suppose that f is the unary operation on U defined by $f(i) = i - 1$ for all $i > 0$ and $f(0) = 0$. Let $\mathbf{A}$ be the algebra with universe U^2 (whose elements we write as columns), and with three types of defining operations. First, we equip $\mathbf{A}$ with all unary operations h of the form

$$h\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} r(x) \\ s(y) \end{bmatrix}$$

where r and s are unary operations on U with $|r(U)|, |s(U)| \leq 2$. Next we add all unary operations h' of the form

$$h'\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} r(y) \\ s(x) \end{bmatrix}$$

where again r and s are unary operations on U with $|r(U)|, |s(U)| \leq 2$. Finally we add the binary operation g, defined by

$$g\left(\begin{bmatrix} x \\ y \end{bmatrix}, \begin{bmatrix} u \\ v \end{bmatrix}\right) = \begin{bmatrix} f(y) \\ f(u) \end{bmatrix}.$$

It is easy to verify that $\mathbf{A}$ is simple and strongly abelian. In particular, it is strongly nilpotent. By Theorem 8.1, to show that $\mathbf{A}$ generates a residually large variety it suffices to prove that the generated variety fails to be rectangular. For this we must exhibit a maximal term which does not have enough strong klukovits subsets.

It is obvious that $g(x, y)$ is a maximal term since it depends on both variables and (it can be argued that) no term of $\mathbf{A}$ depends on more than two variables. If $g(x, y)$ had a one-element strong klukovits subset, then there would have to be a strong klukovits term k such that $k(z, g(x, y)) = g(z, y)$ or $g(x, z)$. The term k would have to be essentially binary and the range of k would have to contain the range of g, which is $(U - \{n\}) \times (U - \{n\})$. Analyzing the possibilities, one finds that up to equivalence the only essentially binary terms of $\mathbf{A}$ whose range contains $(U - \{n\}) \times (U - \{n\})$ are $g(x, y)$, and $g(y, x)$. An easy computation shows that neither of these are strong klukovits terms for g in either variable. According to Theorem 7.7 the variety $\mathcal{V}(\mathbf{A})$ is not rectangular.

References

[1] J. Berman, E. W. Kiss, P. Pröhle and Á. Szendrei. The set of types of a finitely generated variety. *Discrete Math*, 112:1–20, (1993).

[2] P. Crawley and R. P. Dilworth. *Algebraic Theory of Lattices*, Prentice Hall, 1973.

[3] G. Higman. The orders of relatively free groups. *Proc. Intern. Conf. Theory of Groups,* Austral. Nat. Univ. Canberra, 153–165, (1965).

[4] D. Hobby and R. McKenzie. *The Structure of Finite Algebras*, volume 76 of *Contemporary Mathematics.* American Mathematical Society, 1988.

[5] K. Kearnes. An order-theoretic property of the commutator. *International Journal of Algebra and Computation*, 3:491–533, (1993).

[6] E. W. Kiss. An easy way to minimal algebras. *International Journal of Algebra and Computation*, 7:55–75, (1997).

[7] E. W. Kiss and M. Valeriote. Strongly Abelian varieties and the Hamiltonian property. *The Canadian Journal of Mathematics*, 43:331–346, (1991).

[8] E. W. Kiss and M. Valeriote. Abelian varieties and the Hamiltonian property. *Journal of Pure and Applied Algebra*, 87:37–49, (1993).

[9] L. Klukovits. Hamiltonian varieties of universal algebras. *Acta. Sci. Math.*, 37:11–15, (1975).

[10] R. McKenzie. The residual bound of a finite algebra is not computable. *International Journal of Algebra and Computation*, 6:29–48, (1996).

[11] R. McKenzie, G. McNulty, and W. Taylor. *Algebras, Lattices, Varieties Volume 1.* Wadsworth and Brooks/Cole, Monterey, California, 1987.

[12] P. M. Neumann. Some indecomposable varieties of groups. *Quart. J. Math. Oxford*, 14:46–50, (1963).

[13] J. Płonka. Diagonal algebras. *Fund. Math.*, 58:309–321, (1966).

[14] R. Pöschel, M. Reichel. Projection algebras and rectangular algebras. In *General Algebra and Applications. Research and Exposition in Math.* 20:180–194. Heldermann Verlag, Berlin, 1993.

[15] J. Shapiro. Finite algebras with Abelian properties. *Algebra Universalis*, 25:334–364, (1988).

(Keith A. Kearnes) Department of Mathematics, University of Louisville, Louisville, Kentucky 40292, USA

(Emil W. Kiss) Eötvös University, Department of Algebra and Number Theory, 1088 Budapest, Múzeum krt. 6–8, Hungary

E-mail address, Keith A. Kearnes: `kakear01@homer.louisville.edu`

E-mail address, Emil W. Kiss: `ewkiss@cs.elte.hu`