APPROXIMATE DISTRIBUTIVE LAWS AND FINITE EQUATIONAL BASES FOR FINITE ALGEBRAS IN CONGRUENCE-DISTRIBUTIVE VARIETIES

KIRBY A. BAKER AND JU WANG

Abstract. For a congruence-distributive variety, Maltsev's construction of principal congruence relations is shown to lead to approximate distributive laws in the lattice of equivalence relations on each member. In the case of a variety generated by a finite algebra, these approximate laws then yield two known results: the boundedness of the complexity of unary polynomials needed in Maltsev's construction, from which follows the finite equational basis theorem for such a variety. An algorithmic version of the construction is included.

1. Introduction

We present a calculus of equivalence relations that quantifies Maltsev's construction of principal congruence relations ([16], Theorem 1.20) to show how, in a congruence-distributive variety, distributive laws hold for equivalence relations after they have been adjusted by unary polynomial functions of a certain nesting depth. This theory illuminates two results. The first, due to the second author [18], is that in a congruence-distributive variety generated by a finite algebra, the nesting depth of the unary polynomials involved in Maltsev's construction of principal congruence relations can be bounded. The second is the theorem, due to the first author [4], that a congruence-distributive variety generated by a finite algebra of finite type has a finite equational basis. Several proofs of this result are in the literature [4, 14, 17, 13, 8]; see also [7]. In [19] it is shown how this theorem follows from the boundedness theorem, to produce an explicit equational basis while appealing neither

Date: August 30, 2001.

1991 *Mathematics Subject Classification.* Primary: 08B10; Secondary: 08A30, 08B26.

Key words and phrases. congruence distributive, principal congruence, finite basis.

Work of the second author supported by Chinese National Technology Project 97-3.

to Ramsey's theorem as in [4] or to the compactness theorem of first-order logic in some form [14, 17, 13, 8, 15]. An improved version of this construction is included in algorithmic form.

For a history of the question of finite equational bases for finite algebras, see [4, 6, 15]. For general terminology and standard concepts, see [16, 9, 11].

2. Approximate distributive laws for equivalence relations

By an *operational* (or *basic*) *unary polynomial* on an algebra $\mathbf{A}$ we mean any polynomial function obtained by freezing all entries except one in a basic operation of $\mathbf{A}$. Let O_A be the set of all operational unary polynomials on $\mathbf{A}$. For $\alpha \in \mathbf{Eqv}(A)$, the lattice of equivalence relations on the universe A of $\mathbf{A}$, let $\mathcal{O}\alpha$ be the equivalence relation on A generated by $\alpha \cup \{\langle p(a), p(b)\rangle : p \in O_A, \langle a, b\rangle \in \alpha\}$. Thus $\mathcal{O}$ is an operator on $\mathbf{Eqv}(A)$. We write $\mathcal{O}_A$ for $\mathcal{O}$ when needed for clarity.

Observations.

2.1: $\mathcal{O}$ is a complete join-endomorphism of $\mathbf{Eqv}(A)$. In particular, $\mathcal{O}$ is isotone, and therefore

2.2: $\mathcal{O}(\alpha \cap \beta) \subseteq \mathcal{O}\alpha \cap \mathcal{O}\beta$ for $\alpha, \beta \in \mathbf{Eqv}(A)$.

2.3: The fixed points of $\mathcal{O}$ are the congruence relations of $\mathbf{A}$.

2.4: $\mathrm{Cg}\,\alpha = \cup_{k=0}^{\infty} \mathcal{O}^k\alpha$, where $\mathrm{Cg}\,\alpha$ denotes the smallest congruence relation on $\mathbf{A}$ that contains α. (This is a variant of Maltsev's construction of principal congruence relations; see [16], Theorem 1.20, [10] and §4 below.)

2.5: If p is a unary polynomial function on $\mathbf{A}$ obtained by freezing all entries but one in a term of depth D, then $\langle a, b\rangle \in \alpha$ implies $\langle p(a), p(b)\rangle \in \mathcal{O}^D\alpha$. (Here "depth" is nesting depth in the construction of the term.)

2.6: For an integer $M \geq 0$, by a *linear unary polynomial of depth* M let us mean a composition of M operational unary polynomials of $\mathbf{A}$ (where a composition of no polynomials is the identity function). If some linear unary polynomial of depth at most M takes a pair $\langle a_0, a_1\rangle$ to the pair $\langle b_0, b_1\rangle$ in $\mathbf{A}$, let us say that $\langle b_0, b_1\rangle$ is *weakly projective* to $\langle a_0, a_1\rangle$ in at most M steps, written $\langle a_0, a_1\rangle \to_M \langle b_0, b_1\rangle$. Then $\mathcal{O}^M\alpha$ is the equivalence relation on A generated by all pairs weakly projective to pairs in α in at most M steps.

2.7: If $f : \mathbf{A} \to \mathbf{B}$ is a surjective homomorphism, then $f \circ \mathcal{O}_A = \mathcal{O}_B \circ f$.

Here $f(\alpha)$ for $\alpha \in \mathbf{Eqv}(A)$ means the smallest member of $\mathbf{Eqv}(B)$ containing all pairs $\langle f(a_1), f(a_2)\rangle$ for $\langle a_1, a_2\rangle \in \alpha$.

2.8 Lemma. Let V be a congruence-distributive variety, fix Jónsson terms $t_0, \ldots, t_n$, and let D be the maximum of their depths. Then for $\mathbf{A} \in V$ and $\alpha, \beta_1, \cdots \beta_m \in \mathbf{Eqv}(A)$ we have these variants of the distributive laws:

(i): $\alpha \cap (\beta_1 \vee \cdots \vee \beta_m) \subseteq (\mathcal{O}^D\alpha \cap \mathcal{O}^D\beta_1) \vee \cdots \vee (\mathcal{O}^D\alpha \cap \mathcal{O}^D\beta_m)$

(ii): $(\alpha \vee \beta_1) \cap \cdots \cap (\alpha \vee \beta_m) \subseteq \mathcal{O}^{2DL(m)}\alpha \vee (\mathcal{O}^{2DL(m)}\beta_1 \cap \cdots \cap \mathcal{O}^{2DL(m)}\beta_m)$, where $L(m) = \lceil \log_2 m \rceil$.

Proof of Lemma 2.8. For (i): Suppose $\langle a, c\rangle \in \alpha \cap (\beta_1 \vee \cdots \vee \beta_\ell)$. Then $\langle a, c\rangle \in \alpha$ and also a and c are connected by a sequence $a = b_0, \ldots, b_m = c$ such that for each $j = 1, \ldots, m$ we have $\langle b_{j-1}, b_j\rangle \in \beta_{k(j)}$ for some $k(j)$. Let $s_{ij} = t_i(a, b_j, c)$ for each relevant i, j. As in [2], Jónsson's laws relating the t_i give a "zig-zag" sequence $a = s_{10}, s_{11}, \ldots, s_{1m}{=}s_{2m}, s_{2,m-1}, \ldots, s_{20}{=}s_{30}, s_{31}, \ldots$, etc., with the second subscript alternately increasing and decreasing, ending with c in the guise of $s_{n-1,m}$ if n is even or of $s_{n-1,0}$ if n is odd. Let us examine relations between adjacent terms $s_{i,j-1}$ and s_{ij}. Since these terms are obtained by evaluating the unary polynomial $t_i(a, \cdot, c)$ at b_{j-1} and b_j respectively, by Observation 2.5 we have $\langle s_{i,j-1}, s_{ij}\rangle \in \mathcal{O}^D\beta_{k(j)}$. By the same observation applied to the unary polynomial $t_i(a, b_j, \cdot)$ evaluated at c and a, we have $\langle s_{ij}, a\rangle \in \mathcal{O}^D\alpha$ (since $t_i(a, b_j, a) = a$); similarly $\langle s_{i,j-1}, a\rangle \in \mathcal{O}^D\alpha$, so $\langle s_{i,j-1}, s_{ij}\rangle \in \mathcal{O}^D\alpha$. Via the zig-zag sequence, then, $\langle a, c\rangle \in (\mathcal{O}^d\alpha \cap \mathcal{O}^D\beta_1) \vee \cdots \vee (\mathcal{O}^d\alpha \cap \mathcal{O}^D\beta_\ell)$, as required.

For (ii): For $m = 2$, if α, β_1, β_2 were actually congruence relations, by congruence-distributivity we would have the derivation $(\alpha \vee \beta_1) \cap (\alpha \vee \beta_2) \subseteq [(\alpha\vee\beta_1)\cap\alpha]\vee[(\alpha\vee\beta_1)\cap\beta_2] \subseteq \alpha\vee[(\alpha\cap\beta_2)\vee(\beta_1\cap\beta_2)] \subseteq \alpha\vee(\beta_1\cap\beta_2)$. Since α, β_1, β_2 are not necessarily congruence relations, however, we invoke (i) twice and use Observations 2.1 and 2.2 to distribute powers of $\mathcal{O}$ through meets and joins: $(\alpha \vee \beta_1) \cap (\alpha \vee \beta_2) \subseteq [(\mathcal{O}^D\alpha \vee \mathcal{O}^d\beta_1) \cap \mathcal{O}^D\alpha] \vee [(\mathcal{O}^D\alpha \vee \mathcal{O}^D\beta_1) \cap \mathcal{O}^D\beta_2]$
$\subseteq \mathcal{O}^D\alpha \vee [(\mathcal{O}^{2D}\alpha \cap \mathcal{O}^{2D}\beta_2) \vee (\mathcal{O}^{2D}\beta_1 \cap \mathcal{O}^{2D}\beta_2)] \subseteq \mathcal{O}^{2D}\alpha \vee (\mathcal{O}^{2D}\beta_1 \cap \mathcal{O}^{2D}\beta_2)$.

For general m, it suffices to check the case where m is a power of 2, which is accomplished by using the case $m = 2$ recursively: $(\alpha \vee \beta_1) \cap \cdots\cap(\alpha\vee\beta_{2m}) \subseteq [\mathcal{O}^{2DL(m)}\alpha\vee(\mathcal{O}^{2DL(m)}\beta_1\cap\cdots\cap\mathcal{O}^{2DL(m)}\beta_m)]\cap[\mathcal{O}^{2DL(m)}\vee(\mathcal{O}^{2DL(m)}\beta_{m+1} \cap \cdots \cap \mathcal{O}^{2DL(m)}\beta_{2m})] \subseteq \mathcal{O}^{2D+2DL(m)}\alpha \vee (\mathcal{O}^{2D+2DL(m)}\beta_1 \cap \cdots \cap \mathcal{O}^{2D+2DL(m)}\beta_{2m}) = \mathcal{O}^{2DL(2m)}\alpha \vee (\mathcal{O}^{2DL(2m)}\beta_1 \cap \cdots \cap \mathcal{O}^{2DL(2m)}\beta_{2m})$.
□

2.9 Definition. For $\alpha \in \mathbf{Eqv}(A)$, $\mathcal{O}^{-1}\alpha$ is the largest $\beta \in \mathbf{Eqv}(A)$ with $\mathcal{O}\beta \subseteq \alpha$, or equivalently (in view of Observation 2.1), $\mathcal{O}^{-1}\alpha = \vee\{\theta \in \mathbf{Eqv}(A) : \mathcal{O}\theta \subseteq \alpha\}$.

Observations.

2.10 : $\mathcal{O}^{-1}$ is a complete meet-endomorphism of $\mathbf{Eqv}(A)$.

2.11 : The fixed points of $\mathcal{O}^{-1}$ are the congruence relations on $\mathbf{A}$.

2.12 : $\mathcal{O}^{-k}\alpha$, in the sense of $\mathcal{O}^{-1}(\mathcal{O}^{-1}(\cdots(\mathcal{O}^{-1}(\alpha))\cdots))$ (k times), equals $\vee\{\theta \in \mathbf{Eqv}(A) : \mathcal{O}^k\theta \subseteq \alpha\}$.

Here is another kind of approximate distributive law, one that will actually be used in what follows; its virtue is that the exponent of α does not depend on m:

2.13 Lemma.
$\mathcal{O}^{-mD}[(\alpha \vee \beta_1) \cap \cdots \cap (\alpha \vee \beta_m)] \subseteq \mathcal{O}^D\alpha \vee (\mathcal{O}^{mD}\beta_1 \cap \cdots \cap \mathcal{O}^{mD}\beta_m)$.

Proof. For convenience, for $k = 0, \ldots, m$ write $\rho_k = \mathcal{O}^{kD}\beta_1 \cap \cdots \cap \mathcal{O}^{kD}\beta_k$. (Thus $\rho_0 = 1 \in \mathbf{Eqv}(A)$ and ρ_m occurs in the statement of the Lemma.) Let θ be such that $\mathcal{O}^{mD}\theta \subseteq (\alpha \vee \beta_1) \cap \cdots \cap (\alpha \vee \beta_m)$; we must show that $\theta \subseteq \mathcal{O}^D\alpha \vee (\mathcal{O}^{mD}\beta_1 \cap \cdots \cap \mathcal{O}^{mD}\beta_m)$. We first prove this claim:

(2.14) $\quad \mathcal{O}^D\alpha \vee (\mathcal{O}^{(k-1)D}\theta \cap \rho_{k-1}) \subseteq \mathcal{O}^D\alpha \vee (\mathcal{O}^{kD}\theta \cap \rho_k)$ for $k = 1, \ldots, m$.

The claim depends on the equation and inclusions

(2.15)
$$\begin{aligned} \mathcal{O}^{(k-1)D}\theta \cap \rho_{k-1} &= (\mathcal{O}^{(k-1)D}\theta \cap \rho_{k-1}) \cap (\alpha \vee \beta_k) \\ &\subseteq [\mathcal{O}^D(\mathcal{O}^{(k-1)D}\theta \cap \rho_{k-1}) \cap \mathcal{O}^D\alpha] \vee \\ &\quad [\mathcal{O}^D(\mathcal{O}^{(k-1)D}\theta \cap \rho_{k-1}) \cap \mathcal{O}^D\beta_k] \\ &\subseteq \mathcal{O}^D\alpha \vee [\mathcal{O}^{kD}\theta \cap \rho_k]. \end{aligned}$$

Here the equality follows from $\mathcal{O}^{(k-1)D}\theta \cap \rho_{k-1} \subseteq \mathcal{O}^{mD}\theta \subseteq \alpha \vee \beta_k$. The first inclusion follows from (i) of Lemma 2.8. In the second inclusion, it is harmless to delete all but α in the bracketed expression on the left; for the bracketed expression on the right, Observation 2.2 is used to distribute $\mathcal{O}^D$ to the lowest-level constituents, even within ρ_{k-1}. To complete the proof of the claim 2.14, it suffices to take the join of $\mathcal{O}^D\alpha$ with the first and last expressions in 2.14.

Since $\rho_0 = 1$, by the claim 2.14 we have $\theta = \theta \cap 1 = \theta \cap \rho_0 \subseteq \mathcal{O}^D\alpha \vee (\theta \cap \rho_0) \subseteq \mathcal{O}^D\alpha \vee (\theta \cap \rho_1) \subseteq \cdots \subseteq \mathcal{O}^D\alpha \vee (\theta \cap \rho_m) \subseteq \mathcal{O}^D\alpha \vee \rho_m$, which gives the Lemma. $\square$

3. Varieties of bounded Maltsev depth

3.1 Definition. An algebra $\mathbf{A}$ has *Maltsev depth at most M* if $\mathcal{O}^{M+1} = \mathcal{O}^M$ on $\mathbf{Eqv}(A)$, in which case $\operatorname{Cg}\alpha = \mathcal{O}^M\alpha$ for each $\alpha \in \mathbf{Eqv}(A)$. A variety V can be said to have Maltsev depth at most M if $\mathcal{O}^{M+1} = \mathcal{O}^M$ on $\mathbf{Eqv}(A)$ for all $\mathbf{A} \in V$. An algebra or variety has *bounded Maltsev depth* if it has Maltsev depth at most M for some M.

3.2 Theorem (improving Ju Wang [19]). Let V be a congruence-distributive locally finite variety. If the finite subdirectly irreducible members of V have bounded Maltsev depth, then so does V itself. Specifically, if the Maltsev depths of finite subdirectly irreducible members are bounded by N, then all members of V have Maltsev depth bounded by $N + D$, where D is the maximum depth of the designated Jónsson terms for V.

The proof appears as 3.6 below. The explicit bound on the Maltsev depth of V is the improvement to [19].

3.3 Corollary (Ju Wang [18]). If V is a congruence-distributive variety generated by a finite algebra then there is a bound M such that for all $\mathbf{A} \in V$ and all $a, b \in A$, $\mathrm{Cg}(a, b) = \mathrm{Cg}_M(a, b)$. Here $\mathrm{Cg}(a, b)$ denotes the principal congruence relation obtained by identifying a and b and $\mathrm{Cg}_M(a, b)$ denotes the equivalence relation generated by $\{\langle p(a), p(b)\rangle : p \in O_A^M\}$ (M-fold compositions).

Proof of Corollary 3.3 from Theorem 3.2: For $\mathbf{A} \in V$ and $a, b \in A$, apply Observation 2.4 to $\delta(a, b)$, the atomic equivalence relation that identifies only a and b. □

3.4 Lemma. Suppose that $f : \mathbf{A} \to \mathbf{B}$ is a surjective homomorphism and that $\mathbf{B}$ has Maltsev depth at most M. Then for each $\alpha \in \mathbf{Eqv}(A)$ we have $\mathrm{Cg}\,\alpha \subseteq \mathcal{O}^M \alpha \vee \ker f$.

In other words, in computing the congruence closure of an equivalence relation we can bound powers of $\mathcal{O}$ in $\mathbf{Eqv}(A)$ at the cost of an adjustment by $\ker f$.

Proof. By various Observations, $f(\mathrm{Cg}\,\alpha) = f(\cup_{k=0}^{\infty} \mathcal{O}^k \alpha) = \cup_{k=0}^{\infty} f(\mathcal{O}^k \alpha) = \cup_{k=0}^{\infty} \mathcal{O}^k f(\alpha) = \mathcal{O}^M f(\alpha) = f(\mathcal{O}^M \alpha)$. Therefore $\mathrm{Cg}\,\alpha \subseteq f^{-1} f(\mathcal{O}^M \alpha) = \mathcal{O}^M \alpha \vee \ker f$. □

3.5 Lemma. Let V be a congruence-distributive variety. If $\mathbf{A} \in V$ is the subdirect product of finitely many factors each with Maltsev depth at most N, then $\mathbf{A}$ has Maltsev depth at most $N + D$, where D is the maximum depth of designated Jónsson terms for V.

Proof. Suppose $\mathbf{A}$ is a subdirect product of $\mathbf{B}_1, \ldots, \mathbf{B}_m$, and let $f_i : \mathbf{A} \to \mathbf{B}_i$ be the corresponding coordinate projections. For $\alpha \in \mathbf{Eqv}(A)$, Lemma 3.4 gives $\mathrm{Cg}\,\alpha \subseteq (\mathcal{O}^N \alpha \vee \ker f_1) \cap \cdots \cap (\mathcal{O}^N \alpha \vee \ker f_m)$. Then

$\mathrm{Cg}\,\alpha \subseteq \mathcal{O}^D\mathcal{O}^N\alpha \vee (\ker f_1 \cap \cdots \cap \ker f_m) = \mathcal{O}^{N+D}\alpha \vee 0 = \mathcal{O}^{N+D}\alpha$, by Lemma 2.13, taking into account the fact that $\mathrm{Cg}\,\alpha$ and the $\ker f_i$ are congruence relations and so are fixed points of $\mathcal{O}$ and $\mathcal{O}^{-1}$. □

3.6. *Proof of Theorem 3.2.* Let $\mathbf{A} \in V$ and $\alpha \in \mathbf{Eqv}(A)$ be given. First consider the case where $\alpha = \delta(a, b)$. Let $\langle r, s\rangle \in \mathrm{Cg}(a, b)$. Because Maltsev's construction (as in Observation 2.4) is finitary, we still have $\langle r, s\rangle \in \mathrm{Cg}(a, b)$ inside some finitely generated subalgebra $\mathbf{S}$ of $\mathbf{A}$. Since V is locally finite, S is finite and is therefore the subdirect product of finitely many factors. By Lemma 3.5 applied to $\mathbf{S}$, we have $\langle r, s\rangle \in \mathcal{O}^{M+D}\delta(a, b)$, as desired.

For general α, the same bound follows from Observation 2.1 and the fact that every element of $\mathbf{Eqv}(A)$ is the supremum of a set of atomic equivalence relations, i.e., relations of the form $\delta(a, b)$. □

3.7 Example. Consider varieties of lattices. Since only one nontrivial Jónsson term is needed, such varieties have $D = 2$.

1. In the variety $\mathcal{D}$ of distributive lattices, the only subdirectly irreducible member is the 2-element chain $\mathbf{2}$, which has Maltsev depth 0, so $\mathcal{D}$ has Maltsev depth at most $0 + 2 = 2$. This bound is actually achieved in the cube $\mathbf{2}^3$.
2. The five-element nonmodular lattice N_5 has Maltsev depth 2, so the variety generated by N_5 has Maltsev depth at most 4.
3. The five-element modular nondistributive lattice M_3 has Maltsev depth 2, so the variety generated by M_3 has Maltsev depth at most 4.

In the last two examples there are weak projectivities of length 3 that cannot be shortened, but by using transitivities Maltsev depth 2 can be achieved.

4. An algorithmic approach

We undertake a direct proof of Lemma 3.5, by means of a recursive construction. It is based on the observation that when a Maltsev scheme [10] is pulled back through a surjection, the weak projectivities pull back suitably but equalities needed for the connecting sequence may fail, resulting in a longer sequence that has "gaps". The following framework provides for the gaps.

In any algebra $\mathbf{A} \in \mathcal{V}$, for a finite sequence of elements $c = c_0, \ldots, c_m = d$, by an *even link* of the sequence (or an *odd link*) let us mean a pair $\langle c_i, c_{i+1}\rangle$, where i is even (or odd, respectively). For $a, b, \in A$

and an integer N, let us say that such a sequence has depth (at most) N relative to a, b if

(i) m is odd, so that the number of terms is even; and

(ii) for all odd links $\langle c_i, c_{i+1}\rangle$ of the sequence, $\langle a, b\rangle \to_N \langle c_i, c_{i+1}\rangle$.

Let us call an even link $\langle c_i, c_{i+1}\rangle$ a *gluing* if $c_i = c_{i+1}$ or a *gap* if not. Let us say that the sequence is *end-consistent* if
$\langle c, c_i\rangle \in \mathrm{Cg}(c, d)$ for all i.

Here are some examples, in all of which it is assumed that $a, b, c, d \in A$:

1. The *trivial sequence:* The sequence c, d itself has depth 0 relative to a, b, since there are no odd links.
2. A *doubled sequence:* If $c = c_0, c_1, \ldots, c_k = d$ is a Maltsev sequence connecting c to d, with $\{a, b\} \to_N \{c_{i-1}, c_i\}$ for each $i = 1, \ldots, k$, then the doubled sequence $c = c_0, c_0, c_1, c_1, \ldots, c_k, c_k = d$ has depth N relative to a, b.

   Conversely, observe that a sequence from c to d of depth N relative to a, b that has no gaps, only gluings, gives a Maltsev sequence of depth at most N witnessing $(c, d) \in \mathrm{Cg}(a, b)$.
3. An *image sequence:* If $f : \mathbf{A} \to \mathbf{B}$ is a homomorphism and $c = c_0, c_1, \ldots, c_m = d$ is a sequence in $\mathbf{A}$ of depth N relative to a, b, then $\bar{c} = \bar{c}_0, \bar{c}_1, \ldots, \bar{c}_m = \bar{d}$ is a sequence in $\mathbf{B}$ of depth N relative to $\bar{a}, \bar{b}$, where bars denote images.
4. A *pullback sequence:* If $f : \mathbf{A} \to \mathbf{B}$ is a surjection such that in $\mathbf{B}$ the images $\bar{c}, \bar{d}$ are connected by a sequence of depth N relative to $\bar{a}, \bar{b}$, then this sequence pulls back to a sequence in $\mathbf{A}$ of depth N relative to a, b. Indeed, the same terms can be used for the polynomials, with an arbitrary choice of pre-images of the auxiliary elements involved.
5. A *lifted sequence:* If $c = c_0, \ldots, c_m = d$ is a sequence of depth N relative to a, b, then the zig-zag sequence
   $c = c_0 = t_1(c, c_0, d),\ t_1(c, c_1, d),\ \ldots,\ t_1(c, c_m, d) = t_2(c, c_m, d),\ \ldots,$
   $c_m = d$
   has depth $N + D$ relative to a, b, where D is the maximum depth of the terms t_i.

   Observe that the zig-zag sequence has the virtue of being end-consistent, at the cost of an increase in depth. Observe also that if the original sequence has no gaps, neither does the lifted sequence.
6. A *patched sequence* If $c = c_0, \ldots, c_m = d$ is a sequence connecting c and d and for some even i we have another sequence $r_0 = c_i, r_1, \ldots, r_k = c_{i+1}$, where k is even, then we say that the combined sequence $c = c_0, \ldots, c_i, c_i = r_0, r_1, \ldots, r_k = c_{i+1}, c_{i+2}, \ldots, c_m$

has been obtained by "patching" the second sequence into the first at the link $\langle c_i, c_{i+1}\rangle$. We generally re-index the patched sequence.

Observe that if the original two sequences are end-consistent, so is the patched sequence. Observe also that if the original two sequences have depth at most N relative to a, b then so does the patched sequence.

Re-proof of Lemma 3.5: Given $\langle c, d\rangle \in \mathrm{Cg}(a, b)$, the plan is to start with the trivial sequence c, d and modify it repeatedly by patching gaps, always keeping the result end-consistent and of depth at most $N+D$ relative to a, b, until finally such a sequence is obtained with no gaps. Then we are done.

To describe the modification step, suppose that we currently have an end-consistent sequence $c = c_0, \ldots, c_m = d$ of depth at most $N+D$ relative to a, b. By a "gap split by π_j", where $\pi_j : \mathbf{A} \to \mathbf{B}_j$ is the coordinate projection, let us mean a gap $\langle c', d'\rangle = \langle c_i, c_{i+1}\rangle$ for which the images $\pi_j(c_i), \pi_j(c_{i+1})$ are distinct—certainly any gap has some such j. We patch this gap "via π_j" as follows. By end-consistency in $\mathbf{A}$, in $\mathbf{B}_j$ we have $\langle \pi_j(c'), \pi_j(d')\rangle \in \mathrm{Cg}(\pi_j(c), \pi_j(d)) \subseteq \mathrm{Cg}(\pi_j(a), \pi_j(b))$. By hypothesis, there is a Maltsev sequence in $\mathbf{B}_j$ connecting $\pi_j(c')$ and $\pi_j(d')$ with depth at most N relative to $\pi_j(a), \pi_j(b)$. Pull the double of this sequence back to $\mathbf{A}$ and lift, to obtain an end-consistent sequence in $\mathbf{A}$ connecting c' and d', of depth at most $N + D$ relative to a, b. Finally, patch this lifted sequence into the current sequence to obtain a new sequence. By construction, the segment of the new sequence between c' and d' has no gaps split by π_j, only gluings. Moreover, if later a new end-consistent sequence is patched in somewhere in that segment, the resulting patched sequence too will have no gaps split by π_j between c' and d', because by end-consistency all the even links will be in $\ker \pi_j$.

A convenient overall organization is to patch all gaps split by π_1, via π_1, then to patch all gaps split by π_2, via π_2, and so on. Because, as noted, further patching does not introduce more gaps, eventually all gaps will have been patched at all π_j. Since any gap must be split by some π_j, there are no gaps left and the algorithm terminates. $\square$

5. The finite basis theorem

5.1 Theorem [4]. A finite algebra of finite type that generates a congruence-distributive variety is finitely based.

The proof appears as 5.4 below.

5.2 Lemma. A variety V has bounded Maltsev depth M if and only if this property is finitely equationally expressible, in the sense that there is a finite set Σ of laws of V all of whose models have Maltsev depth at most M.

Proof. The "if" implication is trivial; let us consider "only if". Because we are constructing laws, this discussion will distinguish between three contexts: term algebras, free algebras in V, and arbitrary algebras in V. We notate elements of free algebras as images of terms. Thus for a term algebra $\mathbf{T}$ generated by variable symbols $x, y, \ldots$, the free algebra in V with corresponding generators will be denoted $\mathbf{F}_V(\bar{x}, \bar{y}, \ldots)$, where the bar denotes the natural epimorphism of $\mathbf{T}$ onto the free algebra. The proof of the lemma will consist of examining carefully how the relation $\mathcal{O}^{M+1} = \mathcal{O}^M$ in a suitable free algebra becomes equational in V.

By a *protolinear term*, let us mean a term that is a formal composition of operation symbols using variable symbols $x, z_1, \ldots, z_m$, each appearing once, where x occupies the "argument" entry and the auxiliary variable symbols $z_1, z_2, \ldots$ appear consecutively left to right up to some point and do not appear thereafter, and where every subterm is either a variable or includes x. For example, if the type consists of a single binary operation with symbol b and if $m \geq 2$, then one protolinear term is $\ell(x, z_1, \ldots, z_m) = b(z_1, b(x, z_2))$. In an algebra in V with designated elements $c_1, \ldots, c_m$ there is a corresponding unary polynomial $a \mapsto \ell(a, c_1, \ldots, c_m) = b(c_1, b(a, c_2))$. In fact, every linear unary polynomial in every algebra in V has this form, for a suitable m. Let us choose m large enough that protolinear terms in $x, z_1, \ldots, z_m$ are adequate to induce any linear unary polynomial of depth at most $M+1$ in any member of V; such a choice is $m = (M+1)(k-1)$, where let k be the maximum arity of operation symbols in the type of V. Let Λ_{M+1} be the set of protolinear terms ℓ in $x, z_1, \ldots, z_m$ of depth $M+1$. Since V is of finite type, Λ_{M+1} is finite.

Take any $\ell \in \Lambda_{M+1}$. In the free algebra $\mathbf{F} = \mathbf{F}_V(\bar{x}_0, \bar{x}_1, \bar{z}_1, \ldots, \bar{z}_m)$, observe that $\langle \bar{x}_0, \bar{x}_1 \rangle \to_{M+1} \langle \ell(\bar{x}_0, \bar{z}_1, \ldots, \bar{z}_m), \ell(\bar{x}_1, \bar{z}_1, \ldots, \bar{z}_m) \rangle$. Then by the choice of M, $\langle \ell(\bar{x}_0, \bar{z}_1, \ldots, \bar{z}_m), \ell(\bar{x}_1, \bar{z}_1, \ldots, \bar{z}_m) \rangle \in \mathcal{O}^M \delta(\bar{x}_0, \bar{x}_1)$, which by Observation 2.6 is the equivalence relation generated by all pairs $\langle p(a), p(b) \rangle$ for all linear p that are compositions of at most M operational unary polynomials on $\mathbf{F}$. Therefore in $\mathbf{F}$ there is a finite sequence connecting $\ell(\bar{x}_0, \bar{z}_1, \ldots, \bar{z}_m)$, $\ell(\bar{x}_1, \bar{z}_1, \ldots, \bar{z}_m)$, of depth at most M relative to $\bar{x}_0, \bar{x}_1$ and with no gaps. The even links, giving equations in a free algebra, constitute a set Σ_ℓ of laws of V. Let $\Sigma = \bigcup_{\ell \in \Lambda_{M+1}} \Sigma_\ell$. Then $V \models \Sigma$.

Further, if in some model $\mathbf{A}$ of Σ we have $\langle a_0, a_1 \rangle \to_{M+1} \langle e_0, e_1 \rangle$, then there exist $\ell \in \Lambda_{M+1}$ and constants $c_1, \ldots, c_m$ such that $e_j = \ell(a_j, c_1, \ldots, c_m)$ for $j = 0, 1$. The laws of Σ_ℓ then give a recipe for building a Maltsev scheme in $\mathbf{A}$ to show $\langle e_0, e_1\rangle \in \mathcal{O}^M\delta(a_0, a_1)$. Formally, we represent $\mathbf{A}$ as a homomorphic image of $\mathbf{F}$, pull the arrow back to $\mathbf{F}$, regard it as a congruence scheme, replace it by a congruence scheme of depth at most M using the laws of Σ_ℓ, and then map forward to $\mathbf{A}$. By the observation of 2.6, this argument proves that $\mathbf{A}$ has Maltsev depth at most M. $\square$

5.3 *Remark.* The construction just presented yields explicit laws, individually not complex but possibly numerous.

5.4 *Proof of Theorem 5.1.*

Let $\mathbf{A}$ be a finite algebra of finite type, generating a congruence-distributive variety V. By Theorem 3.2, V has bounded Maltsev depth M, so that Lemma 5.2 applies. Let us build an equational basis for V by including various finite sets of laws in turn to get smaller and smaller varieties, ending with V. First, let us take the finite set Ψ of laws of Jónsson [12] satisfied by the chosen Jónsson terms for V. Second, let us take the finite set Σ of laws constructed in Lemma 5.2; the variety defined by $\Psi \cup \Sigma$ includes V and has members of Maltsev depth at most M.

Third, by Jónsson [12], V has only finitely many subdirectly irreducible (SI) algebras, all finite; let K be their maximum cardinality. Let us take the set of laws $\Delta_{K,M}$ obtained by applying the construction of §4 of [2] for the case of the disjunction $(\forall x_0)\cdots(\forall x_K)(\mathrm{OR}_{i<j} x_i \approx x_j)$, to a maximum depth M. The set of laws $\Psi \cup \Sigma \cup \Delta_{K,M}$ defines a variety containing V of which all SI members have at most K elements.

Fourth, a finite set Γ of additional laws will suffice to exclude the finitely many SI models of $\Psi \cup \Sigma \cup \Delta_{K,M}$ that are not in V. The equational basis of V, then, is $\Psi \cup \Sigma \cup \Delta_{K,M} \cup \Gamma$. $\square$

6. Problems

1. Determine whether Lemma 3.5 can be extended to the case of infinitely many subdirect factors. This is unlikely to be the case, even for lattices, but a counterexample is elusive. One approach would be to look for a sequence of finite lattices, each with $\mathcal{O}^{M+1} = \mathcal{O}^M$ for the same bound M, but where the Maltsev schemes producing this reduction require longer and longer strings of transitivities.
2. Can the approximate distributive law 2.13 be simplified while still retaining a constant bound on the power of $\mathcal{O}$ applied to α? What

about the case of lattices? Can the power of $\mathcal{O}$ in 2.8-(ii) be reduced?

3. From each operation $*_i$ on pairs described in [2] we can define an operation $\alpha *_i \beta$ on equivalence relations, whose value is the equivalence closure of the obvious set of pairs. Incorporate these operations in the theory developed in §2. (Cf. [4, 17].)
4. The method of constructing a basis used in §5 is still not very economical in terms of the number of laws produced. Find a more economical approach—one that approaches known equational bases in small examples.
5. The method of [4] was actually carried further, to a finite basis theorem for varieties whose subdirectly irreducible members form an elementary class. This approach is distilled especially well in Jónsson [13]. Can this more general theory be tied to the methods of the present paper?

The authors are grateful to the referee and editor for valuable suggestions.

References

[1] Baker, K. A., *Equational axioms for classes of lattices*, Bull. Amer. Math. Soc. **77** (1971), 97- 102.
[2] Baker, K. A., *Primitive satisfaction and equational problems for lattices and other algebras*, Trans. Amer. Math. Soc. **190** (1974), 125-150.
[3] Baker, K. A. and Pixley, A. F., *Polynomial interpolation and the Chinese remainder theorem for algebraic systems*, Math. Z. **143** (1975), 165-174.
[4] Baker, K. A., *Finite equational bases for finite algebras in a congruence-distributive equational class*, Advances in Math. **24** (1977), 207-243.
[5] Baker, K. A., *Nondefinability of projectivity in lattice varieties*, Algebra Universalis **17** (1983), 267-274.
[6] Baker, K. A., *Bjarni Jónsson's contributions in algebra*, Algebra Universalis **31** (1994), 306-336.
[7] Baker, K. A., and Wang, J., *Definable principle subcongruences*, Algebra Universalis (to appear).
[8] Burris, S., *On Baker's finite basis theorem for congruence distributive varieties*, Proc. Amer. Math. Soc. **73** (1979), 141-148.
[9] Burris, S., and Sankappanavar, H. P., *A Course in Universal Algebra*, Graduate Texts in Mathematics, Springer-Verlag, New York, 1981.
[10] Fried, E., Grätzer, G., and Quackenbush, R., *Uniform congruence schemes*, Algebra Universalis **10** (1982) 176-188.
[11] Grätzer, G., *Universal Algebra*, 2nd. ed., Springer-Verlag, New York, 1979.
[12] Jónsson, B., *Algebras whose congruence lattices are distributive*, Math. Scand. **21** (1967), 110-121.
[13] Jónsson, B., *On finitely based varieties of algebras*, Colloq. Math. **42** (1979), 255-261.

[14] Makkai, M., *A proof of Baker's finite-base theorem on equational classes generated by finite elements of congruence distributive varieties*, Algebra Universalis **3** (1973), 174-181.

[15] McKenzie, R., *Finite equational bases for congruence modular varieties*, Algebra Universalis **24** (1987), 224-250.

[16] McKenzie, R., McNulty, G., and Taylor, W., *Algebras, Lattices, Varieties*, Wadsworth, Inc., Belmont, MA, 1987.

[17] Taylor, W., *Baker's finite basis theorem*, Algebra Universalis **8** (1978), 191-196.

[18] Wang, Ju, *A proof of the Baker conjecture*, Acta Math. Sinica **33** (1990), 626-633 (Chinese).

[19] Wang, Ju, *Uniform minimal pairs, uniform boundedness, and the finite basis theorem*, Adv. in Math. (China) **24** (1995), 320-326 (Chinese).

(K. A. Baker) University of California, Box 951555, Los Angeles, CA 90095-1555, USA

E-mail address: `baker@math.ucla.edu`

(J. Wang) Institute of Software, Academia Sinica, Beijing, 100080, China