POLYNOMIAL IDENTITIES FOR BERNSTEIN ALGEBRAS OF SIMPLE MENDELIAN INHERITANCE

MURRAY R. BREMNER*,†, YUNFENG PIAO AND SHELDON W. RICHARDS

Abstract. The Bernstein algebra $D^n G_\alpha$ of simple Mendelian inheritance is the n-fold duplicate of the gametic algebra G_α with basis $A_1, \dots, A_\alpha$ and structure constants $A_i A_j = \frac{1}{2} A_i + \frac{1}{2} A_j$. We simplify and generalize results of Bernad et al. and Peresi by showing that every identity for G_α follows from Costa's shape identity C, and that every identity for the zygotic algebra DG_α follows from the recombination identity R. We use computer algebra to determine the identities of degree ≤ 7 for the copular algebra $D^2 G_\alpha$.

1. Introduction

Hardy [9] and Weinberg [19] described the genotype distribution for Mendelian inheritance, and showed that stability is achieved in the second generation. The problem of finding all coefficients of heredity for which the distribution remains unchanged after the second generation was partially solved by Bernstein [3]. The nonassociative structures called genetic algebras were introduced by Etherington [7], and Bernstein's problem was formulated in algebraic terms by Lyubich [12]. The definition of Bernstein algebras was given by Holgate [10], and the generalization to n-th order algebras was made by Abraham [1]. Surveys of this theory can be found in Wörz-Busekros [20], Lyubich [13] and Reed [16]. Mathematical population genetics primarily uses methods of probability theory, stochastic processes, and differential equations; see for example, Ewens [8], Svirezhev and Passekov [18]. For connections between this approach and the methods of nonassociative algebra, see the remarks in [8, pages 208, 216, 250], [18, page 233] and [13, Preface].

In this paper we relate Bernstein algebras to the theory of DNA computing. The operation of intermolecular recombination was formalized by Landweber and Kari [11], and the recombination identity R of degree 4 was discovered by Bremner [5]; see also Sverchkov [17]. Bernad, González, Martínez and Iltyakov [2] found a complete set of polynomial identities for the gametic and zygotic algebras on two alleles; this set of generators was simplified by Peresi [15]. For the gametic algebras, we simplify and generalize [2, Theorem 2.1] by reducing the identities to R and another identity Q of degree 4, proving that R and Q hold for any number of alleles, and showing that R and Q are collectively equivalent to the single shape identity C discovered by Costa [6]. For the zygotic algebras, we simplify and generalize [2, Theorem 3.2] by reducing the identities to R and proving that R holds for any number of alleles. For the copular algebras, we use computer algebra to determine the identities of

2000 *Mathematics Subject Classification.* Primary 17D92; Secondary 17-04, 17A30, 92D10.
Key words and phrases. Bernstein algebras, polynomial identities, computer algebra.
* Corresponding author. Email: bremner@math.usask.ca.
† Partially supported by the Natural Sciences and Engineering Research Council of Canada.

degree ≤ 7: the first identities occur in degree 5; the values of R give absolute zero-divisors; there are no new identities in degrees 6 or 7; and the space of identities for two alleles properly contains the space of identities for three or more alleles.

2. Preliminaries

We recall basic results on nonassociative algebras and polynomial identities.

Definition 2.1. An *algebra* is a vector space $\mathcal{A}$ over the field $\mathbb{Q}$ of rational numbers with a bilinear product $\mathcal{A} \times \mathcal{A} \to \mathcal{A}$ denoted $(x, y) \mapsto xy$. We say $\mathcal{A}$ is *commutative* if $xy = yx$ for all $x, y \in \mathcal{A}$. A *baric algebra* is an algebra with a *weight function*: a nonzero linear map $\omega\colon \mathcal{A} \to \mathbb{Q}$ such that $\omega(xy) = \omega(x)\omega(y)$ for all $x, y \in \mathcal{A}$. If $\mathcal{A}$ is commutative then the *duplicate* $D\mathcal{A}$ is the algebra whose underlying vector space is the symmetric square $S^2(A)$ and whose product is $(x_1 \otimes x_2)(y_1 \otimes y_2) = x_1x_2 \otimes y_1y_2$ extended bilinearly. (We do not define $(x_1 \otimes x_2)(y_1 \otimes y_2) = x_1y_1 \otimes x_2y_2$.)

Lemma 2.2. [20, Theorem 6.15] *If $\mathcal{A}$ is commutative then $D\mathcal{A}$ is commutative. If $\mathcal{A}$ is a baric algebra with weight function ω, then $D\mathcal{A}$ is a baric algebra with weight function $\omega(x_1 \otimes x_2) = \omega(x_1)\omega(x_2)$ extended bilinearly.*

Definition 2.3. The *plenary powers* of $x \in \mathcal{A}$ are $x^{[0]} = x$ and $x^{[n+1]} = (x^{[n]})^2$ for $n \geq 0$. A *Bernstein algebra* of order n is a commutative baric algebra $\mathcal{A}$ in which $x^{[n+1]} = \omega(x)^{2^n} x^{[n]}$ for all $x \in \mathcal{A}$. The *gametic algebra* G_α on α alleles is the baric algebra with *natural basis* $A_1, A_2, \dots, A_\alpha$, product $A_iA_j = \frac{1}{2}A_i + \frac{1}{2}A_j$ and weight function $\omega\big(\sum_i x_iA_i\big) = \sum_i x_i$. The *canonical basis* is $C_1 = A_1$, $C_i = A_i - A_1$ ($i \geq 2$) with product $C_1^2 = C_1$, $C_1C_i = \frac{1}{2}C_i$ ($i \geq 2$), $C_iC_j = 0$ ($i, j \geq 2$) and weight function $\omega\big(\sum_i x_iC_i\big) = x_1$. The duplicate of G_α is the *zygotic algebra* DG_α; the duplicate of DG_α is the *copular algebra* D^2G_α.

Lemma 2.4. *G_α is a Bernstein algebra of order 0: $x^2 = \omega(x)x$ for all $x \in G_\alpha$.*

Lemma 2.5. [14, Theorem 2.1] *If $\mathcal{A}$ is a Bernstein algebra of order n then its duplicate $D\mathcal{A}$ is a Bernstein algebra of order $n{+}1$.*

Lemma 2.6. *For $n \geq 0$, D^nG_α is a Bernstein algebra of order n.*

Definition 2.7. A *nonassociative monomial* is a fully parenthesized string of variables; a *nonassociative polynomial* I is a linear combination of nonassociative monomials. We say I is *multilinear* of degree n if each variable $x_1, \dots, x_n$ occurs exactly once in every monomial. We say $I(x_1, \dots, x_n)$ is a *polynomial identity* for the algebra $\mathcal{A}$ if $I(x_1, \dots, x_n) = 0$ for all $x_1, \dots, x_n \in \mathcal{A}$.

Lemma 2.8. [21, Chapter 1] *Any polynomial identity over a field of characteristic 0 is equivalent to a finite set of multilinear identities.*

Example 2.9. Since D^nG_α is commutative, we restrict our attention to multilinear commutative nonassociative monomials. The difference between monomials and commutative monomials can be illustrated as follows. In degree 3 there are two association types $(**)*$ and $*(**)$; combining these with 6 permutations of 3 variables gives 12 nonassociative monomials. Assuming commutativity we need only the first type and 3 permutations: $(ab)c$, $(ac)b$, $(bc)a$. In degree 4 there are five association types $((**)*)*$, $(*(**))*$, $(**)(**)$, $*((**)*)$ and $*(*(**))$; combining these with 24 permutations of 4 variables gives 120 nonassociative monomials. Assuming commutativity we need only the first and third types, with 12 permutations for the first, $((ab)c)d$, $((ab)d)c$, $((ac)b)d$, $((ac)d)b$, $((ad)b)c$, $((ad)c)b$, $((bc)a)d$, $((bc)d)a$, $((bd)a)c$, $((bd)c)a$, $((cd)a)b$, $((cd)b)a$, and 3 permutations for the third, $(ab)(cd)$, $(ac)(bd)$, $(ad)(bc)$.

Definition 2.10. We inductively define a *total order on commutative association types.* Let t be a commutative association type in degree d, and let x be the product $a_1 \cdots a_d$ with association type t. By commutativity we may assume that $x = x'x''$ uniquely where x', x'' have association types t', t'' respectively, and that either (i) $\deg x' > \deg x''$ or (ii) $\deg x' = \deg x''$ and t' precedes t'' in the total order already defined on association types of lower degree. If t_i $(i = 1, 2)$ are commutative association types in degree d with factorizations $x_i = x_i'x_i''$ where x_i' and x_i'' have association types t_i' and t_i'' respectively, then t_1 precedes t_2 if and only if either (i) t_1' precedes t_2' or (ii) $t_1' = t_2'$ and t_1'' precedes t_2''.

Example 2.11. Listed in the above order, the commutative association types up to degree 5 are $*$, $**$, $(**)*$, $((**)*)*$, $(**)(**)$, $(((**)*)*)*$, $((**)(**))*$, $((**)*)(**)$.

Definition 2.12. Any commutative association type t in degree d has a number s of *symmetries* arising from commutativity, giving $d!/2^s$ *equivalence classes* of monomials with type t. As *standard representative* of each class we choose the monomial whose underlying permutation comes first in lexicographical order. We impose a *total order on the equivalence classes* by using the lexicographical order on the standard representatives.

Example 2.13. In degree 3 there is one symmetry $(ab)c = (ba)c$ giving the 3 ordered monomials of Example 2.9. In degree 4, type 1 has one symmetry $((ab)c)d = ((ba)c)d$, and type 2 has three symmetries $(ab)(cd) = (ba)(cd) = (ab)(dc) = (cd)(ab)$, giving the $12 + 3 = 15$ ordered monomials of Example 2.9.

Lemma 2.14. [4, Proposition 1] *In degree d, there are $(2d-3)!!$ multilinear commutative nonassociative monomials where $k!! = (k)(k-2)\cdots(3)(1)$, k odd.*

Definition 2.15. We write P_d for the vector space of multilinear commutative nonassociative polynomials of degree d: all linear combinations over $\mathbb{Q}$ of the multilinear commutative nonassociative monomials of degree d. The symmetric group S_d acts on P_d: omitting parentheses, we set $\sigma(a_{i_1} \cdots a_{i_d}) = a_{\sigma(i_1)} \cdots a_{\sigma(i_d)}$; we then straighten the monomial (replace it by the standard representative of its equivalence class). This action of S_d does not change the association types and gives P_d the structure of an S_d-module.

Lemma 2.16. *The subspace $I(\mathcal{A}) \subseteq P_d$ consisting of the polynomial identities satisfied by the algebra $\mathcal{A}$ is an S_d-submodule of P_d.*

Proof. If I is an identity for $\mathcal{A}$ then so is σI, and any linear combination of identities for $\mathcal{A}$ is again an identity for $\mathcal{A}$. □

Definition 2.17. Let $I(a_1, \ldots, a_d)$ be an element of P_d and let a_{d+1} be a new variable. We apply multiplication, and substitution for $i = 1, \ldots, d$, to get $d{+}1$ elements of P_{d+1}: $I(a_1, \ldots, a_d)a_{d+1}$, and $I(a_1, \ldots, a_i a_{d+1}, \ldots, a_d)$ for $1 \le i \le d$. These $d{+}1$ *liftings* of I generate the submodule of P_{d+1} consisting of the identities in degree $d{+}1$ which are consequences of the identity I in degree d.

3. Gametic Algebras

In this section we study polynomial identities satisfied by the gametic algebras G_α.

Definition 3.1. We define polynomials $I, J, Q, R \in P_4$ and $K \in P_5$:

$$\begin{aligned}
I &= 2((ab)c)d - 2((ac)b)d - ((ad)b)c + ((ad)c)b\,,\\
J &= ((ab)c)d + ((ab)d)c - ((ac)b)d - ((ad)b)c + ((cd)a)b - (ab)(cd)\,,\\
Q &= 3((ab)c)d - ((ab)d)c - 3((ac)b)d + ((ac)d)b\,,\\
R &= 2((ab)c)d - ((ab)d)c - ((ac)b)d - ((bc)a)d + (ab)(cd)\,,\\
K &= 2(((ab)c)d)e - 2(((ac)b)d)e - ((ab)c)(de) + ((ac)b)(de)\,.
\end{aligned}$$

I, J, K are equivalent to the identities of [2, Theorem 2.1], and R is the identity satisfied by intermolecular recombination [5, Theorem 3].

Theorem 3.2. [2, Theorem 2.1] *Every identity satisfied by G_2, the gametic algebra on two alleles, follows from commutativity, I, J and K.*

Lemma 3.3. *I and J follow from Q and R, and conversely.*

Proof. We have

$$\begin{aligned}
4I &= 3Q(a,b,c,d) + Q(a,b,d,c) - Q(a,c,d,b)\,,\\
8J &= 6Q(a,b,c,d) + 3Q(a,b,d,c) + Q(a,c,d,b) + 3Q(b,a,c,d) - 3Q(c,a,d,b)\\
&\quad - Q(c,b,d,a) - 8R(a,b,c,d)\,,\\
Q &= 2I(a,b,c,d) - I(a,b,d,c) + I(a,c,d,b)\,,\\
2R &= 2I(a,b,c,d) + I(a,c,d,b) + 2I(b,a,c,d) - I(b,a,d,c) + I(b,c,d,a)\\
&\quad - I(c,a,d,b) - 2J(a,b,c,d)\,.
\end{aligned}$$

Thus I and J are in the submodule generated by Q and R, and conversely. □

Lemma 3.4. *K follows from Q.*

Proof. We have

$$\begin{aligned}
48K = {} & 36\,Q(a,b,c,d)e + 12\,Q(a,b,d,c)e + 21\,Q(a,b,e,c)d - 15\,Q(a,b,e,d)c\\
& - 12\,Q(a,c,d,b)e - 21\,Q(a,c,e,b)d + 15\,Q(a,c,e,d)b + 7\,Q(a,d,e,b)c\\
& - 7\,Q(a,d,e,c)b - 21\,Q(b,a,e,c)d + 15\,Q(b,a,e,d)c + 17\,Q(b,c,e,a)d\\
& - 8\,Q(b,c,e,d)a - 3\,Q(b,d,e,a)c + 21\,Q(c,a,e,b)d - 15\,Q(c,a,e,d)b\\
& - 17\,Q(c,b,e,a)d + 8\,Q(c,b,e,d)a + 3\,Q(c,d,e,a)b + 9\,Q(d,a,e,b)c\\
& - 9\,Q(d,a,e,c)b + 3\,Q(d,b,e,a)c - 3\,Q(d,c,e,a)b + 25\,Q(ae,b,c,d)\\
& + 20\,Q(ad,b,c,e) - 9\,Q(ae,b,d,c) + 12\,Q(ad,b,e,c) + 9\,Q(ae,c,d,b)\\
& - 12\,Q(ad,c,e,b) - 24\,Q(be,a,c,d) + 8\,Q(be,a,d,c) + 24\,Q(ce,a,b,d)\\
& - 8\,Q(ce,a,d,b) + 6\,Q(a,de,b,c) - 6\,Q(a,de,c,b) - 18\,Q(a,b,c,de)\,.
\end{aligned}$$

This equation can be verified with a computer algebra system. There are 36 terms on the right side; each expands to a linear combination of 4 monomials. We obtain 144 monomials, and must replace 24 by the standard representative of the equivalence class; only 52 distinct monomials remain. Let C be the 52×36 matrix in which C_{ij} is the coefficient of monomial i in the expansion of term j. Taking the row sums of C gives the coefficients in the expansion of the left side of the equation. $\square$

Example 3.5. This is how we obtained the equation in the proof of Lemma 3.4. There are 5 liftings of Q, namely

$$Q(a,b,c,d)e,\ Q(ae,b,c,d),\ Q(a,be,c,d),\ Q(a,b,ce,d),\ Q(a,b,c,de),$$

and five liftings of R. Let L be a (105+120) $\times$ 105 matrix, initially zero. For each lifting, we apply all permutations of the variables, store the results in the lower 120×105 block of L, and compute the row canonical form. The final rank is 100: the dimension of the submodule $U \subset P_5$ consisting of the identities implied by Q and R. Only five liftings (generators of U) increase the rank (see Algorithm 5.5 below):

$$Q(a,b,c,d)e,\ Q(ae,b,c,d),\ Q(a,be,c,d),\ Q(a,b,c,de),\ R(a,b,c,d)e.$$

Let M be a 105×600 matrix with five 105×120 blocks. In column j of block k we put permutation j of generator k. In the row canonical form, the columns which contain the leading 1 of a row give a basis of U. Let N be a 105×101 matrix with this basis of U in columns 1–100 and K in column 101. From the row canonical form we get the coefficients of K as a linear combination of permutations of liftings of Q and R. (Note that R does not appear in the final result.)

Theorem 3.6. *For $\alpha \geq 2$, every identity for the gametic algebra G_α follows from commutativity, Q and R.*

Proof. Theorem 3.2 and Lemmas 3.3 and 3.4 imply the result for $\alpha = 2$. Since $G_2 \subseteq G_\alpha$ for all α, every identity for G_α is an identity for G_2: thus in each degree the identities for G_α are a subspace of the identities for G_2. It remains to show that G_α satisfies Q and R. Since Q and R are multilinear, it suffices to consider basis elements. Evaluating the association types in degree 4 we get

$$\begin{aligned}\big((A_iA_j)A_k\big)A_\ell &= \tfrac{1}{8}A_i + \tfrac{1}{8}A_j + \tfrac{1}{4}A_k + \tfrac{1}{2}A_\ell\\ \big(A_iA_j\big)\big(A_kA_\ell\big) &= \tfrac{1}{4}A_i + \tfrac{1}{4}A_j + \tfrac{1}{4}A_k + \tfrac{1}{4}A_\ell\,.\end{aligned}$$

Using these we obtain

$$\begin{aligned}Q\big(A_i,A_j,A_k,A_\ell\big) &= 3\big(\tfrac{1}{8}A_i+\tfrac{1}{8}A_j+\tfrac{1}{4}A_k+\tfrac{1}{2}A_\ell\big) - \big(\tfrac{1}{8}A_i+\tfrac{1}{8}A_j+\tfrac{1}{2}A_k+\tfrac{1}{4}A_\ell\big)\\ &\quad - 3\big(\tfrac{1}{8}A_i+\tfrac{1}{4}A_j+\tfrac{1}{8}A_k+\tfrac{1}{2}A_\ell\big) + \big(\tfrac{1}{8}A_i+\tfrac{1}{2}A_j+\tfrac{1}{8}A_k+\tfrac{1}{4}A_\ell\big)\,,\\ R\big(A_i,A_j,A_k,A_\ell\big) &= 2\big(\tfrac{1}{8}A_i+\tfrac{1}{8}A_j+\tfrac{1}{4}A_k+\tfrac{1}{2}A_\ell\big) - \big(\tfrac{1}{8}A_i+\tfrac{1}{8}A_j+\tfrac{1}{2}A_k+\tfrac{1}{4}A_\ell\big)\\ &\quad - \big(\tfrac{1}{8}A_i+\tfrac{1}{4}A_j+\tfrac{1}{8}A_k+\tfrac{1}{2}A_\ell\big) - \big(\tfrac{1}{4}A_i+\tfrac{1}{8}A_j+\tfrac{1}{8}A_k+\tfrac{1}{2}A_\ell\big)\\ &\quad + \big(\tfrac{1}{4}A_i+\tfrac{1}{4}A_j+\tfrac{1}{4}A_k+\tfrac{1}{4}A_\ell\big)\,.\end{aligned}$$

In both expressions all terms cancel. $\square$

Costa [6, Theorem 2, page 127] has shown that the genetic algebra G_α satisfies the following identity:

$$C' = -2((ab)c)d + (a(bc))d + 2(ab)(cd) + a((bc)d) - 2a(b(cd)).$$

Definition 3.7. Since G_α is commutative, we can write $C = -C'$ in the form

$$C = 2((ab)c)d - ((bc)a)d - ((bc)d)a + 2((cd)b)a - 2(ab)(cd).$$

The referee made the following conjecture.

$$
M_1 = \begin{bmatrix}
2 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 2 & -2 & 0 & 0 \\
0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 2 & -2 & 0 & 0 \\
0 & 0 & 2 & 0 & 0 & 0 & -1 & -1 & 0 & 2 & 0 & 0 & 0 & -2 & 0 \\
0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 2 & -1 & -1 & 0 & -2 & 0 \\
0 & 0 & 0 & 0 & 2 & 0 & 0 & 2 & -1 & -1 & 0 & 0 & 0 & 0 & -2 \\
0 & 0 & 0 & 0 & 0 & 2 & 0 & 2 & 0 & 0 & -1 & -1 & 0 & 0 & -2 \\
2 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & -2 & 0 & 0 \\
0 & 2 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 2 & 0 & -2 & 0 & 0 \\
0 & 0 & -1 & -1 & 0 & 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 \\
0 & 0 & 0 & 0 & 0 & 2 & 0 & 2 & 0 & 0 & -1 & -1 & 0 & 0 & -2 \\
0 & 0 & 0 & 2 & -1 & -1 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & -2 & 0 \\
0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 2 & -1 & -1 & 0 & -2 & 0 \\
-1 & -1 & 2 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & -2 & 0 \\
0 & 0 & 0 & 2 & -1 & -1 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & -2 & 0 \\
-1 & -1 & 0 & 0 & 2 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 \\
0 & 0 & 0 & 0 & 2 & 0 & 0 & 2 & -1 & -1 & 0 & 0 & 0 & 0 & -2 \\
0 & 2 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 2 & 0 & -2 & 0 & 0 \\
0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 2 & -2 & 0 & 0 \\
-1 & -1 & 0 & 0 & 2 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 \\
0 & 0 & -1 & -1 & 0 & 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 \\
-1 & -1 & 2 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & -2 & 0 \\
0 & 0 & 2 & 0 & 0 & 0 & -1 & -1 & 0 & 2 & 0 & 0 & 0 & -2 & 0 \\
2 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & -2 & 0 & 0 \\
2 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 2 & -2 & 0 & 0
\end{bmatrix}
$$

TABLE 1. Identity C generates the row space of the matrix M_1

$$
M_2 = \left[\begin{array}{rrrrrrrrrrrrrrr}
2 & -1 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
-1 & 2 & 0 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 \\
-1 & 0 & 2 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & -1 & 2 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 \\
0 & -1 & 0 & 0 & 2 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & -1 & -1 & 2 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 \\
2 & -1 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
-1 & 2 & 0 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 \\
-1 & 0 & -1 & 0 & 0 & 0 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & 0 & -1 & 0 & -1 & 0 & 0 & 1 \\
0 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 2 & -1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 2 & 0 & -1 & 0 & 1 & 0 \\
-1 & 0 & 2 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & -1 & 2 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 \\
-1 & 0 & -1 & 0 & 0 & 0 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & 0 & -1 & 0 & -1 & 0 & 0 & 1 \\
0 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 2 & -1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & -1 & 2 & 1 & 0 & 0 \\
0 & -1 & 0 & 0 & 2 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & -1 & -1 & 2 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 \\
0 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 2 & -1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 2 & 0 & -1 & 0 & 1 & 0 \\
0 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 2 & -1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & -1 & 2 & 1 & 0 & 0 \\
\hline
3 & -1 & -3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 3 & 0 & 0 & -3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-3 & 1 & 3 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & 3 & 1 & -3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & -3 & 0 & 0 & 3 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & -3 & -1 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
3 & -1 & 0 & 0 & 0 & 0 & -3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & -3 & 1 & 0 & 0 & 0 & 0 & 0 \\
-3 & 1 & 0 & 0 & 0 & 0 & 3 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 3 & 1 & -3 & 0 & 0 & 0 & 0 & 0 \\
1 & -3 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & -3 & -1 & 3 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 3 & -1 & 0 & 0 & -3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & -3 & 1 & 0 & 0 & 0 \\
0 & 0 & -3 & 1 & 0 & 0 & 3 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 3 & 0 & 0 & 1 & -3 & 0 & 0 & 0 \\
0 & 0 & 1 & -3 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & -3 & 0 & 0 & -1 & 3 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 3 & -1 & 0 & 0 & -3 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & 3 & 0 & 0 & 0 & 0 & -3 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -3 & 1 & 0 & 0 & 3 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 3 & 1 & -3 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & -3 & 0 & 0 & 0 & 0 & 3 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -3 & -1 & 3 & 0 & 0 & 0
\end{array}\right]
$$

TABLE 2. Identities R and Q generate the row space of the matrix M_2

$$
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 3/2 & -3/2 & 0 & 0 & -3 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 & -1/2 & 5/2 & 0 & 0 & -1 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 3/2 & -5/2 & 0 & 0 & -3 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & -1/2 & -1/2 & 0 & 0 & -1 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -3 & -3/2 & 7/2 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & -3/2 & 3/2 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 3 & 2 & -3 & 0 & 0 & -3 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & -2 & 0 & 0 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -3 & -1 & 3 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1
\end{bmatrix}
$$

TABLE 3. Common row canonical form of the matrices M_1 and M_2

Lemma 3.8. *Assuming commutativity, R and Q are collectively equivalent to C.*

Proof. Let M_1 be the 24×15 matrix in which the columns are labeled by the ordered list of 15 monomials from the end of Example 2.9, and the rows are labeled by the 24 permutations of a, b, c, d in lexicographical order. In row i of M_1 we put the coefficient vector of the identity obtained by applying permutation i to the terms of C (and straightening the monomials). The matrix M_1 is displayed in Table 1.

Let M_2 be the 48×15 matrix consisting of two 24×15 blocks; the columns are labeled by the ordered list of 15 monomials from the end of Example 2.9, and in each block the rows are labeled by the 24 permutations of a, b, c, d in lexicographical order. In row i of the upper block we put the coefficient vector of the identity obtained by applying permutation i to the terms of R; in row i of the lower block we put the coefficient vector of the identity obtained by applying permutation i to the terms of Q. The matrix M_2 is displayed in Table 2.

It suffices to prove that the matrices M_1 and M_2 have the same row space. We compute the row canonical forms; both equal the matrix displayed in Table 3. □

We can now simplify Theorem 3.6 as follows.

Theorem 3.9. *For $\alpha \geq 2$, every identity for the gametic algebra G_α follows from commutativity and Costa's identity C.*

4. Zygotic Algebras

The natural basis of DG_α is $A_{ij} = A_i \otimes A_j$ $(1 \leq i \leq j \leq \alpha)$ with product

$$A_{ij} A_{k\ell} = \tfrac{1}{4} A_{ik} + \tfrac{1}{4} A_{i\ell} + \tfrac{1}{4} A_{jk} + \tfrac{1}{4} A_{j\ell} \,.$$

Definition 4.1. We define polynomials L, M, N (degree 4) and O, P (degree 5):

$$
\begin{aligned}
L &= ((a^2)b)a - a^2(ab)\,, \\
M &= -2((ab)b)a + (b^2a)a + a^2b^2\,, \\
N &= ((ab)c)d - ((ab)d)c - ((bc)a)d + ((bc)d)a + ((bd)a)c - ((bd)c)a\,, \\
O &= -2(((ab)c)c)d + 2(((bc)a)c)d - ((c^2b)a)d + ((ab)c^2)d\,, \\
P &= (((ab)c)d)e - (((ad)c)b)e - (((bc)a)d)e + (((cd)a)b)e\,.
\end{aligned}
$$

L is the Jordan identity. We linearize L and M (but not O) as follows:

$$L(a,b,c,d) = ((ac)b)d + ((ad)b)c + ((cd)b)a - (ab)(cd) - (ac)(bd) - (ad)(bc)\,,$$

$$M(a,b,c,d) = ((ab)d)c + ((ad)b)c + ((bc)d)a - ((bd)a)c - ((bd)c)a + ((cd)b)a - 2(ac)(bd)\,.$$

Theorem 4.2. [2, Theorem 3.2] *Every identity satisfied by DG_2, the zygotic algebra on two alleles, follows from commutativity, L, M, N, O and P.*

Remark 4.3. Peresi [15] has shown that every identity satisfied by DG_2 follows from commutativity, M and N.

Lemma 4.4. *L, M and N follow from R, and conversely.*

Proof. We have

$$\begin{aligned}
L &= -R(a,a,a,b)\,,\\
M &= R(a,a,b,b) + R(a,b,a,b) - R(a,b,b,a)\,,\\
4N &= 4R(a,b,c,d) + 4R(a,b,d,c) - 4R(a,c,d,b) + 3R(d,a,b,c) - 3R(d,a,c,b)\\
&\quad + 5R(d,b,a,c) - R(d,b,c,a) - 5R(d,c,a,b) - 3R(d,c,b,a)\,.
\end{aligned}$$

Conversely,

$$\begin{aligned}
2R(a,b,c,d) &= -2L(a,b,c,d) + L(a,c,b,d) - L(a,d,b,c) + M(a,b,c,d)\\
&\quad + M(a,b,d,c) + 2N(a,b,c,d) - N(a,c,b,d) + N(a,d,b,c)\,.
\end{aligned}$$

This completes the proof. □

Lemma 4.5. *O and P follow from R.*

Proof. We have

$$\begin{aligned}
O &= 3R(a,b,d,c)c - R(a,c,b,c)d - 4R(a,c,d,b)c + R(a,c,d,c)b\\
&\quad + 2R(a,d,b,c)c - 2R(a,d,c,b)c + R(b,c,a,c)d + R(b,c,c,a)d\\
&\quad + 3R(b,d,a,c)c + R(b,d,c,a)c - 2R(c,d,a,b)c - R(c,d,a,c)b\\
&\quad - R(c,d,b,a)c - R(ab,c,c,d) + R(ab,c,d,c) + R(ac,b,c,d)\\
&\quad - R(ac,b,d,c) + R(ac,d,b,c) - R(ac,d,c,b) - R(cd,a,b,c)\\
&\quad + R(cd,a,c,b)\,,\\
4P &= 4R(a,b,c,d)e + 4R(a,b,d,c)e - 4R(a,c,d,b)e + 3R(a,d,b,c)e\\
&\quad - 5R(a,d,c,b)e + 6R(a,d,e,b)c + 2R(a,e,b,d)c + 6R(a,e,d,b)c\\
&\quad + R(b,d,a,c)e + 3R(b,d,c,a)e - 8R(b,d,e,a)c + 8R(b,d,e,c)a\\
&\quad - 2R(b,e,a,d)c + 2R(b,e,c,d)a - 4R(b,e,d,a)c + 4R(b,e,d,c)a\\
&\quad - 3R(c,d,a,b)e - 3R(c,d,b,a)e - 6R(c,d,e,b)a - 2R(c,e,b,d)a\\
&\quad - 6R(c,e,d,b)a + 4R(d,e,a,b)c - 4R(d,e,b,a)c + 4R(d,e,b,c)a\\
&\quad - 4R(d,e,c,b)a - 2R(ad,b,c,e) + 2R(ad,b,e,c) + 2R(bd,a,c,e)\\
&\quad - 2R(bd,a,e,c) - 2R(bd,c,a,e) + 2R(bd,c,e,a) + 2R(bd,e,a,c)\\
&\quad - 2R(bd,e,c,a) - 2R(be,d,a,c) + 2R(be,d,c,a) - 2R(cd,b,e,a)\\
&\quad + 2R(cd,b,a,e)\,.
\end{aligned}$$

This completes the proof. □

Theorem 4.6. *For $\alpha \geq 2$, every identity for the zygotic algebra DG_α follows from commutativity and R.*

Proof. Theorem 4.2 and Lemmas 4.4 and 4.5 imply the result for $\alpha = 2$. Since $DG_2 \subseteq DG_\alpha$ for all α, every identity for DG_α is an identity for DG_2: thus in each degree the identities for DG_α are a subspace of the identities for DG_2. It remains to show that DG_α satisfies R. We ignore the assumption that $i \leq j$ and identify A_{ij} with A_{ji}. We first compute the general product in degree 3:

$$\big(A_{ij}A_{k\ell}\big)A_{mn} = \tfrac{1}{8}\big(A_{im}+A_{in}+A_{jm}+A_{jn}+A_{km}+A_{kn}+A_{\ell m}+A_{\ell n}\big)\,.$$

For the first association type in degree 4 we have

$$\big(\big(A_{ij}A_{k\ell}\big)A_{mn}\big)A_{pq} = \tfrac{1}{16}\big(A_{ip}+A_{iq}+A_{jp}+A_{jq}\big) + \tfrac{1}{16}\big(A_{kp}+A_{kq}+A_{\ell p}+A_{\ell q}\big) + \tfrac{1}{8}\big(A_{mp}+A_{mq}+A_{np}+A_{nq}\big)\,,$$

and therefore

$$2\big(\big(A_{ij}A_{k\ell}\big)A_{mn}\big)A_{pq} - \big(\big(A_{ij}A_{k\ell}\big)A_{pq}\big)A_{mn} - \big(\big(A_{ij}A_{mn}\big)A_{k\ell}\big)A_{pq} - \big(\big(A_{k\ell}A_{mn}\big)A_{ij}\big)A_{pq} = -\tfrac{1}{16}\big(A_{im}+A_{in}+A_{jm}+A_{jn}+A_{ip}+A_{iq}+A_{jp}+A_{jq} + A_{km}+A_{kn}+A_{\ell m}+A_{\ell n}+A_{kp}+A_{kq}+A_{\ell p}+A_{\ell q}\big)\,.$$

For the second association type in degree 4 we have

$$\big(A_{ij}A_{k\ell}\big)\big(A_{mn}A_{pq}\big) = \tfrac{1}{16}\big(A_{im}+A_{in}+A_{ip}+A_{iq}+A_{jm}+A_{jn}+A_{jp}+A_{jq} + A_{km}+A_{kn}+A_{kp}+A_{kq}+A_{\ell m}+A_{\ell n}+A_{\ell p}+A_{\ell q}\big)\,.$$

Adding the last two results gives $R\big(A_{ij}, A_{k\ell}, A_{mn}, A_{pq}\big) = 0$. □

Remark 4.7. Costa [6, Corollary 1, page 130] has shown that the zygotic algebras DG_α satisfy the identity $C(a,b,c,d)e$ in degree 5, where C is the identity of Definition 3.7.

5. Copular Algebras

The natural basis of D^2G_α consists of $A_{ijk\ell} = A_{ij} \otimes A_{k\ell}$ with $1 \leq i \leq j \leq \alpha$, $1 \leq k \leq \ell \leq \alpha$, and either $i < k$ or $i = k$, $j < \ell$. If we identify $A_{ijk\ell}$, $A_{ij\ell k}$, $A_{jik\ell}$, $A_{ji\ell k}$, $A_{k\ell ij}$, $A_{\ell kij}$, $A_{k\ell ji}$, $A_{\ell kji}$ then the product is

$$A_{ijk\ell}A_{mnop} = \tfrac{1}{16}\big(A_{ikmo}+A_{ikmp}+A_{ikno}+A_{iknp}+A_{i\ell mo}+A_{i\ell mp}+A_{i\ell no}+A_{i\ell np} + A_{jkmo}+A_{jkmp}+A_{jkno}+A_{jknp}+A_{j\ell mo}+A_{j\ell mp}+A_{j\ell no}+A_{j\ell np}\big)\,.$$

Little is known about polynomial identities of D^2G_α: even for $\alpha = 2$, a generating set for the identities has not yet been determined.

In this section our proofs are primarily computational. We use arithmetic modulo $p = 101$ so that each matrix entry fits in one byte. Modular arithmetic is essential in higher degrees when the matrices are large; it saves memory, but requires reconstruction of rational coefficients from modular coefficients.

Lemma 5.1. *For $\alpha \geq 2$, every identity of degree ≤ 4 for the copular algebras D^2G_α follows from commutativity. In particular, D^2G_α does not satisfy Costa's identity C or the recombination identity R.*

Proof. Since an identity in degree d implies identities in degree $d+1$, it suffices to prove that D^2G_α has no identities in degree 4. Since $D^2G_2 \subseteq D^2G_\alpha$, the identities for D^2G_α are a subspace of the identities for D^2G_2, and so it suffices to show that D^2G_2 has no identities in degree 4. We use modular arithmetic, since non-existence over $\mathbb{F}_p$ implies non-existence over $\mathbb{Q}$. (If I is a nontrivial identity over $\mathbb{Q}$, we multiply by the LCM of the denominators of the coefficients, divide by the GCD of the coefficients, and reduce mod p to get a nontrivial identity over $\mathbb{F}_p$.) We abbreviate the basis of D^2G_2 by

$$U = A_{1111},\ V = A_{1112},\ W = A_{1122},\ X = A_{1212},\ Y = A_{1222},\ Z = A_{2222}.$$

For $p = 101$ we have $U^2 = U$, $UV = VU = 51V$, $UX = XU = 76W$, $V^2 = 76X$, $VX = XV = 38Y$, $X^2 = 19Z$, and all other products are 0. There are 15 monomials M_i in degree 4 (see Example 2.9), so we need 15 linearly independent constraints on their coefficients x_i. Setting $abcd = UUVV$ in each M_i gives a multiple of W or X; for example, $((ab)c)d = ((UU)V)V = 38X$, $(ab)(cd) = (UU)(VV) = 19W$. We obtain these constraints:

$$60x_4 + 60x_6 + 60x_8 + 60x_{10} + 19x_{13} = 0\,,$$
$$38x_1 + 38x_2 + 19x_3 + 19x_5 + 19x_7 + 19x_9 + 19x_{14} + 19x_{15} = 0\,.$$

Setting $abcd = UVUV$, $UVVU$, $VUUV$, $VUVU$, $VVUU$, we get another 10 constraints. Setting $abcd = UUVX$, $UVUX$, $UVXU$ in each M_i gives a multiple of Y, and we get another 3 constraints. The coefficient matrix of the corresponding linear system has full rank. □

We now define two large matrices and two algorithms for processing them.

Definition 5.2. The *expansion matrix* in degree d for the commutative nonassociative algebra $\mathcal{A}$ of dimension r is the matrix $\mathcal{E}_d$ with $(2d{-}3)!! + r$ rows and $(2d{-}3)!!$ columns; the upper block has $(2d{-}3)!!$ rows and the lower block has r rows. The columns correspond to the ordered multilinear monomials.

The expansion matrix contains in its nullspace the polynomial identities for an algebra; for this matrix we use a fill and reduce process which terminates when the rank has stabilized.

Algorithm 5.3 (Fill and Reduce)**.** *Input:* Positive integers d (degree of identities), p (upper limit of pseudorandom integers), s (number of iterations required with stable rank) and a procedure to evaluate the product in the algebra $\mathcal{A}$ of dimension r. *Output:* Once the rank of the expansion matrix has stabilized, its nullspace contains polynomial identities of degree d satisfied by the algebra $\mathcal{A}$.

(1) Initialize the expansion matrix $\mathcal{E}_d$ to zero.
(2) Repeat until the rank of $\mathcal{E}_d$ has not increased for s iterations:
    (a) Generate d pseudorandom elements $x_1, \dots, x_d \in \mathcal{A}$ represented as $r \times 1$ column vectors with components in $\{0, 1, \dots, p{-}1\}$.
    (b) For $j = 1, 2, \dots, (2d-3)!!$ evaluate the j-th monomial with $a_k = x_k$ $(k = 1, \dots, d)$ and store the result in column j of the lower block of $\mathcal{E}_d$.
    (c) Compute the row canonical form of $\mathcal{E}_d$; the lower block is now zero.

Definition 5.4. The *generator matrix* for commutative nonassociative identities in degree d is the matrix $\mathcal{G}_d$ with $(2d{-}3)!! + d!$ rows and $(2d{-}3)!!$ columns; the upper block has $(2d{-}3)!!$ rows and the lower block has $d!$ rows. The columns correspond to the ordered multilinear monomials. The rows of the lower block correspond to the permutations of the d variables.

We use the generator matrix to find a subset of the nullspace basis which generates all the identities in the sense that every identity is a linear combination of permutations of the generating identities.

Algorithm 5.5 (Module Generators)**.** *Input:* An ordered set of generators $I_1, \ldots, I_k$ for a submodule $U \subseteq P_d$ of identities in degree d satisfied by the algebra $\mathcal{A}$. *Output:* The ordered subset which generates U and is minimal (each element does not belong to the submodule generated by the previous elements).

(1) Initialize the generator matrix $\mathcal{G}_d$ to zero. Set oldrank $\leftarrow 0$.
(2) For $\ell = 1, \ldots, k$ do:
  (a) Set $i \leftarrow 0$.
  (b) For each permutation σ of the d variables do:
    (i) Increment i.
    (ii) Apply σ to each monomial in I_ℓ and replace the resulting monomial by the standard representative of its equivalence class.
    (iii) Store the permuted identity σI_ℓ in row i of the lower block.
  (c) Compute the row canonical form of $\mathcal{G}_d$; the lower block is now zero.
  (d) Set newrank $\leftarrow$ rank$(\mathcal{G}_d)$.
  (e) If newrank $>$ oldrank then record I_ℓ as a new generator.
  (f) Set oldrank $\leftarrow$ newrank.

We can now find polynomial identities for the copular algebras D^2G_α. Unlike the gametic and zygotic algebras, we must distinguish two cases. For $\alpha \geq 3$, D^2G_α satisfies a space of dimension 60 of identities in degree 5 (Theorem 5.7). For $\alpha = 2$, D^2G_2 satisfies a space of dimension 65 of identities in degree 5 which includes the identities for $\alpha \geq 3$ (Theorem 5.9).

Definition 5.6. We define polynomials $S, T \in P_5$:

$$\begin{aligned} S &= ((ab)d)(ce) - ((ab)e)(cd) + ((ac)d)(be) - ((ac)e)(bd) + ((bc)d)(ae) \\ &\quad - ((bc)e)(ad)\,, \\ T &= ((ab)d)(ce) - ((ac)e)(bd) + ((ad)b)(ce) - ((ae)c)(bd) - ((bd)a)(ce) \\ &\quad + ((ce)a)(bd)\,. \end{aligned}$$

Theorem 5.7. *For $\alpha \geq 3$, every identity of degree ≤ 7 for the copular algebras D^2G_α follows from commutativity, $R(a,b,c,d)e$, S and T. (The last three identities also hold for $\alpha = 2$.)*

Proof. We use the `Maple` package `LinearAlgebra[Modular]` with $p = 101$. The nonzero products in D^2G_3 (mod p) are:

$$\begin{array}{lll} A_{1111}A_{1111} = A_{1111}\,, & A_{1111}A_{1112} = 51A_{1112}\,, & A_{1111}A_{1113} = 51A_{1113}\,, \\ A_{1111}A_{1212} = 76A_{1122}\,, & A_{1111}A_{1213} = 76A_{1123}\,, & A_{1111}A_{1313} = 76A_{1133}\,, \\ A_{1112}A_{1112} = 76A_{1212}\,, & A_{1112}A_{1113} = 76A_{1213}\,, & A_{1112}A_{1212} = 38A_{1222}\,, \\ A_{1112}A_{1213} = 38A_{1223}\,, & A_{1112}A_{1313} = 38A_{1233}\,, & A_{1113}A_{1113} = 76A_{1313}\,, \\ A_{1113}A_{1212} = 38A_{1322}\,, & A_{1113}A_{1213} = 38A_{1323}\,, & A_{1113}A_{1313} = 38A_{1333}\,, \\ A_{1212}A_{1212} = 19A_{2222}\,, & A_{1212}A_{1213} = 19A_{2223}\,, & A_{1212}A_{1313} = 19A_{2233}\,, \\ A_{1213}A_{1213} = 19A_{2323}\,, & A_{1213}A_{1313} = 19A_{2333}\,, & A_{1313}A_{1313} = 19A_{3333}\,. \end{array}$$

Step 1: We show that D^2G_3 satisfies identities in degree 5 over $\mathbb{F}_p$; we find a basis for the space of identities and a set of generators for the module of identities.

There are 105 monomials in degree 5 and D^2G_3 has dimension 21, so the expansion matrix has size 126×105. The first iteration of Algorithm 5.3 generates these 5 pseudorandom elements of D^2G_3 (as row vectors):

$$\begin{bmatrix} 58 & 91 & 36 & 15 & 16 & 99 & 0 & 77 & 62 & 33 & 57 & 93 & 50 & 44 & 49 & 49 & 33 & 76 & 96 & 5 & 79 \\ 15 & 43 & 54 & 6 & 11 & 7 & 32 & 52 & 36 & 97 & 72 & 89 & 0 & 16 & 77 & 65 & 50 & 94 & 100 & 65 & 28 \\ 54 & 6 & 37 & 71 & 68 & 65 & 12 & 61 & 72 & 61 & 98 & 81 & 90 & 65 & 85 & 77 & 39 & 93 & 9 & 2 & 65 \\ 14 & 54 & 45 & 87 & 41 & 55 & 7 & 63 & 43 & 98 & 25 & 17 & 29 & 65 & 3 & 63 & 57 & 72 & 90 & 89 & 11 \\ 31 & 34 & 83 & 19 & 100 & 54 & 2 & 27 & 6 & 75 & 49 & 74 & 31 & 7 & 8 & 26 & 84 & 77 & 69 & 26 & 62 \end{bmatrix}$$

We evaluate the monomials on these elements, store the results in the lower block of the expansion matrix, and compute the row canonical form; the rank is 21. The second iteration generates another 5 elements; we again fill and reduce the expansion matrix; the rank is 40. After the third iteration the rank is 45; we perform another 100 iterations but the rank does not increase. Thus the space of identities has dimension 60. From the row canonical form we extract the canonical basis for the nullspace, and sort the vectors by increasing number of nonzero components. Using Algorithm 5.5 we find that identity 1 generates a submodule of dimension 45, identities 2–35 belong to the submodule generated by identity 1, identity 36 increases the dimension to 55, and identity 37 increases the dimension to 60 (the entire nullspace). Thus every identity of degree 5 follows from commutativity and identities 1, 36, 37: these are $R(a,b,c,d)e$, S, T.

Step 2a. Each of the three identities I in degree 5 produces six liftings:

$$I(a,b,c,d,e)f,$$
$$I(af,b,c,d,e),\ I(a,bf,c,d,e),\ I(a,b,cf,d,e),\ I(a,b,c,df,e),\ I(a,b,c,d,ef).$$

We apply Algorithm 5.5 to these 18 identities and find seven module generators:

$$(R(a,b,c,d)e)f,\ R(af,b,c,d)e,\ R(a,b,c,d)(ef),\ S(af,b,c,d,e),\ S(a,b,c,df,e),$$
$$T(af,b,c,d,e),\ T(a,bf,c,d,e).$$

These identities cumulatively generate submodules of dimensions 270, 570, 705, 786, 826, 835 and 840.

Step 2b: There are 945 monomials in degree 6, so the expansion matrix has size 966×945. The first seven iterations of Algorithm 5.3 produce ranks 21, 41, 60, 75, 90, 102, 105; we then perform another 100 iterations but the rank does not increase. Thus the nullspace of identities for D^2G_3 has dimension 840. Since this is also the dimension of the space of lifted identities, it follows that every identity in degree 6 follows from commutativity and the identities in degree 5.

Step 3: We extend these computations to degree 7; the expansion matrix has size 10416×10395. The liftings of the identities from degree 6 generate a module of dimension 10185. The stable rank of the expansion matrix is 210, and so the nullspace of identities also has dimension 10185. Thus every identity in degree 7 follows from commutativity and the identities in degree 5.

Step 4: Since $D^2G_3 \subseteq D^2G_\alpha$ for $\alpha \geq 3$, the identities of D^2G_α are a subspace of the identities for D^2G_3. We verify that the three identities in degree 5 are satisfied by D^2G_α in characteristic 0 by independent computations with the copular algebra on 20 alleles, as follows. Each identity is multilinear in 5 variables, and each basis element has 4 subscripts, so in D^2G_{20} we take

$$a = A_{1,2,3,4}\,,\ b = A_{5,6,7,8}\,,\ c = A_{9,10,11,12}\,,\ d = A_{13,14,15,16}\,,\ e = A_{17,18,19,20}\,.$$

(Subscript j stands for an arbitrary subscript i_j.) An identity in degree 5 involves at most 20 alleles for arbitrary α, so this is the general case. For any α and for any 5 basis elements of D^2G_α which generate a subalgebra $H \subseteq D^2G_\alpha$, there is a surjective homomorphism $\Phi\colon D^2G_{20} \to H$ defined as follows:

$$\text{for any } \phi\colon \{1,\dots,20\} \to \{1,\dots,\alpha\} \text{ we set } \Phi(A_{j_1j_2j_3j_4}) = A_{\phi(j_1)\phi(j_2)\phi(j_3)\phi(j_4)}.$$

We use rational arithmetic to evaluate the three identities in degree 5 on the elements $a, b, c, d, e \in D^2G_{20}$, and find that all three collapse to 0. Thus the three identities in degree 5 are satisfied over $\mathbb{Q}$ by D^2G_α for $\alpha \geq 3$, and hence the identities for D^2G_3 are the same as the identities for D^2G_α for $\alpha \geq 3$. □

Definition 5.8. We define polynomials $U, V \in P_5$:

$$\begin{aligned} U &= ((ab)c)(de) - ((ab)d)(ce) - ((ac)b)(de) + ((ac)d)(be) + ((ad)b)(ce) \\ &\quad - ((ad)c)(be)\,, \\ V &= 3((ab)c)(de) + 2((ac)b)(de) - 2((ac)e)(bd) + ((ad)b)(ce) - 4((ad)e)(bc) \\ &\quad - 3((bc)a)(de) + 3((cd)b)(ae)\,. \end{aligned}$$

Theorem 5.9. *For $\alpha = 2$, every identity of degree ≤ 7 for the copular algebra D^2G_2 follows from commutativity, $R(a, b, c, d)e$, U and V. (The two identities U and V do not hold for D^2G_α with $\alpha \geq 3$.)*

Proof. The proof of the first claim is similar to that of Theorem 5.7. For the second claim, since $D^2G_3 \subseteq D^2G_\alpha$ for $\alpha \geq 3$, it suffices to show that U and V do not hold for D^2G_3. We set $a = A_{1111}$, $b = A_{1111}$, $c = A_{1112}$, $d = A_{1113}$, $e = A_{1112}$ and evaluate $U(a, b, c, d, e)$, $V(a, b, c, d, e)$; both equal

$$15A_{1122} + 86A_{1123} + 86A_{1212} + 15A_{1213} + 15A_{1223} + 86A_{1322}\,.$$

Since the result is nonzero over $\mathbb{F}_p$, it is also nonzero over $\mathbb{Q}$. □

Remark 5.10. Costa [6, Corollary 1, page 130] has shown that the copular algebras D^2G_α satisfy the identity $(C(a, b, c, d)e)f$ in degree 6, where C is the identity of Definition 3.7.

Corollary 5.11. *For any $\alpha \geq 2$ and any $i_1, \dots, i_{16} \in \{1, \dots, \alpha\}$, the following element is an absolute zero-divisor in the copular algebra D^2G_α over $\mathbb{Q}$:*

$$\begin{aligned} &R(A_{i_1i_2i_3i_4}, A_{i_5i_6i_7i_8}, A_{i_9i_{10}i_{11}i_{12}}, A_{i_{13}i_{14}i_{15}i_{16}}) = \\ &-\frac{1}{64}\sum_{p=1}^{4}\sum_{q=5}^{8}\sum_{r=13}^{14}\sum_{s=15}^{16} A_{i_pi_qi_ri_s} - \frac{1}{128}\sum_{p=1}^{8}\sum_{q=13}^{16}\sum_{r=9}^{10}\sum_{s=11}^{12} A_{i_pi_qi_ri_s} \\ &+\frac{1}{256}\sum_{p=1}^{4}\sum_{q=5}^{8}\sum_{r=9}^{12}\sum_{s=13}^{16} A_{i_pi_qi_ri_s} + \frac{1}{128}\sum_{p=1}^{8}\sum_{q=9}^{12}\sum_{r=13}^{14}\sum_{s=15}^{16} A_{i_pi_qi_ri_s}\,. \end{aligned}$$

Proof. The identity $R(a, b, c, d)e = 0$ says that the values of the recombination identity are absolute zero-divisors in every copular algebra D^2G_α. □

Example 5.12. The algebra D^2G_2 has dimension 6 and basis A_{1111}, A_{1112}, A_{1122}, A_{1212}, A_{1222}, A_{2222}. For every assignment of these basis elements to a, b, c, d we compute $R(a, b, c, d)$, and find that the space of absolute zero-divisors has dimension 3 and the following basis:

$$R(A_{1111}, A_{1111}, A_{1112}, A_{1112}) = \tfrac{1}{16}A_{1111} - \tfrac{1}{8}A_{1112} + \tfrac{1}{16}A_{1122}\,,$$

$$R(A_{1111}, A_{1111}, A_{1112}, A_{1212}) =$$
$$\tfrac{1}{16}A_{1111} - \tfrac{1}{16}A_{1112} + \tfrac{1}{16}A_{1122} - \tfrac{1}{8}A_{1212} + \tfrac{1}{16}A_{1222}\,,$$
$$R(A_{1111}, A_{1112}, A_{1112}, A_{1212}) =$$
$$\tfrac{5}{128}A_{1111} - \tfrac{1}{32}A_{1112} + \tfrac{3}{64}A_{1122} - \tfrac{3}{32}A_{1212} + \tfrac{1}{32}A_{1222} + \tfrac{1}{128}A_{2222}\,.$$

Acknowledgement. The authors thank the referee for a very helpful report; in particular, for suggesting Lemma 3.8 and for bringing to our attention the papers of Costa [6] and Peresi [15].

References

[1] Abraham, V. M.: Linearizing quadratic transformations in genetic algebras. Proc. London Math. Soc. (3) 40 (1980), no. 2, 346–363.
[2] Bernad, J., González, S., Martínez, C., Iltyakov, A. V.: Polynomial identities of Bernstein algebras of small dimension. J. Algebra 207 (1998), no. 2, 664–681.
[3] Bernstein, S.: Solution of a mathematical problem connected with the theory of heredity. Ann. Math. Statistics 13 (1942) 53–61.
[4] Bremner, M. R.: Varieties of anticommutative n-ary algebras. J. Algebra 191 (1997), no. 1, 76–88.
[5] Bremner, M. R.: Jordan algebras arising from intermolecular recombination. SIGSAM Bull. 39 (2005), no. 4, 106–117.
[6] Costa, R.: Shape identities in genetic algebras. Linear Alg. Appl. 214 (1995) 119–131.
[7] Etherington, I. M. H.: Genetic algebras. Proc. Roy. Soc. Edinburgh 59 (1939) 242–258.
[8] Ewens, W. J.: Mathematical Population Genetics, I: Theoretical Introduction. Second Edition. Interdisciplinary Applied Mathematics, 27. Springer-Verlag, New York, 2004.
[9] Hardy, G. H.: Mendelian proportions in a mixed population. Science 28 (1908) 49–50.
[10] Holgate, P.: Genetic algebras satisfying Bernstein's stationarity principle. J. London Math. Soc. (2) 9 (1974/75) 612–623.
[11] Landweber, L., Kari, L.: The evolution of cellular computing: nature's solution to a computational problem. Biosystems 52 (1999), nos. 1–3, 3–13.
[12] Lyubich, Y. I.: Basic concepts and theorems of the evolutionary genetics of free populations. Russian Math. Surveys 26 (1971), no. 5, 51–124.
[13] Lyubich, Y. I.: Mathematical Structures in Population Genetics. Translated from the 1983 Russian original by D. Vulis and A. Karpov. Biomathematics, 22. Springer-Verlag, Berlin, 1992.
[14] Ouattara, M.: Sur les algèbres de Bernstein d'ordre 2. Linear Algebra Appl. 144 (1991) 29–38.
[15] Peresi, L. A.: A basis of identities for Bernstein algebras. Int. J. Math. Game Theory Algebra 14 (2004), no. 3, 179–188.
[16] Reed, M. L.: Algebraic structure of genetic inheritance. Bull. Amer. Math. Soc. (N.S.) 34 (1997), no. 2, 107–130.
[17] Sverchkov, S. R.: Structure and representations of Jordan algebras arising from intermolecular recombination. To appear in: Algebras, Representations and Applications (Maresias, Brazil, August 2007), Contemporary Mathematics, American Mathematical Society. See also: Jordan Theory Preprint Archives, Number 236, 25 May 2007.
[18] Svirezhev, Y. M., Passekov, V. P: Fundamentals of Mathematical Evolutionary Genetics. Translated from the Russian by Alexey A. Voinov and Dmitrii O. Logofet. Mathematics and its Applications (Soviet Series), 22. Kluwer Academic Publishers, Dordrecht, 1990.
[19] Weinberg, W.: On the demonstration of heredity in man. In: Boyer, Samuel H. (editor): Papers on Human Genetics, pages 4–15. Prentice-Hall, Englewood Cliffs, 1963.
[20] Wörz-Busekros, A.: Algebras in Genetics. Lecture Notes in Biomathematics, 36. Springer-Verlag, Berlin, 1980.
[21] Zhevlakov, K. A., Slinko, A. M., Shestakov, I. P., Shirshov, A. I.: Rings That Are Nearly Associative. Translated from the Russian by Harry F. Smith. Pure and Applied Mathematics, 104. Academic Press, New York, 1982.

Department of Mathematics and Statistics, University of Saskatchewan, Canada