# On the derivations of gametic algebras for polyploidy with multiple alleles

*R. Costa*

## 1. Introduction

1.1. In this paper we obtain some results concerning derivations of a class of genetic algebras. Let us set down our terminology and notation.

First of all, our base field is the field $\mathbb{R}$ of real numbers. All genetic algebras are algebras over this field and "algebra" means $\mathbb{R}$-algebra, "linear mapping" means "$\mathbb{R}$-linear mapping". We adopt Gonshor's definition: A genetic algebra is a commutative algebra for which there exists a basis $C_0, C_1, \ldots, C_n$ with a multiplication table satisfying the following conditions:

If $C_iC_j = \sum_{k=0}^{n} \lambda_{ijk} C_k \quad (i,j = 0, \ldots, n)$ then:

(i) $\lambda_{000} = 1$
(ii) $\lambda_{0jk} = 0$ if $k < j$
(iii) for $i > 0$ and $j > 0$, $\lambda_{ijk} = 0$ if $k \leq \max\ \{i,j\}$.

Any basis of $A$, satisfying conditions (i), (ii) and (iii), is called a canonical basis of A. In general, $A$ may have many canonical bases.

It is well known ([10]) that, for a given genetic algebra $A$, there exists a unique non-zero homomorphism of algebras $w: A \to R$. This homomorphism is defined by $w(C_0) = 1$ and $w(C_i) = 0$ $(i \geq 1)$ and it is called the weight function of $A$. The algebra $A$ is then a baric algebra. The kernel of $w$, which is an $n$-dimensional ideal of $A$, has $C_1, \ldots, C_n$ as a basis. This ideal will be indicated by $N$.

It is known that the numbers $\lambda_{000} = 1$, $\lambda_{011}, \ldots, \lambda_{0nn}$ are, in fact, independent of the canonical basis of $A$. They are called the train roots or shortly the $t$-roots of A. The proof of this consists in the observation that these numbers are the proper values of all the linear mappings $x \mapsto bx$, where $b$ is an arbitrary element of weight 1.

Recebido em 25/8/82. Versão definitiva em 30/1/82.

The concept of genetic algebra was introduced by Etherington ([1], [2]) in order to investigate the behavior of gametic and zygotic populations, in the case of diploid and tetraploid individuals with one locus and two alleles. Later, Gonshor ([3], [4]) carried these investigations to cover the case of multiallelism and polyploidy with mutations. He showed, in particular, that the gametic and zygotic algebra of such a population is a genetic algebra. Nowadays, many genetic systems are described by such algebras ([10]).

1.2. Let us indicate by $G(n+1, 2m)$ the gametic algebra of a $2m$-ploid population with $n+1$ alleles, which we shall denote here by $A_0, A_1, \ldots, A_n$. This algebra has a natural basis consisting of all monomials of degree $m$ in the "variables" $A_0, A_1, \ldots, A_n$. Each one of these monomials represents one of the gametic types of the population. The multiplication of two of these monomials is an algebraic representation of the distribution of gametic types obtained by the mating of the gametic types corresponding to the given monomials. The number of such monomials is $\binom{m+n}{n}$, the binomial number, and so the dimension of $G(n+1, 2m)$ is $\binom{n+m}{m}$. The basis consisting of these monomials is not a canonical basis of $G(n+1, 2m)$ but the set of monomials $A_0^{i_0}(A_0-A_1)^{i_1}\ldots(A_0-A_n)^{i_n}$ with $i_0 + i_1 + \ldots + i_n = m$ in the variables $A_0, A_0-A_1, \ldots, A_0-A_n$ is a canonical basis, as shown by Gonshor ([4]). He showed that these monomials form a canonical basis, where the multiplication table is given by:

$$[A_0^{i_0}(A_0-A_1)^{i_1}\ldots(A_0-A_n)^{i_n}]\,[A_0^{j_0}(A_0-A_1)^{j_1}\ldots(A_0-A_n)^{j_n}]$$

$$= \begin{cases} \binom{2m}{m}^{-1}\binom{i_0+j_0}{m} A_0^{i_0+j_0-m}(A_0-A_1)^{i_1+j_1}\ldots(A_0-A_n)^{i_n+j_n} & \text{if } i_0+j_0 \geq m \\ 0 \text{ if } i_0+j_0 < m. \end{cases}$$

In particular,

$$A_0^m(A_0^{j_0}(A_0-A_1)^{j_1}\ldots(A_0-A_n)^{j_n}) = \binom{2m}{m}^{-1}\binom{m+j_0}{m} A_0^{j_0}(A_0-A_1)^{j_1}\ldots(A_0-A_n)^{j_n}$$

and so the $t$-roots of $G(n+1, 2m)$ are the real numbers $\binom{2m}{m}^{-1}\binom{m+j}{m}$ for $j = m, m-1, \ldots, 1, 0$ (in this order). The $t$-root $\binom{2m}{m}^{-1}\binom{m+j}{m}$ of $G(n+1, 2m)$ has multiplicity $\binom{m-j+n-1}{m-j}$, cf. Gonshor [4]; this is the number of monomials of degree $m-j$ in $(n+1)-1 = n$ variables; or, what is the same, the algebraic multiplicity of the proper value $\binom{2m}{m}^{-1}\binom{m+j}{m}$ of the linear mapping $x \mapsto A_0^m x$, $x \in G(n+1, 2m)$. Especially for $n=1$, i.e., two alleles in the locus under consideration, all $t$-roots are simple.

Of special importance in genetics are the algebras $G(n+1, 2)$ and $G(2, 2m)$. They correspond to multiallelism and polyploidy respectively. We give some emphasis to these two sequences of gametic algebras.



1.3. Given an algebra $A$ over the real field $\mathbb{R}$, a derivation of $A$ is a linear mapping $d: A \to A$ such that $d(ab) = ad(b) + d(a)b$ for any $a, b$ in $A$. If $d_1$ and $d_2$ are derivations of $A$, the linear mapping $d_1d_2 - d_2d_1$ is also a derivation of $A$. So the set of all derivations of $A$ may be equipped with the operation $(d_1, d_2) \to d_1d_2 - d_2d_1$. This set of derivations is a real Lie algebra, called the derivation algebra of $A$ (cf. [8]).

We shall be concerned here with finding a basis of the derivation algebra of $G(n+1, 2m)$. It should be interesting to find the genetic interpretations of the results presented here.

## 2. Multiallelism only

In this case, the algebras $G(n+1, 2)$ describe the gametic population of a diploid and multiallelic population. The natural basis of $G(n+1, 2)$ is the set of monomials of degree 1, namely $A_0, A_1, \ldots, A_n$. The multiplication table is $A_iA_j = 1/2\,A_i + 1/2\,A_j$, which reads genetically as "the gametes produced by a zygote resulting from the mating of gametes $A_i$ and $A_j$ will be $A_i$ and $A_j$, with equal probability". One canonical basis is defined by $C_0 = A_0$, $C_i = A_0 - A_i$ $(i \geq 1)$ and now the multiplication table is $C_0^2 = C_0$, $C_0C_i = 1/2\,C_i$, $C_iC_j = 0$ if $i \geq 1$ and $j \geq 1$. The $t$-roots of $G(n+1, 2)$ are $1, 1/2, 1/2, \ldots, 1/2$. The weight function $w$ of $G(n+1, 2)$ is given by $w(C_0) = 1$, $w(C_i) = 0$ $(i \geq 1)$ or by $w(A_i) = 1$ $(i = 0, 1, \ldots, n)$. It is well known that $G(n+1, 2)$ satisfies the polynomial equation $x^2 = w(x)x$ for every $x$ in $G(n+1, 2)$. This identity may be linearized to give the two variables identity $2xy = w(x)y + w(y)x$.

We prove in this paragraph that $G(n+1, 2)$ has the greatest derivation algebra among all genetic algebras of dimension $n+1$ and is the only with this property. First of all, we consider the class of baric (commutative) algebras which have a *unique* weight function.

**Theorem 1.** *Let $A$ be a baric algebra having a unique weight function $w$. For every derivation $d$ of $A$ we have $w \circ d = 0$.*

*Proof.* Call $\sigma_t$ the automorphism of A defined by

$$\sigma_t = 1_A + td + \frac{t^2d^2}{2!} + \ldots + \frac{t^nd^n}{n!} + \ldots = e^{td},$$

for each real number $t$ (Jacobson [8]) where $1_A$ is the identity of $A$. Hence $w \circ \sigma_t$ is an algebra homomorphism from $A$ to $\mathbb{R}$, and is clearly non zero. By our hypothesis, $w \circ \sigma_t = w$, for all $t \in \mathbb{R}$. If we take derivatives in both sides of this equality we obtain $w \circ d\sigma_t/dt = 0$; in particular, when $t = 0$, we have $w \circ d = 0$.

**Corollary 1.** *Let $A$ be a baric algebra of dimension $n+1$ with unique weight homomorphism. Then the dimension of the derivation algebra of $A$ is not greater than $n(n+1)$. In particular for every genetic algebra of dimension $n+1$, its derivation algebra has dimension not greater than $n(n+1)$.*

*Proof.* Each derivation maps $A$ into the kernel of the weight function, which has dimension $n$.

**Corollary 2.** *Let $A$ be a baric algebra with a unique weight function $w$. Then the derivation algebra of $A$ can be identified with a subalgebra of the Lie algebra of matrices $(a_{ij})$ $(0 \le i, j \le n)$ such that $a_{0i} = 0$ $(i = 0, \dots, n)$.*

*Proof.* Construct a basis $c_0, c_1, \dots, c_n$ of $A$ such that $w(c_0) = 1$ and $w(c_i) = 0$ $(i = 1, \dots, n)$. For each derivation $d$ of $A$, $d(c_i)$ doesn't depend on $c_0$.

**Proposition 1.** *The derivations of $G(n+1, 2)$ are exactly those linear mappings $d$ such that $w \circ d = 0$.*

*Proof.* It is enough to prove that, if $w \circ d = 0$ then $d$ is a derivation of $G(n+1, 2)$. In fact, if $x, y \in G(n+1, 2)$, then $2d(xy) = d(2xy) = d(w(x)y + w(y)x) = w(x)d(y) + w(y)d(x) = w(x)d(y) + xw(d(y)) + w(y)d(x) + yw(d(x)) = 2xd(y) + 2d(x)y$ and so $d(xy) = d(x)y + xd(y)$.

**Proposition 2.** *Let $A_0, A_1, \dots, A_n$ be the natural basis of $G(n+1, 2)$ and let $d_{ij}$ $(i \ne j)$ be defined by*

$$d_{ij}(A_k) = \begin{cases} A_i - A_j & \text{if } k = i \\ 0 & \text{otherwise.} \end{cases}$$

*Then the elements $d_{ij}$ $(i \ne j,\ 0 \le i, j \le n)$ form a basis of the derivation algebra of $G(n+1, 2)$.*

*Proof.* By Prop. 1, each $d_{ij}$ is a derivation. It is enough to prove that they are linearly independent. Suppose $\lambda_{ij} \in \mathbb{R}$ and $\Sigma \lambda_{ij}\ d_{ij} = 0$ (here $i, j = 0, 1, \dots, n$ and $i \ne j$). Then for a fixed $k$,

$$0 = \sum_{\substack{i,j \\ i \ne j}} \lambda_{ij} d_{ij}(A_k) = \sum_{\substack{j=0 \\ j \ne k}}^{n} \lambda_{kj} d_{kj}(A_k) = \sum_{\substack{j=0 \\ j \ne k}}^{n} \lambda_{kj}(A_k - A_j) \text{ and}$$

so $\left(\sum_{\substack{j=0 \\ j \ne k}}^{n} \lambda_{kj}\right) A_k = \sum_{\substack{j=0 \\ j \ne k}}^{n} \lambda_{kj} A_j$. By comparing coefficients we get $\lambda_{kj} = 0$.

The condition $wod = 0$, which characterizes the derivations of $G(n+1, 2)$, is still characteristic for the derivations of this algebra in the



sense of the following theorem, which is similar to a theorem of Gonshor [4], giving $G(n+1, 2)$ as the only (up to isomorphism) baric algebra of dimension $n+1$ such that every linear mapping preserving weight is a homomorphism of algebras.

**Theorem 2.** *Let $A$ be a commutative baric algebra of dimension $n+1$, with weight function $w$. Suppose that every linear mapping $d: A \to A$ such that $wod = 0$, is a derivation of $A$. Then $A$ is (isomorphic to) $G(n+1, 2)$.*

*Proof.* We will show that every element $a \in A$ such that $w(a) = 1$ is an idempotent. In fact, given $a \in A$ with $w(a) = 1$, consider $d_a: A \to A$ given by $d_a(x) = w(x)a - x$. We have $w(d_a(x)) = w(w(x)a - x) = w(x)w(a) - w(x) = 0$. Hence, by our hypothesis, $d_a$ is a derivation. Hence $d_a(a^2) = 2ad_a(a)$. But $d_a(a^2) = w(a^2)a - a^2 = a - a^2$ and $d_a(a) = w(a)a - a = 0$. Hence $a = a^2$. Taking now a basis $A_0, A_1, \dots, A_n$ of $A$ such that $w(A_i) = 1$, we have $w(1/2\,(A_i + A_j)) = 1$ and so $1/2(A_i + A_j) = [1/2(A_i + A_j)]^2$ from what follows $A_iA_j = 1/2A_i + 1/2\,A_j$ and the isomorphism is clear.

## 3. Polyploidy only

In this case, the algebras $G(2, 2m)$ describe the gametic population corresponding to a $2m$-ploid and diallelic population. The natural basis of $G(2, 2m)$ is the set of monomials of degree $m$ in the two variables $A_0$ and $A_1$. They are $A_0^m, A_0^{m-1} A_1, \dots, A_0 A_1^{m-1}$, $A_1^m$, so the dimension of $G(2, 2m)$ is $m+1$. The product of two of such monomials is given by

$$(A_0^i\, A_1^{m-i})(A_0^j\, A_1^{m-j}) = \binom{2m}{m}^{-1} \sum_{k=0}^{m} \binom{i+j}{k}\binom{2m-i-j}{m-k} A_0^k\, A_1^{m-k}$$

which is an algebraic way of expressing the distribution of probability for the gametes produced by the zygote obtained by mating the gametes $A_0^i\, A_1^{m-i}$ and $A_0^j\, A_1^{m-j}$.

A canonical basis for $G(2, 2m)$ is the set of monomials $A_0^i(A_0 - A_1)^{m-i}$ $(0 \le i \le m)$, with multiplication given by

$$[A_0^i\,(A_0 - A_1)^{m-i}]\,[A_0^j\,(A_0 - A_1)^{m-j}] = \begin{cases} \binom{2m}{m}^{-1} \binom{i+j}{m} A_0^{i+j-m} (A_0 - A_1)^{2m-i-j} & \text{if } i+j \ge m \\ 0 & \text{otherwise.} \end{cases}$$

If we call $A_0^m = c_0$, $A_0^{m-1}(A_0 - A_1) = c_1, \dots, (A_0 - A_1)^m = c_m$ then we have $c_i c_j = \binom{2m}{i+j}^{-1} \binom{m}{i+j} c_{i+j}$ if $i+j \le m$ and 0 otherwise. The $t$-roots are $t_i = \binom{2m}{i}^{-1} \binom{m}{i}$ $(i = 0, 1, \dots, m)$ and so $1 = t_0 > t_1 = 1/2 > t_2 > \dots > t_m$.

We construct a derivation $\delta$ of $G(2, 2m)$ by defining $\delta$ to be the linear mapping $\delta(c_i) = ic_i (i = 0, 1, \ldots, m)$. It is easy to verify that $\delta$ is indeed a derivation.

A second derivation $\eta$ is defined in the following way:

$$\eta(c_i) = \frac{t_{i+1}}{t_i - t_{i+1}} c_{i+1} \text{ for } 0 \leq i \leq m-1 \text{ and } \eta(c_m) = 0.$$

(Observe that $\eta$ is nilpotent). In order to prove that $\eta$ is a derivation, it is enough to prove that $\eta(c_i c_j) = c_i \eta(c_j) + \eta(c_i) c_j$ holds for all $i$ and $j$. If $i + j > m$ than $c_i c_j = 0$ so $\eta(c_i c_j) = 0$; on the other hand,

$$\eta(c_i) c_j + c_i \eta(c_j) = \frac{t_{i+1}}{t_i - t_{i+1}} c_{i+1} c_j + \frac{t_{j+1}}{t_j - t_{j+1}} c_i c_{j+1} = 0$$

because $i + j + 1 > m$. If $i + j = m$, then $\eta(c_i c_j) = \eta(c_0 c_m) = \eta(t_m c_m) = = t_m \eta(c_m) = 0$ and $c_i \eta(c_j) + \eta(c_i) c_j = c_i \dfrac{t_{j+1}}{t_j - t_{j+1}} c_{j+1} + \dfrac{t_{i+1}}{t_i - t_{i+1}} c_{i+1} c_j = 0$ again because $i + j + 1 = m + 1 > m$. Now for the case $i + j < m$, we must observe that the sequence $t_1/(t_0 - t_1), t_2/(t_1 - t_2), \ldots, t_m/(t_{m-1} - t_m)$ is nothing but the sequence $1, 1 - 1/m, 1 - 2/m, \quad , 1/m$. With this, we have

$$\eta(c_i c_j) = \eta(c_0 c_{i+j}) = \eta(t_{i+j} c_{i+j}) =$$

$$= t_{i+j} \frac{t_{i+j+1}}{t_{i+j} - t_{i+j+1}} c_{i+j+1} \text{ and } \eta(c_i) c_j + c_i \eta(c_j) =$$

$$= \frac{t_{i+1}}{t_i - t_{i+1}} c_{i+1} c_j + \frac{t_{j+1}}{t_j - t_{j+1}} c_i c_{j+1} = \left( \frac{t_{i+1}}{t_i - t_{i+1}} + \frac{t_{j+1}}{t_j - t_{j+1}} \right) c_0 c_{i+j+1} =$$

$$= \left( \frac{t_{i+1}}{t_i - t_{i+1}} + \frac{t_{j+1}}{t_j - t_{j+1}} \right) t_{i+j+1} \; c_{i+j+1}.$$ We must prove that

$$\frac{t_{i+j}}{t_{i+j} - t_{i+j+1}} = \frac{t_{i+1}}{t_i - t_{i+1}} + \frac{t_{j+1}}{t_j - t_{j+1}}.$$ We have $$\frac{t_{i+j}}{t_{i+j} - t_{i+j+1}} =$$

$$= \frac{t_{i+j+1} - t_{i+j+1} + t_{i+j}}{t_{i+j} - t_{i+j+1}} = 1 + \frac{t_{i+j+1}}{t_{i+j} - t_{i+j+1}} = \frac{t_{i+1}}{t_i - t_{i+1}} + \frac{t_{j+1}}{t_j - t_j + 1}$$

(because in any arithmetic sequence $\alpha_1, \alpha_2, \ldots$ we have $\alpha_1 + \alpha_{i+j+1} = \alpha_{i+1} + + \alpha_{j+1}$). This proves that $\eta$ is a derivation of $G(2, 2m)$.

It is also clear that $\delta$ and $\eta$ are linearly independent.

**Theorem 3.** *Every derivation $d$ of $G(2, 2m)$ is a linear combination of $\delta$ and $\eta$. Precisely, if $\alpha_{10}$ and $\alpha_{11}$ are the components of $d(c_0)$ and $d(c_1)$ in the direction of $c_1$, then $d = \alpha_{10}\eta + \alpha_{11}\delta$.*



*Proof.* First of all, we recall that for any derivation $d$ of $G(2, 2m)$, we have $w \circ d = 0$, where $w$ is the weight function of $G(2, 2m)$. This implies that for $0 \leq j \leq m$, $d(c_j)$ is a linear combination of $c_1, \ldots, c_m$, say $d(c_j) = = \sum_{i=1}^{m} \alpha_{ij} c_i$. From $c_0^2 = c_0$ (again!) we have

$$\sum_{i=1}^{m} \alpha_{i0} c_i = d(c_0) = 2c_0 d(c_0) = 2c_0 \left( \sum_{i=1}^{m} \alpha_{i0} c_i \right) = \sum_{i=1}^{m} 2\alpha_{i0} c_0 c_i = \sum_{i=1}^{m} 2\alpha_{i0} t_i c_i.$$

Comparing coefficients and remembering that $t_1 = 1/2$ and $t_i < 1/2$ for $i \geq 2$, we conclude that $\alpha_{10}$ is arbitrary but $\alpha_{20} = \alpha_{30} = \ldots = \alpha_{m0} = 0$. This means that

$$d(c_0) = \alpha_{10} c_1 \tag{1}$$

From the equality $c_0 c_1 = t_1 c_1$ we get, applying $d$,

$$\alpha_{10} c_1 c_1 + c_0 \left( \sum_{i=1}^{m} \alpha_{i1} c_i \right) = t_1 \left( \sum_{i=1}^{m} \alpha_{i1} c_1 \right) \text{ or}$$

$\alpha_{10} t_2 c_2 + \sum_{i=1}^{m} \alpha_{i1} t_i c_i = \sum_{i=1}^{m} \alpha_{i1} t_1 c_i$. The comparison of coefficients gives the relations:

$$\begin{cases} \alpha_{10} t_2 + \alpha_{21} t_2 = \alpha_{21} t_1 \\ \alpha_{i1} t_i \;= \alpha_{i1} t_1 \qquad (i > 2), \end{cases}$$

from which we deduce that $\alpha_{21} = t_2/(t_1 - t_2)\ \alpha_{10}$ and $\alpha_{i1} = 0$ for $i > 2$. Hence:

$$d(c_1) = \alpha_{11} c_1 + \alpha_{10} \frac{t_2}{t_1 - t_2} c_2. \tag{2}$$

Taking now $c_0 c_2 = t_2 c_2$ we get:

$$\alpha_{10} c_1 c_2 + c_0 \left( \sum_{i=1}^{m} \alpha_{i2} c_i \right) = t_2 \left( \sum_{i=1}^{m} \alpha_{i2} c_i \right) \text{ or}$$

$$\alpha_{10} t_3 c_3 + \sum_{i=1}^{m} \alpha_{i2} t_i c_i = \sum_{i=1}^{m} \alpha_{i2} t_2 c_i \text{ and so}$$

$$\begin{cases} \alpha_{12} t_1 \;= \alpha_{12} t_2 \\ \alpha_{10} t_3 + \alpha_{32} t_3 = \alpha_{32} t_2 \\ \alpha_{i2} t_i \;= \alpha_{i2} t_2 \qquad (i > 3) \end{cases}$$

from which we obtain $\alpha_{12} = 0$, $\alpha_{32} = \dfrac{t_3}{t_2 - t_3} \alpha_{10}$ and $\alpha_{i2} = 0$ for $i > 3$. Hence, for the moment, we have the equality

(3) $$d(c_2) = \alpha_{22}c_2 + \alpha_{10}\frac{t_3}{t_2 - t_3}c_3$$

Naturally, by repetition of this argument, we will conclude that

(4) $$d(c_k) = \alpha_{kk}c_k + \alpha_{10}\frac{t_{k+1}}{t_k - t_{k-1}}c_{k+1}$$

for any $k \le m-1$ and $d(c_m) = \alpha_{mm}c_m$. We prove now that

(5) $$\alpha_{kk} = k\alpha_{11}\ (k = 1, 2, \ldots, m)$$

by induction on $k$. The case $k = 1$ is obvious. Suppose now that $1 < k \le m-1$ and that we have already proved that $\alpha_{k-1,k-1} = (k-1)\alpha_{11}$. This means that

$$d(c_{k-1}) = (k-1)\alpha_{11}c_{k-1} + \alpha_{10}\frac{t_k}{t_{k-1} - t_k}c_k.$$

From the equality $c_1c_{k-1} = t_kc_k$ we get $d(c_1)c_{k-1} + c_1d(c_{k-1}) = t_kd(c_k)$ and so:

$$\left[\alpha_{11}c_1 + \alpha_{10}\frac{t_2}{t_1 - t_2}c_2\right]c_{k-1} + c_1\left[(k-1)\alpha_{11}c_{k-1} + \alpha_{10}\frac{t_k}{t_{k-1} - t_k}c_k\right] =$$

$= t_k\left[\alpha_{kk}c_k + \alpha_{10}\frac{t_{k+1}}{t_k - t_{k+1}}c_{k+1}\right]$. This gives:

$$\alpha_{11}t_kc_k + \alpha_{10}\frac{t_2}{t_1 - t_2}t_{k+1}c_{k+1} + \alpha_{11}(k-1)t_kc_k +$$

$$+\alpha_{10}\frac{t_k}{t_{k-1} - t_k}t_{k+1}c_{k+1} = t_k\alpha_{kk}c_k + t_k\alpha_{10}\frac{t_{k+1}}{t_k - t_{k+1}}c_{k+1}$$ and so:

$$\begin{cases}\alpha_{11} + \alpha_{11}(k-1) = \alpha_{kk}\\ \dfrac{t_2}{t_1 - t_2} + \dfrac{t_k}{t_{k-1} - t_k} = \dfrac{t_k}{t_k - t_{k+1}}\end{cases}$$

The first is the desired equality and the second is an identity for every $k$ (as in the proof that $\eta$ was a derivation). The relations (1) to (5) show exactly that $d = \alpha_{10}\eta + \alpha_{11}\delta$.

**Corollary 1.** *For any $m \ge 1$, the derivation algebra of $G(2, 2m)$ is isomorphic to the non abelian Lie algebra of dimension 2.*

*Proof.* Clear, since $\delta$ and $\eta$ do not commute. In fact, $\delta\eta - \eta\delta = \eta$.

**Corollary 2.** *The proper values of a non-nilpotent derivation of $G(2, 2m)$ are of the form $0, \alpha, 2\alpha, \ldots, m\alpha$, for some non-zero real number $\alpha$.*



## 4. Multiallelism and Polyploidy

As we know, $G(n+1, 2m)$ has a canonical basis consisting of all monomials $A_0^{i_0}(A_0 - A_1)^{i_1}\ldots(A_0 - A_n)^{i_n}$ with $i_0 + i_1 + \ldots + i_n = m$. The $t$-roots are given by

$$t_0 = 1,\ t_1 = 1/2, \ldots, t_j = \frac{\binom{2m-j}{m}}{\binom{2m}{m}}, \ldots, t_m = \frac{1}{\binom{2m}{m}}$$

with multiplicities $1, n, \ldots, \binom{j+n-1}{j}, \ldots, \binom{m+n-1}{m}$ respectively.

In order to simplify notations, we denote the variables $A_0, A_0 - A_1, \ldots, A_0 - A_n$ by $X_0, X_1, \ldots, X_n$. So we have the multiplication table of $G(n+1, 2m)$:

$$(X_0^{i_0}X_1^{i_1}\ldots X_n^{i_n})\,(X_0^{j_0}X_1^{j_1}\ldots X_n^{j_n})$$

$$= \begin{cases}\binom{2m}{m}^{-1}\binom{i_0+j_0}{m}X_0^{i_0+j_0-m}X_1^{i_1+j_1}\ldots X_n^{i_n+j_n} & \text{if } i_0 + j_0 \ge m\\ 0 \text{ otherwise}\end{cases}$$

The weight function of $G(n+1, 2m)$ is defined to be 1 on $X_0^m$ and 0 on the other monomials.

Let us call now $V_i$ the subspace of $G(n+1, 2m)$ generated by the monomials of degree $m-i$ in the variable $X_0$. So $V_0$ is generated by $X_0^m$, $V_1$ is generated by $X_0^{m-1}X_i$ (i $= 1, \ldots, n$), and so on. Each subspace $V_i$ is the proper subspace of the linear mapping $x \mapsto X_0^m x$, $x \in G(n+1, 2m)$, corresponding to the proper value $t_i$. We have the direct sum decomposition $V_0 \oplus V_1 \oplus \ldots \oplus V_m = G(n+1, 2m)$. Observe also that $V_1 \oplus \ldots \oplus V_m$ is the kernel of $w$. When $n = 1$ (polyploidy only), each subspace $V_i$ has dimension 1, because the $t$-roots are all simple.

**Lemma 1.** *For any derivation $d$ of $G(n+1, 2m)$, $d(X_0^m) = A \in V_1$, that is, $d(V_0) \subset V_1$.*

*Proof.* We already know that $w \circ d = 0$. So we may write $d(X_0^m) = A + v_2 + + v_3 + \ldots + v_{m-1} + v_m$, with $A \in V_1$ and $v_i \in V_i$ $(i \ge 2)$. The equality $(X_0^m)^2 = = X_0^m$ implies that $2X_0^m(A + v_2 + \ldots + v_{m-1} + v_m) = A + v_2 + \ldots + v_{m-1} + v_m$. But $X_0^m v_i = t_iv_i$ $(i = 1, \ldots, m)$ and $t_1 = 1/2$, which imply that

$$A + 2t_2v_2 + \ldots + 2t_{m-1}v_{m-1} + 2t_mv_m = A + v_2 + \ldots + v_{m-1} + v_m$$ and so $$v_2 = v_3 = \ldots = v_{m-1} = v_m = 0,$$ that is $d(X_0^m) = A \in V_1$.

**Lemma 2.** *For any derivation $d$ of $G(n+1, 2m)$ and any $1 \le j \le n$, we have $d(X_0^{m-1}X_j) = B_j + \dfrac{1}{t_1 - t_2}A(X_0^{m-1}X_j)$, where $B_j$ is some element of $V_1$ and $A = d(X_0^m)$. In particular, $d(V_1) \subset V_1 \oplus V_2$.*

*Proof*: Call $d(X_0^{m-1} X_j) = B_j + C_j + v_3 + \ldots + v_{m-1} + v_m$, with $B_j \in V_1$, $C_j \in V_2$ and $v_i \in V_i$ $(i \geq 3)$. From the equality $X_0^m(X_0^{m-1} X_j) = t_1(X_0^{m-1} X_j)$ we obtain:

$A(X_0^{m-1} X_j) + X_0^m(B_j + C_j + v_3 + \ldots + v_m) = t_1(B_j + C_j + v_3 + \ldots + v_m)$,
that is, $A(X_0^{m-1} X_j) + t_1 B_j + t_2 C_j + t_3 v_3 + \ldots + t_m v_m = t_1(B_j + C_j + v_3 + \ldots + v_m)$.

But $A(X_0^{m-1} X_j)$ belongs to $V_2$ so we must have:

$$\begin{cases} t_1 C_j = t_2 C_j + A(X_0^{m-1} X_j) \\ t_i v_i \;= t_1 v_i \quad (i \geq 3) \end{cases} \quad \text{or}$$

$$\begin{cases} C_j = \dfrac{1}{t_1 - t_2} \; A(X_0^{m-1} X_j) \\ v_i \;= 0 \quad (i \geq 3) \end{cases}$$

which gives $d(X_0^{m-1} X_j) = B_j + \dfrac{1}{t_1 - t_2} A(X_0^{m-1} X_j)$.

A repetition of the same argument will prove that for any derivation $d$ of $G(n+1, 2m)$ and any $1 \leq j, k \leq m$,

$$d(X_0^{m-2} X_j X_k) = P_{jk} + \frac{1}{t_2 - t_3} A(X_0^{m-2} X_j X_k)$$

where $P_{jk}$ is some element of $V_2$, depending on $X_j$ and $X_k$. More generally, we will have the formulae

$$d(X_0^{m-p} X_{i_1} \ldots X_{i_p}) = P_{i_1 \cdots i_p} + \frac{1}{t_p - t_{p+1}} A(X_0^{m-p} X_{i_1} \ldots X_{i_p})$$

which describe, at least partially, the action of $d$ on the basis of $V_p$, for every $p \leq m-1$. Here $P_{i_1 \cdots i_p}$ is some element of $V_p$, depending on $X_{i_1}, \ldots, X_{i_p}$.

**Lemma 3.** *For any derivation $d$ of $G(n+1, 2m)$ we have*

$$d(X_{i_1} \ldots X_{i_m}) = P_{i_1} \ldots {}_{i_m} \in V_m.$$

*Proof.* We have $X_0^m(X_{i_1} \ldots X_{i_m}) = t_m(X_{i_1} \ldots X_{i_m})$ from which we obtain $A(X_{i_1} \ldots X_{i_m}) + X_0^m d(X_{i_1} \ldots X_{i_m}) = t_m d(X_{i_1} \ldots X_{i_m})$. But $A(X_{i_1} \ldots X_{i_m}) = 0$ because $A$ is a linear combination of monomials like $X_0^{m-1} X_i$. If we decompose $d(X_{i_1} \ldots X_{i_m})$ as $v_1 + v_2 + \ldots + v_{m-1} + P_{i_1} \cdots {}_{i_m}$, with $v_i \in V_i$ $(i = 1, \ldots, m-1)$ and $P_{i_1} \ldots {}_{i_m} \in V_m$, we have $t_1 v_1 + t_2 v_2 + \ldots + t_{m-1} v_{m-1} + t_m P_{i_1 \cdots i_m} = t_m(v_1 + v_2 + \ldots + v_{m-1} + P_{i_1 \cdots i_m})$ and so $v_1 = v_2 = \ldots = v_{m-1} = 0$.

We need now a better understanding of the elements $P_{i_1 \cdots i_p}$ $(2 \leq p \leq m)$ which appeared above.



**Lemma 4.** *For any derivation $d$ of $G(n+1, 2m)$ and any monomial $X_0^{m-2} X_j X_k$ in $V_2$, we have:*

$$d(X_0^{m-2} X_j X_k) = \frac{1}{t_2} [(X_0^{m-1} X_j) B_k + (X_0^{m-1} X_k) B_j] + \frac{1}{t_2 - t_3} A(X_0^{m-2} X_j X_k),$$

*where $B_j$ and $B_k$ are those elements of Lemma 2.*

*Proof.* Take the equality $(X_0^{m-1} X_j)\,(X_0^{m-1} X_k) = t_2(X_0^{m-2} X_j X_k)$. By derivation, we obtain:

$$(X_0^{m-1} X_j)(B_k + C_k) + (X_0^{m-1} X_k)(B_j + C_j) = t_2(P_{jk} + \frac{1}{t_2 - t_3} A(X_0^{m-2} X_j X_k))$$

where $C_k$ and $C_j$ are those elements of Lemma 2. But $(X_0^{m-1} X_j) C_k$ and $(X_0^{m-1} X_k) C_j$ are in $V_3$, hence we must have $(X_0^{m-1} X_j) B_k + (X_0^{m-1} X_k) B_j = t_2 P_{jk}$.

Lemma 4 is the first induction step in the following:

**Lemma 5.** *For any derivation $d$ of $G(n+1, 2m)$ and any monomial $X_0^{m-p} X_{i_1} \ldots X_{i_p}$ in $V_p$ $(2 \leq p \leq m)$ we have the following recurrence relation:*

$$d(X_0^{m-p} X_{i_1} \ldots X_{i_p}) = \frac{1}{t_p} [P_{i_1 \cdots i_{p-1}} (X_0^{m-1} X_{i_p}) +$$

$$+ B_{i_p} X_0^{m-(p-1)} X_{i_1} \ldots X_{i_{p-1}}] + \frac{1}{t_p - t_{p+1}} A(X_0^{m-p} X_{i_1} \ldots X_{i_p}).$$

*Proof.* Start with $(X_0^{m-(p-1)}) X_{i_1} \ldots X_{i_{p-1}}) (X_0^{m-1} X_{i_p}) = t_p(X_0^{m-p} X_{i_1} \ldots X_{i_p})$ and use induction.

**Remark.** The recurrence relation expressed in lemma 5 can be resolved to give the following explicit formula for $d$:

$$d(X_0^{m-p} X_{i_1} \ldots X_{i_p}) = \frac{1}{t_p} \sum_{j=1}^{p} X_0^{m-p} X_{i_1} \ldots \hat{X}_{i_j} \ldots X_{i_p} B_{i_j} +$$

$$+ \frac{1}{t_p - t_{p+1}} A(X_0^{m-p} X_{i_1} \ldots X_{i_p}),$$ where $\wedge$ indicates absence.

The preceding lemmas describe the action of a derivation on the canonical basis of $G(n+1, 2m)$. These formulae have some indeterminates, which are the elements $A, B_1, \ldots, B_n$ of $V_1$. Let us call this sequence the *characteristic sequence* of the derivation $d$. Hence, given a derivation $d$ of $G(n+1, 2m)$ we can associate to it an element of the vector space $(V_1)^{n+1}$ (direct product of $V_1$, $(n+1)$-times). Conversely, given any element of

$(V_1)^{n+1}$, we define a derivation $d$ of $G(n+1, 2m)$ in the following (recurrent) way: If we give $(A, B_1, \ldots, B_n)$ in $(V_1)^{n+1}$, then $d(A_0^m) = A$, $d(A_0^{m-1} X_i) =$

$$= B_i + \frac{1}{t_1 - t_2} A(X_0^{m-1} X_i),\ d(X_0^{m-2} X_j X_k) = \frac{1}{t_2}(X_0^{m-1} X_j)B_k +$$

$+ (X_0^{m-1} X_k)B_j + \dfrac{1}{t_2 - t_3} A(X_0^{m-2} X_j X_k)$, and so on. The proof that this $d$ is a derivation offers no difficulty.

On the other hand, the correspondence $d \mapsto (A, B_1, \ldots, B_n)$ is linear, as it becomes clear by looking to the formulae above. As we can see easily, the zero derivation corresponds to the sequence $(0, 0, \ldots, 0)$.

A basis for the derivation algebra is obtained by standard ways. If we take $B_1 = \ldots = B_n = 0$ than we will have $n$ derivations corresponding to the choices $A = X_0^{m-1} X_i$ $(i = 1, \ldots, n)$. These derivations are all nilpotent. (These derivations look like $\eta$ in theorem 3). Then taking $A = 0$, $B_i = X_0^{m-1} X_i$, $B_j = 0$ $(j \neq i)$ we obtain another $n$ derivations, which look like $\delta$ in theorem 3. The remaining derivations appear when we take $A = 0$, $B_i = X_0^{m-1} X_j$ $(j \neq i)$, $B_k = 0$ $(k \neq i)$. These are nilpotent. We have proved

**Theorem 4.** *The derivation algebra of $G(n+1, 2m)$ has dimension $n(n+1)$. In fact it has a basis consisting of $n$ diagonalisable derivations and $n^2$ nilpotent derivations.*

**Remark.** One of the referees has simplified the proofs in our earlier version of this paper, by proving that a genetic algebra with weight function $w$ over any field of characteristic 0 whose $t$-roots $\gamma_{011}, \ldots, \gamma_{0nn}$ are all different from 1, satisfies $w o d = 0$ for all derivations $d$. In the case the field is $\mathbb{R}$, this result appears as a particular case of our corollary to th.1.

## References


[1] Etherington, I. M. H.; *Genetic algebras*, Proc. Roy. Soc. Edinburgh 59 (1939) 242-258.

[2] Etherington, I. M. H.; *Non-associative algebras and the symbolism of genetics*, Proc. Roy. Soc. Edinburgh B 61 (1941) 24-42.

[3] Gonshor, H.; *Special train algebras arising in genetics*, Proc. Edinburgh Math. Soc. (2) 12 (1960) 41-53.

[4] Gonshor, H.; *Special train algebras arising in genetics II*, Proc. Edinburgh Math. Soc. (2) 14 (1965) 333-338.

[5] Heuch, I.; *An explicit formula for frequency changes in genetic algebras*, J. Math. Biol. 5 (1977) 43-53.

[6] Heuch, I.; *The genetic algebras for polyploidy with an arbitrary amount of double reduction*, J. Math. Biol. 6 (1978) 343-352.





[7] Holgate, P.; *Genetic algebras associated with polyploidy*, Proc. Edinburg Math. Soc. (2) 15 (1966) 1-9.

[8] Jacobson, N.; *Lie algebras*, Interscience Publishers, New York, London, Sydney, 1962.

[9] Wörz-Busekros, A.; *Polyploidy with an arbitrary mixture of chromosome and chromatid segregation*, J. Math. Biol. (1978) 353-365.

[10] Wörz-Busekros, A.; *Algebras in genetics*, Lecture Notes in Biomathematics, 36, Springer-Verlag, 1980.



Instituto de Matemática e Estatística – USP
Cidade Universitária
Caixa Postal 20570 (Ag. Iguatemi)
CEP 05508 São Paulo, Brasil