Invitation to Research of New Mathematics from Biology: Evolution Algebras

Jianjun Paul Tian

Department of Mathematical Sciences

New Mexico State University

Las Cruces, NM 88001, USA

Email: jtian@nmsu.edu

Abstract

In this expository article, I propose new research topics and open problems on evolution algebras, meanwhile I summarize my research on this new subject in order to give a coherent perspective. Most of my results were published in my book [4], and a few results are newly derived. For most of the open problems or topics, I provide my suggestions about how to study them. I also describe a background story of evolution algebras where the reader may enjoy the interaction history between algebras and genetics. Evolution algebras have raised some interests in mathematics community. However, it is my hope that the open research projects in this invitation will attract more attention from both applied and pure mathematicians, be a welcome letter to anyone who is interested in conducting research in this area, and eventually add new dimensions to mathematics itself.

1 Project objectives

In today's cutting edge research area of mathematical biology, although much has been emphasized on the importance of mathematics for research in the life science in this century, it remains skeptical to most mathematicians whether it will have substantial impact on mathematics itself. Some mathematicians are still unconvinced that biology will ever lead to new theorems or theories. Unlikely as it may appear to be, a new algebra, which is called evolution algebra, is a new mathematics subject that has actually already arisen from biology.

Evolution algebra is a new type of algebra that is inspired from biological phenomena. When I was studying stochastic processes and genetics, it occurred to me that there exists an intrinsic and general mathematical structure behind the neutral Wright-Fisher models in population genetics, asexual reproduction or generally non-Mendelian inheritance including intracellular population genetics, and Markov chains. Then I defined them as a type of algebra - evolution algebras. Evolution algebras are non-associative and non-power-associative Banach algebras. Indeed, they are natural examples of non-associative complete normed algebras arising from natural science. Evolution algebras, as abstract mathematical representations, on the one hand, formulate our understanding of natural processes in the biological world, and have many applications in problems of the real world. On the other hand, it is established as a new mathematical subject. It turns out that evolution algebras have many application to or connections with different mathematical fields.

Being a new mathematics subject, evolution algebras have quite a few topics that require further study. As the developer of evolution algebras, I also hope to demonstrate that evolution algebras have extensive connections with or applications to various fields of mathematics, besides evolution algebras as an effective tool in the research of biology and Markov chains. By using evolution algebras, we might be able to see problems in many mathematical fields from a new perspective. A further study of evolution algebras themselves and exploration of connections with other mathematics is worthwhile, since it will also render evolution algebras more understandable, acceptable, and applicable in mathematical fields. So, this invitation will present open projects that mainly focus on the mathematical aspect of evolution algebras.

The objectives of this invitation to open research projects are:

1. To generalize several important theorems in evolution algebras
2. To explore connections between evolution algebras and graph theory, and lay a foundation for a new "algebraic graph theory"
3. To explore connections between evolution algebras and group theory, knot theory, and 3-manifolds
4. To explore connections between evolution algebras and Ihara-Selberg zeta function
5. To establish structure theorems for infinite-dimensional evolution algebras
6. To develop continuous evolution algebras

2 Background

In history, mathematicians and geneticists once used non-associative algebras to study Mendelian genetics. Mendel [6] first exploited symbols that is quite algebraically suggestive to express his genetic laws. It was later termed "Mendelian algebras" by several other authors. In the 1920s and 30s, genetic algebras were introduced. Serebrowsky [7] was the first to give an algebraic interpretation of the sign "$\times$", which indicated sexual reproduction, and to give a mathematical formulation of Mendel's laws. Glivenkov [8] introduced the so-called Mendelian algebras for diploid populations with one locus or two unlinked loci. Independently, Kostitzin [9] also introduced a "symbolic multiplication" to express Mendel's laws. The systematic study of algebras occurring in genetics was due to I. M. H. Etherington. In his series of papers [10], he succeeded in giving a precise mathematical formulation of Mendel's laws in terms of non-associative algebras. Besides Etherington, important contributions have been made by Gonshor [11], Schafer [12], Holgate [13, 14], Hench [15], Reiser [16], Abraham [17], Lyubich [22, 23], and Worz-Busekos [21]. It is worth mentioning two unpublished work in the field. One is the Ph.D thesis of Claude Shannon, the founder of the modern information theory, which was submitted in 1940 (MIT) [18]. Shannon developed an algebraic method to predict the genetic makeup in future generations of a population starting with arbitrary frequencies. The other one is Charles Cotterman's Ph.D thesis that was also submitted in 1940 (Ohio State University) [19, 20]. Cotterman developed a similar system as Shannon did. He also put forward a concept of derivative genes, now called "identical by descent".

During the early days in this area, it appeared that the genetic algebras or broadly defined genetic algebras, could be developed into a field of independent mathematical interest, because these algebras are in general not associative and do not belong to any of the well-known classes of non-associative algebras such as Lie algebras, alternative algebras, or Jordan algebras. They possess some distinguishing properties that lead to many interesting mathematical results. For example, baric algebras, which have nontrivial representations over the underlying field, and train algebras, whose coefficients of rank equations are just functions of the images under these representations, are new concepts for mathematicians. Until recent year, there are still some good research articles in different directions of genetic algebras in literature although it is hard to study nonassociative algebras generally. Ljubic published an algebraic methodology paper [25]. Fortini and Barakat studied algebraic characters of autotetraploid populations [26]. Campos and Holgate studied algebraic isotopy in genetics [27]. Gonshor considered derivations in genetic algebras [28]. Bremner studies some polynomial identities in genetic algebras [29]. There is some continuous time models in genetic algebras [30, 31]. These models can be considered as an effort of introducing a sort of continuous algebras. For the most comprehensive reference in this area, we refer to Worz-Busekros's book [21]. More recent results, such as genetic evolution in genetic algebras, can be found in Lyubich's book [22]. A good survey is Reed's article [24].

Genetic algebras are the product of interaction between biology and mathematics. Mendelian genetics introduced a new subject to mathematics: genetic algebras. The study of these algebras reveals algebraic structures of Mendelian genetics, which always simplifies and shortens the way to understand genetic and evolutionary phenomena. Indeed, it is the interplay between purely mathematical structures and the corresponding genetic properties that makes this area so fascinating. However, after Baur [32] and Correns [33] first detected that chloroplast inheritance departed from Mendel's rules, and much later, mitochondrial gene inheritance was also identified in the same way, non-Mendelian inheritance of organelle genes was recognized with two features — uniparental inheritance and vegetative segregation [37]. Now, non-Mendelian genetics is a basic language of molecular geneticists. Logically, we may ask what non-Mendelian genetics offer to mathematics. The answer is "evolution algebras" [4].

I defined evolution algebras when observing phenomena in biology including uniparental inheritance [35, 36], particle reactions in physics, and Markov chains. It turns out that these algebras have many unique properties. For instance, each evolution algebra possesses an evolution operator. This evolution operator reveals the dynamic information of an evolution algebra. However, what makes the theory of evolution algebras different from the classical theory of algebras is that in evolution algebras, there are two different types of generators: algebraically persistent generators and algebraically transient generators. The basic notions of algebraic persistency and algebraic transiency, and their relative versions, lead to a hierarchical structure for an evolution algebra. Dynamically, this hierarchical structure displays the direction of the flow induced by the evolution operator. Algebraically, this hierarchical structure is given in the form of a sequence of semi-direct-sum decompositions of an evolution algebra. Thus, this hierarchical structure demonstrates that an evolution algebra is a mixed algebraic and dynamic subject. The algebraic nature of this hierarchical structure allows us to have a skeleton-shape classification of evolution algebras. At the same time, the dynamic nature of this hierarchical structure is what makes the notion of evolution algebra applicable to the study of stochastic processes and many other subjects in different fields. For example, when we apply the structure theorem to the evolution algebras induced by Markov chains, it is easy to see that each Markov chain has a dynamic hierarchy and the probabilistic flow is moving through this hierarchy, and that the collection of Markov chains can be classified by the skeleton-shape classification of their induced evolution algebras. Hierarchical structures of Markov chains may be stated in other terms. But, it is the first time that we show algebraic properties of Markov chains and a complete skeleton-shape classification of Markov chains.

Although evolution algebra theory is an abstract system, it gives insight into understanding of non-Mendelian genetics. For instance, once we apply evolution algebra theory to the inheritance of organelle genes, we can predict all the possible mechanisms to establish the homoplasmy of cell populations [4, 41]. Actually in the current stage of mitochondrial research, these mechanisms are hypothetical [38–40], which might shed light on future biological research in mitochondria. It is also easy to understand the coexistence of triplasmy in tissues of sporadic mitochondrial disorder patients by our algebra theory. Further more, once the algebraic structure of asexual progenies of *Phytophthora infectans* is obtained, we can make certain important predictions and suggestions to botany pathologists [42–44].

Evolution algebras have many connections with other fields of mathematics, such as graph theory, group theory, knot theory, 3-manifolds, and Ihara-Selberg zeta functions. Evolution algebras can be a theoretical framework to unify many phenomena. By using evolution algebras, we might be able to see problems in many mathematical fields from a new perspective. For example, we obtain a theorem of classification of directed graphs. Because evolution algebras hold the intrinsic and coherent relation with graph theory, we will be able to analyze graphs algebraically. The purpose for this is that we try to establish a new theory "algebraic graph theory" to reach the goal of Gian-Carlo Rota — "Combinatorics needs fewer theorems and more theory" [34].

Obviously, we are opening a door to a new subject of the mixture of algebras and dynamics. Many new research topics in evolution algebras are confronting us. In this proposed research, our focus is on mathematics of evolution algebras and applications to and connections with other mathematical fields.

3 Preliminary research

Below is a brief description of what I have done about evolution algebras.

We have defined evolution algebras, investigated their basic properties, established the principal theorem about evolution algebras — the hierarchical structure theorem. Evolution algebras are defined in terms of generators and defining relations. We studied their non-associativity and non-power-associativity. Various algebraic concepts are also investigated, such as evolution subalgebras and evolution homomorphisms, the associative multiplication algebra, the centroid and the derived Lie algebra of an evolution algebra etc.. The occurrence relation among generators of an evolution algebra and the connectedness of an evolution algebra are defined. We utilize the occurrence relation to define the periodicity of generators. We introduced an evolution operator for an evolution algebra that is actually a special right (left) multiplication operator. This evolution operator reveals the dynamic information of an evolution algebra. To describe the evolution flow quantitatively, we introduced a norm for an evolution algebra. Under this norm, a finite evolution algebra becomes a Banach algebra. However, what makes the evolution algebra theory different from the classical algebra theory is that in evolution algebras there are two different categories of generators, algebraically persistent generators and algebraically transient generators. Moreover, the difference between algebraic persistency and algebraic transiency suggests a direction of dynamic flow as it displays in the hierarchy of an evolution algebra. The hierarchy gives a picture of a dynamic process when one takes multiplication in an evolution algebra as time-step in a discrete-time dynamic system. We also obtain the structure theorem for a simple evolution algebra. We use homomorphisms to reduce a "big" evolution algebra to a "small" one that still has the same hierarchy as that of the original algebra does. This reducibility gives a classification of all evolution algebras up to skeleton-shapes. To illustrate, I list several definitions and theorems as follows (for details, please look at my book [4]).

Let's first consider a non-associative and non-commutative free algebra $\Re$ with the set of generators $X = \{e_i \mid i \in \Lambda\}$ over a field K, where Λ is a finite or infinite index set. Note that its elements are polynomials of non-commutative variables e_i with coefficients from K and the basis consists of bracketed words. By a bracketed word, we mean a monomial of variables $e_1, e_2, \cdots$, with brackets inserted so that the order of multiplications in the monomial is uniquely determined. We then specify some polynomials $f_l = e_l^2 + \sum\limits_{k \in \Lambda} a_{lk} e_k$, where almost every $a_{lk} \in K$ is zero except finite many of them; $f_{ij} = e_i e_j$ where $i \neq j$. We consider the ideal I in $\Re$ generated by these elements, f_l, f_{ij}, (i.e. the smallest ideal contains these elements). The factor algebra $\Re / I$ is an evolution algebra defined by the generators and the relations. That is, we define an evolution algebra to be

$$R(X) = \left\langle e_1, \cdots, e_v, \cdots \mid \; e_l^2 + \sum_{k \in \Lambda} a_{lk} e_k, \; e_i e_j, i \neq j; i, j, l \in \Lambda \right\rangle .$$

Let E be an evolution algebra with the generator set $\{e_i \mid i \in \Lambda\}$. We say e_i occurs in $x \in E$, if $\alpha_i \neq 0$, where $x = \sum_{j=1}^{v} \alpha_j e_j$. When e_i occurs in x, we write $e_i \prec x$. Note that if $e_i \prec e_i^{[n]}$, then $\langle e_i \rangle \subseteq \langle x \rangle$, where $\langle x \rangle$ means the evolution subalgebra generated by x. We define a period of a generator e_j to be

$$d = g.c.d. \left\{ \log_2 m \mid e_j \prec \left(e_j^{(m)} \right) \right\},$$

where power $e_j^{(m)}$ is some $k - th$ plenary power, $2^k = m$. If d is 1, we say e_j is aperiodic; if the set $\left\{ \log_2 m \mid e_j \prec (e_j^{(m)}) \right\}$, is empty, we define $d = \infty$.

The evolution operator of E is defined to be a map L from E to itself that sends e_i to e_i^2. Or using a formal notation $\theta = \sum\limits_{i \in \Lambda} e_i$, L can be defined L as follows:

$$L(x) = \theta \cdot x = (\sum_{i \in \Lambda} e_i) \cdot x,$$

for any $x \in E$, a special right multiplication operator.

A norm for an evolution algebra is defined to be $N(x) = \sum_i |a_i|$, where $x \in E$ and $x = \sum_i a_i e_i$. Under this norm, a finite dimensional evolution algebra becomes a Banach algebra.

We say that generator e_j is algebraically persistent if the evolution subalgebra $\langle e_j \rangle$, generated by e_j, is a simple subalgebra, and e_i is algebraically transient if the subalgebra $\langle e_i \rangle$ is not simple.

Theorem 3.1. *All generators have the same period in a nonnegative simple evolution algebra.*

Theorem 3.2. *1) Let E be a connected evolution algebra, then E has a proper evolution subalgebra if and only if E has an algebraically transient generator.*

2) Let E be a connected evolution algebra, then E is a simple evolution algebra if and only if E has no algebraically transient generator.

Theorem 3.3. *Any finite dimensional evolution algebra has a simple evolution subalgebra.*

Theorem 3.4. *Let E be a connected finite-dimensional evolution algebra. As a vector space, E has a decomposition of direct sum of subspaces:*

$$E = A_{0,1} \oplus A_{0,2} \oplus \cdots \oplus A_{0,n_0} \dot{+} B_0,$$

where $A_{0,i}$, $i = 1, 2, \cdots, n_0$, are all simple evolution subalgebras, $A_{0,i} \cap A_{0,j} = \{0\}$ for $i \neq j$, and B_0 is a subspace spanned by algebraically transient generators (which is called the $0-th$ transient space). The summation $A_{0,1} \oplus A_{0,2} \oplus \cdots \oplus A_{0,n_0}$ is also a direct sum of subalgebras. Symbol $\dot{+}$ indicates the summation is not a direct sum of subalgebras, just a direct sum of subspaces. We call this decomposition a semi-direct-sum decomposition of an evolution algebra.

We can establish a semi-direct-sum decomposition for B_0 by using induced relative concepts on B_0. This procedure can continue until where we can not decompose the transient spaces. We get the hierarchy as follows

$$\begin{aligned}
E &= A_{0,1} \oplus A_{0,2} \oplus \cdots \oplus A_{0,n_0} \dot{+} B_0 \\
B_0 &= A_{1,1} \oplus A_{1,2} \oplus \cdots \oplus A_{1,n_1} \dot{+} B_1 \\
B_1 &= A_{2,1} \oplus A_{2,2} \oplus \cdots \oplus A_{2,n_2} \dot{+} B_2 \\
& \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
B_{m-1} &= A_{m,1} \oplus A_{m,2} \oplus \cdots \oplus A_{m,n_m} \dot{+} B_m \\
B_m &= B_{m,1} \oplus B_{m,2} \oplus \cdots \oplus B_{m,h},
\end{aligned}$$

where $A_{k,l}$ is a $k-th$ simple evolution subalgebra, $A_{k,l} \cap A_{k,l'} = \{0\}$ if $l \neq l'$, B_k is the $k-th$ transient space. B_m can be decomposed as a direct sum of $(m+1)-th$ simple evolution subalgebras. We may call these $(m+1)-th$ simple evolution subalgebras the heads of the hierarchy, and h is the number of heads.

Theorem 3.5. *Every evolution algebra E is homomorphic to a unique evolution algebra E_r such that its evolution subalgebras in its hierarchy are all one dimensional subalgebras. This unique evolution algebra E_r is called the skeleton-shape of E.*

Theorem 3.6. *Given the numbers of levels and subalgebras at each level, an evolution algebra can be determined up to the skeleton-shape homomorphism.*

To demonstrate the applicability of evolution algebras, we studied evolution algebras that correspond to or are determined by Markov chains. Denote the collection of all Markov chains with a finite state space by $C(M)$, the collection of all finite-dimensional evolution algebras over the real number field R by $C_R(A)$. We defined a map Φ from $C(M)$ to $C_R(A)$, for each Markov chain X, its image is the algebra defined by taking X's transition probabilities to be the structural coefficients. This map is well-defined. $C_R(A)$ is a category with objectives as evolution algebras and morphisms as homomorphisms of algebras. If we identify a Markov chain X and its image $\Phi(X)$, then $C(M)$ becomes a subcategory of $C_R(A)$. $C(M)$ now has algebraic structures. For instance, homomorphisms between Markov chains, and they preserves skeleton-shape of hierarchy of a Markov chain. By applying Theorem 3.5, we can classify all Markov chains, and each Markov chain in the same class shares the same skeleton-shape of hierarchy. By using evolution algebras, the algebraic properties of Markov chains can also be revealed. For example, a Markov chain is irreducible if and only if its evolution algebra is simple, and a subset of state space of a Markov chain is closed in the sense of probability if and only if it generates an evolution subalgebra. An element has the algebraic period of d if and only if it has the probabilistic period of d. In general, a generator is probabilistically transient if it is algebraically transient, and a generator is algebraically persistent if it is probabilistically persistent. When the dimension of the evolution algebra is finite, algebraic concepts and analytic concepts are equivalent. Although the dynamic behavior of an evolution algebra is embodied by various powers of its elements, the evolution operator seems to represent a "total" principal powers. We studied the spectrum of evolution operators at the $0-th$ level in the hierarchy of $\Phi(X)$. We also gave formulae of the sojourn times during each simple evolution subalgebra at each level on the hierarchy.

4 Proposed open research topics

4.1 To generalize several important theorems in evolution algebras

The framework of evolution algebra theory was already established in my book "Evolution Algebras and Their Applications". Further more detailed studies are needed now. The most important thing to do, however, is to generalize several theorems in order to make evolution algebras more understandable and applicable.

Structures of simple evolution algebras: For finite-dimensional simple evolution algebras over R, we know they do not have hierarchic structure. For finite-dimensional nonnegative simple evolution algebras, however, we have proved they have an interesting cyclic dynamic structure. A nonnegative evolution algebra means coefficients in its defining relations are nonnegative. We have the following two theorems to describe the structures of nonnegative simple evolution algebras.

Theorem 4.1. *All generators have the same period in a nonnegative simple evolution algebra.*

Theorem 4.2. *Let E be a nonnegative simple evolution algebra with generator set $\{e_i \mid i \in \Lambda\}$. There is a partition of generators with d disjointed classes C_0, C_2, $\cdots$, C_{d-1}, such that $L(\Delta_k) \subseteq \Delta_{k+1}(modd)$, or $\Delta_k^2 \subseteq \Delta_{k+1}(modd)$, $k = 1, 2, \cdots d-1$, where d is the period of e_i, $\Delta_k = Span\,(C_k)$, and L is the evolution operator of E, mod is taken with respect to the* *index of the class of generators. There is also a direct sum of linear subspaces*

$$E = \Delta_0 \oplus \Delta_1 \oplus \cdots \oplus \Delta_{d-1}.$$

Since simple evolution algebras are the basic blocks for building general evolution algebras, it is necessary to make a detailed study of the structures of simple evolution algebras. To characterize simple evolution algebras, it is necessary to generalize these two theorems for general finite-dimensional simple evolution algebras. Here the difficulty lies in how to prove the existence of transitive occurrence relations. In order to investigate this question, I suggest to study simple evolution algebras over finite fields. These study will shed light on the study of general proofs. For general simple evolution algebras over R, I suggest to construct an index map over the algebra in order to get a proof. I also suggest to study how to relax the conditions for occurrence relations so that the new definition of occurrence relations will still preserve all old relations.

Structures of simple evolution algebras at high levels: Another basically important question that we would like to propose to study is what is the structure of a simple evolution algebra, when it is not at the $0-$th level of the hierarchy of some evolution algebra, but at a high level of the hierarchy. It is expected that simple evolution algebras at a high level of the hierarchy have similar structures to those at the $0-$th level.

Spectrum study at high levels: For finite-dimensional evolution algebras determined by Markov chains, we have studies their spectral structures in the $0-$th levels in their hierarchy, and got two main theorems:

Theorem 4.3. *Let M_X be a finite-dimensional simple evolution algebra with period d determined by Markov chain X, then the evolution operator has d eigenvalues that are the roots of unity. Each of them has an eigenspace of dimension one. And there are no other eigenvalues of modulus one.*

Theorem 4.4. *Let M_X be an evolution algebra of finite dimension, then the geometric multiplicity of the eigenvalue one of its evolution operator is equal to the number of simple evolution subalgebras of M_X.*

It is important to generalize these results to high levels of their hierarchy in order to better understand the details of dynamic systems. So it is needed to study these generalizations. For the type of evolution algebra that has more than two levels, I suggest to adopt methods of operator decomposition at $0-$th level to study the spectrum in high level. Since scales are different at different levels when we use operator decomposition methods, normalization of the subalgebras at the same level will become necessary. It is expected results similar to these two theorems will be obtained at each level.

4.2 To explore connections between evolution algebras and graph theory, and lay a foundation to establish a new theory — "algebraic graph theory"

For each graph, we can define an evolution algebra as follows.

Definition 4.1. *Let $G = (V,\ E)$ be a graph, V be the set of vertices of G, E be the set of edges of G. We define an algebra as follows: taking $V = \{e_1, e_2, \cdots, e_r\}$ as the generator set and*

$$R = \left\{ e_i^2 = \sum_{e_k \in \Gamma(e_i)} e_k;\ e_i \cdot e_j = 0,\ i \neq j; i, j = 1,\ 2,\ \cdots,\ r \right\}$$

as the set of defining relations, where $\Gamma(e_i)$ is the set of neighbors of e_i. Then the evolution algebra determined by this graph is a quotient algebra

$$\begin{aligned} A\left(G\right) &= \langle V \mid R \rangle \\ &= \left\langle e_1, e_2, \cdots, e_r \mid e_i^2 - \sum_{e_k \in \Gamma(e_i)} e_k;\ e_i \cdot e_j, i \neq j; i, j = 1,\ 2,\ \cdots,\ r \right\rangle. \end{aligned}$$

Let's denote the collection of all finite graphs (the number of vertices is finite and the number of edges is also finite) by $C(G)$, and the collection of all finite-dimensional evolution algebras over the R by $C_R(A)$. Then this definition defines a map Ψ from $C(G)$ to $C_R(A)$ that $\Psi(G) = A(G)$. Below is the theorem to guarantee this map is well defined.

Theorem 4.5. *If graphs G_1 and G_2 are isomorphic as graphs, then $A(G_1)$ and $A(G_2)$ are also isomorphic as evolution algebras.*

If a graph is weighted, weights will be used as coefficients in the defining relations 4.1. If a graph is directed, for a vertex we can choose $+1$ as outgoing weight and -1 as ingoing weight, or other way around. So in one words, for any graph, we can define a corresponding evolution algebra, and this algebra contains all the information about this graph. $C_R(A)$ is a category that has rich algebraic structures. Once we identify a graph G and its image $\Psi(G)$, $C(G)$ becomes a subcategory of $C_R(A)$. Then a nice algebraic concept, algebraic homomorphisms, can be obtained in the $C(G)$, and $C(G)$ thus becomes an algebraic subject. All knowledge about evolution algebras can be applied to the study of graphs. For instance, we can apply Theorem 3.5 to $C(G)$ to give a classification theorem of all finite graphs. However, a deep study for graph hierarchy is required.

Algebraic terms can also be interpreted in term of graph theory. For example, if G is a graph with the vertex set $V(G) = \{e_1, e_2, \cdots, e_r\}$, L is the evolution operator of the corresponding algebra $A(G)$, and suppose

$$L^n(e_i) = n_{i1}e_1 + n_{i2}e_2 + \cdots + n_{ir}e_r,$$

then n_{ij} is the total number of paths with length n from vertex e_i to vertex e_j. If $n_{ij} = 0$, this means there is no path with length n from vertex e_i to vertex e_j. We can also have graph interpretation for the destination operator in evolution algebras

$$D_i = \sum_{k=1}^{\infty} \rho_i L(\rho_i^0 L)^{k-1}.$$

Suppose we start from x and our destination is the vertex e_i. Then, in one step, the number of paths from x to e_i is given by $\rho_i L(x)$; in two steps when the first arrival at e_i happens in the second step, the number of paths from x to e_i is given by $\rho_i L(\rho_i^0 L)\,(x)$; in three steps when the first arrival at e_i happens in the third step, the number of paths from x to e_i is given by $\rho_i L(\rho_i^0 L)^2(x)$; and so on. Within n steps, the total number of paths from x to e_i is given by

$$D_i^n(x) = \sum_{k=1}^{n} \rho_i L(\rho_i^0 L)^{k-1}(x).$$

As we see, the map Ψ converts each graph into an evolution algebra, and by studying this algebra, problems in the graph can be resolved. In this way, evolution algebras will become a systemized method to study graphs. This algebraic method has conspicuous advantages over the combinatorial method in graph theory. A well-known fact about combinatorics is that it lacks a systematized method despite its importance in application. Gian-Carlo Rota hoped about combinatorics — “Combinatorics needs fewer theorems and more theory” [34]. Evolution algebra theory may direct combinatorics towards a trend that will be more systematized and more theory-like.

To establish this new “algebraic graph theory”, I propose the following problems.

1. By using map Ψ, convert or translate every major statement in graph theory into evolution algebras. We know $C(G)$ is a subcategory of $C_R(A)$ under the identification graph G and $\Psi(G)$. The homomorphism of two graphs under Ψ is a homomorphism that preserves hierarchies of the image algebras of graphs. Via Ψ back to graph category, it is expected this homomorphism preserves dynamic connection or hierarchies of graphs. Obviously, algebraic theory render more tools available to study $C(G)$.

2. The study of $\Psi(C(G))$ is important to both graph theory and evolution algebras. For example, we can get a classification of all objects in $\Psi(C(G))$ by algebra theory, a

classification of all graphs. As algebraic subject induced from graphs, $\Psi(C(G))$ must possess some properties that reflect graphs' properties. It is important to dig out these properties and to understand what the hierarchy of evolution algebras means to graphs, which is what I would like to suggest to study.

3. Random walks on graphs, an interesting topic, have many applications. While random walks can be mapped into evolution algebras by Φ, and graphs into evolution algebras by Ψ, then what is the relationships between the evolution algebras determined by random walks on a graph and the evolution algebra determined by the same graph. It is very interesting to investigate this problem, which will add a new landscape in discrete geometry.

4. I once introduced "multi-person random walk"on graph or coalescent random walk on graphs [1, 3]. Now I suggest to study this type of random walk by using evolution algebras. It is actually a study of several evolution algebras with the same generator set.

5. Random graphs and their evolution algebras will be another interesting topic. Random graph theory studies the processes of forming graphs when vertices are given. We may consider a random graph as a type of evolution algebra, whose structure coefficients are random variables. I suggest to study this type of evolution algebra with random coefficients following some given probability distributions. Actually, evolution algebras are Banach algebras, the norm is convenient for studying random variable.

4.3 To explore connections between evolution algebras and group theory, knot theory and 3-manifolds

Any group can be associated with an evolution algebra. Specifically, let G be a group, E be a finite set of generators of the group G, and K be a field. To define an evolution algebra, we take G as the generator set, and the defining relations are:

$$g * g = \sum_{e \in E} k_e g e,$$

and

$$g * h = 0,$$

where $*$ is the algebraic multiplication, $k_e \in K$. One question is how the evolution algebra reflects the properties of the corresponding group. Another question is how we can use evolution algebras to study groups or use group theory to study evolution algebras. I suggest to investigate the relations between groups and the associated evolution algebras.

Here is a simple example for illustration. Let G be a cyclic group with generator g or g^{-1}, and we take the underlying field to be the real number field R. If we define

$$g^r * g^r = g^r(g + g^{-1}) = g^{r+1} + g^{r-1.},$$

after a computation, we have

$$(g^r)^{[n]} = \sum_{k=0}^{n} \left(\begin{array}{c} k \\ n \end{array} \right) g^{r+n-2k}.$$

Now if the group G is infinite, then each element of the group has a period 2, except for the unit element of the group, as we can see form the above formula. If the group has order m, then each element has a period $2+m$, except for the unit. So, infiniteness of cyclic group can be reflected by the period of non-unit elements in the evolution algebra. A very interesting case is for braid groups, which are the fundamental algebraic structures behind 3-manifolds and knot theory. We need to study whether anything deep about knots can be obtained from the associated evolution algebras.

By a different association, we can study how to use evolution algebras associated with braids to compute Burau representations (a Markov chain method in [45]). Given $b \in B_n$, and b is a word of generators σ_i, σ_i^{-1} with length k, $i = 1, 2, \cdots, n-1$. We define the $n-$dimensional evolution algebras with generators $\{e_1, e_2, \cdots, e_n\}$. One multiplication in algebras corresponds to one step move up in the diagram of braids. Let us start from the bottom. When we pass σ_i, we do multiplication $e_i^2 = (1-t)e_i + te_{i+1}$ and $e_{i+1}^2 = e_i$; when we pass σ_i^{-1}, we do multiplication $e_i^2 = e_{i+1}$ and $e_{i+1}^2 = \bar{t}e_i + (1-\bar{t})e_{i+1}$, where $\bar{t} = t^{-1}$. Using non associative computation in evolution algebras, we will get

$$(e_i \cdot e_i) \cdot (e_1 + e_2 + \cdots + e_n)^{[k-1]} = P_{i1}(t)e_1 + P_{i2}(t)e_2 + \cdots + P_{in}(t)e_n,$$

where $P_{ij}(t)$ is a Laurent polynomial, $i = 1, 2, \cdots, n$. **I conjecture that** $P(t) = (P_{ij}(t))_{n\times n}$ **is Burau matrix for the braid** b, which need to be proved.

There is a natural way to connect evolution algebras with 3-manifold structure. For a triangulation t_1 of a 3-manifold M, we can define an evolution algebra $A(M, t_1)$ as we assign an evolution algebra to a graph. When taking the barycentric subdivision t_2 of t_1, we will have another evolution algebra $A(M, t_2)$. We can keep this procedure until get an infinite sequence of evolution algebras

$$\{A\left(M, t_n\right)\}_{n=1}^{\infty}.$$

Here arise a lot of interesting questions: what is the limit of this sequence? what kind of properties of the 3-manifold can be reflected by this sequence of algebras? **My conjecture is that the limit of this sequence is closely related to the Laplace-Beltrami operator of 3-manifold** M. I suggest to explore this conjecture.

4.4 To explore connections between evolution algebras and Ihara-Selberg zeta functions

For a finite graph X, we have the Ihara-Selberg zeta function $Z(u)$ given by $Z(u) = \prod_{w \in W}(1-u^{|w|})^{-1}$, where W denotes the set of all prime, reduced cycles in X, $|w|$ denotes the length of the cycle w [46, 47]. When a graph is mapped as an evolution algebra, we will have an algebraic version of Ihara-Selberg-zeta function. To demonstrate the advantage of evolution algebra methods, I suggest to study this combinatoric topic by evolution algebras.

Let A be an evolution algebra, denote the collection of all its simple evolution subalgebras at all level by $\Sigma(A)$. For any simple evolution algebra $E \in \Sigma(A)$, denote the period of E by $\rho(E)$, the maximum dimension of the cyclic subspace decomposition of E by $\delta(E)$. Then, the Ihara-Selberg zeta function for algebra A is given

$$Z(u) = \prod_{E \in \Sigma(A)} (1-u^{\rho(E)})^{-\delta(E)}.$$

Denote the evolution operator of A by L_A, L_E is the evolution operator of the subalgebras E. In terms of evolution operators, we have

$$(uI - L_A)^{-1} = \prod_{E \in \Sigma(A)} (uI - L_E^{\rho(E)})^{-\delta(E)}.$$

This formula needs to be proved algebraically.

4.5 To establish hierarchic structure theorem for infinite dimensional evolution algebras

For finite-dimensional evolution algebras, the hierarchic structure theorem was already obtained, as stated in Preliminary results section. Now I suggest to generalize this result to infinite-dimensional evolution algebras. Obviously, this is a very difficult problem. I would suggest first to study infinite-dimensional simple evolution algebras. Since there is some similarity between evolution algebras and Lie algebras, it might be helpful to borrow some tools in the well studied field of Lie algebra, particularly, for the case of infinite dimensional algebras. One useful reference is Cartan's work about the classification of infinite-dimensional Lie algebras [48]. I expect an infinite-dimensional simple evolution algebra has a finite period (each generator has the same period) or the maximum of periods of its generators is finite.

To study general infinite dimensional evolution algebras, I would also suggest to borrow some ideas from denumerable Markov chains [49], and infinite matrices theory [50,51]. I also suggest to construct a Hilbert space for any given evolution algebra so that each element of the evolution algebra can be viewed as an operator of this Hilbert space. Further more, this evolution algebra is a subset of or equal to von Neumann algebra of the Hilbert space [52]. Then it is expected that a classification theory of infinite dimensional evolution algebras will be obtained.

4.6 To develop continuous evolution algebras

The continuous version of evolution algebras is of considerable significance in application since most natural processes are continuous-time processes. In genetic algebras, for example, although some effort has been attempted in order to introduce continuous-time evolution over evolutionary history [30, 31], it seems to be impossible to incorporate discrete genotypes in the population with continuous-time evolution. However, within the framework of evolution algebras, it can be possible. Here we propose some basic definitions about continuous evolution algebras.

The theory of continuous-time Markov chains over discrete state space gives us hints to define continuous evolution algebras. One way to define continuous evolution algebras is to take coefficients in the defining relations to be differential functions over the underlying field (we use the real number field R). For example, let the generator set to be $\{e_i | i \in \Lambda\}$, the defining relations are

$$e_i \cdot e_i = \textstyle\sum_{j\in\Lambda} a_{ij}(t)e_j,$$
$$e_i \cdot e_j = 0;$$
$$\textstyle\sum_{k\in\Lambda} a_{ik}(t)a_{kj}(s) = a_{ij}(t+s),$$

where $a_{ij}(t)$ is differentiable function, $i, j \in \Lambda$.

Another way to define continuous evolution algebras is to use differential equations. Then algebraic methods and differential equation methods can be used to study continuous evolution algebras. It is expected that continuous evolution algebras will have similar relations with continuous-time Markov chains, just as (discrete) evolution algebras have the relations with discrete-time Markov chains. From viewpoint of category, I expect the set of all continuous-time Markov chains will be a subcategory of the category of continuous evolution algebras. It is also expected that continuous evolution algebras will be a powerful tool to study evolution of dynamic systems, and biological processes.

Acknowledgment The author would like to thank Utkir Rozikov for providing a list of published research papers related to evolution algebras, although I am unable to include them in this article. The author would like to acknowledge the supports DMS-1446139 of NSF of US and NNSF of Chain 11371048.

References

[1] Jianjun Tian & Xiao-Song Lin, *Continuous time Markov processes on graphs*, Stochastic Analysis and Applications, Vol. **24**(2006), no. 5, 953–972.

[2] Jianjun Tian & Bai-Lian Li, *Coalgebraic structure of genetics inheritance*, Mathematical Biosciences and Engineering, Vol. **1** (2004), No. 2, 243-266.

[3] Jianjun Paul Tian & Zhenqiu Liu, *Coalescent random walks on graphs*, Journal of Computational and Applied Mathematics, **202**(2007), no. 1, 144–154.

[4] Jianjun Paul Tian, **Evolution algebras and their applications**, Lecture Notes in Mathematics, **1921**, Springer-Verlag Berlin Heidelberg 2008.

[5] Jianjun Paul Tian & Petr Vojtechovsky, *Mathematical concepts of evolution algebras in non-Mendelian genetics*, Quasigroups Related Systems, **14**(2006), No. 1, 111–122.

[6] G. Mendel, *Experiments in plant-hybridization*, Classic Papers in Genetics, pages 1-20, J. A. Peter editor, Prentice-Hall Inc. 1959.

[7] A. Serebrowsky, *On the properties of the Mendelian equations*, Doklady A.N.SSSR. **2** (1934), 33-36, (in Russian).

[8] V. Glivenkov, *Algebra Mendelienne comptes rendus*, (Doklady) de l'Acad. des Sci. de l'URSS **4** (1936), 385-386, in Russian).

[9] V.A. Kostitzin, *Sur les coefficients mendeliens d'heredite*, Comptes rendus de l'Acad. des Sci. **206** (1938), 883-885, (in French).

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

[11] H. Gonshor, *Contributions to genetic algebras*, Proc. Edinburgh Math. Soc (2), 273-279, 1973.

[12] R. D. Schafer, *Structure of genetic algebras*, American Journal of Mathematics, **71** (1949), 121-135.

[13] P. Holgate, *Sequences of powers in genetic algebras*, J. London Math., **42** (1967), 489-496.

[14] P. Holgate, *Selfing in gentic algebras*, J. Math. Biology, **6** (1978), 197-206.

[15] I. Hench, *Sequences in genetic algebras for overlapping generations*, Proc. Edinburgh Math. Soc. (2) **18** (1972), 19-29.

[16] O. Reiser, *Genetic algebras studied recursively and by means of differential operators*, Math. Scand. **10** (1962), 25-44.

[17] V.M. Abraham, *Linearising quadratic transformations in genetic algebras*, Thesis, Univ. of London, 1976.

[18] N.J.A. Sloane & A.D. Wyner (editors), Claude Elwood Shannon Collected Papers, IEEE Press, Piscataway, NJ. 1993.

[19] C.W. Cotterma, A calculus for statistico-genetics, Dissertation, The Ohio State University, Columbus, OH. 1940.

[20] P. Ballonoff, Genetics and Social Structure, Dowden, Hutchinson & Ross, Stroudsburg, PA. 1974.

[21] A. Worz-Busekros, Algebras in Genetics, Lecture Notes in Biomath, Vol. 36, Springer-Verlag, Berlin, 1980.

[22] Y.I. Lyubich, Mathematical Structures in Population Genetics, Springer-Verlag, New York, 1992.

[23] Y.I. Lyubich, *A topological approach to a problem in mathematical genetics*, Russian Math. Surveys, **36**:6 (1979), 60-66.

[24] M.L. Reed, *Algebraic structure of genetic inheritance*, Bull. of AMS, 34, **(**2), 107-130, 1997.

[25] J.I. Ljubic, *Algebraic methods in evolutionary genetics*, Biometrical Journal, **20**:5 (2007), 511-529.

[26] P. Fortini & R. Barakat, *Genetic algebras for tetraploidy with several loci*, Journal of Mathematical Biology, **9** (1980), 297-304.

[27] T.M.M. Campos & P. Holgate, *Algebraic isotopy in genetics*, Mathematical Medicine and Biology, **4**:3 (1987), 215-222.

[28] H. Gonshor, *Derivations in genetic algebras*, Communications in Algebra, **16**:8 (1998), 1525-1542.

[29] M. Bremner, *Polynomial identities for Bernstein algebras of simple Mendelian inheritance*, to appear in Communications in Algebra.

[30] I. Heuch, *Genetic algebras and time continuous models*, Theoretical population biology, **4** (1973), 133-144.

[31] H. Gradl, *Long-time behavior of continuous time models in genetic algebras*, Journal of Mathematical Biology, **32** (1994), 269-274.

[32] E. Baur, Zeit. Vererbungsl, **1**, 330-351, 1909.

[33] C. Correns, Zeit. Vererbungsl, **1**, 291-329, 1909.

[34] J. E. Graver & M. E. Watkins, Combinatorics with Emphasis on the Theory of Graphs, Springer-Verlag, New York, 1977.

[35] C.W.Jr. Birky, *The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models*, Annu. Rev. Genet. **35**, 125-148, 2001.

[36] C.W.Jr. Birky, *Inheritance of mitochondrial mutations*, Mitochondrial DNA Mutations and Aging, Disease and Cancer, K.K. Singh, edit, Spring, 1998

[37] N.W. Gillham, Organelle Genes and Genomes, Oxford University Press, 1994

[38] C.F. Emmerson, G.K. Brown & J. Poulton, *Synthesis of mitochondrial DNA in permeabilised human cultured cells*, Nucleic Acids Res. 29, **2**, 2001.

[39] F. Ling & T. Shibata, *Mhr1p-dependent concatemeric mitochondrial DNA formation for generating yeast mitochondrial homoplasmic cells*, Mol. Biol. Cell, vol.15, 310-322, Jan. 2004.

[40] Y. Tang, G. Manfredi, M. Hirano & E.A. Schon, *Maintenance of human rearranged mitochondrial DNAs in long-term transmitochondrial cell lines*, Mol. Biol. Cell, vol.11, 2349-2358, Jul. 2000.

[41] F.M.A. Samen, G.A. Secor, & N.C. Gudmestad, *Variability in virulence among asexual progenies of Phytophthora infestans*, Phytopathology, **93**, 293-304, 2003.

[42] W.E. Fry & S.E. Goodwin, *Re-emergence of potato and tomato late blight in the United States*, Plant Disease, 1349-1357, dec. 1997.

[43] S.B. Goodwin, B.A. Cohen & W.E. Fry, *Panglobal distribution of a single clonal lineage of the Irish potato famine fungus*, Proc. Natl. Acad. Sci. USA, **91**, 11591-11595, 1994.

[44] S.B. Goodwin, *The population genetics of Phytophthora*, Phytopathology, **87**, 462-473, 1997.

[45] S.-X. Lin, F. Tian, & Z. Wang, *Burau representation and random walks on string links*, Pacific J of Mathematics, Vol. 182 (1998), No. 2, 289-302.

[46] S.-X. Lin & Z. Wang, *Random walk on knot diagrams, colored jones polynomial and ihara-selberg zeta function*, in Knots, Braids, and Mapping Class Groups–Papers

Dedicated to Joan S. Birman: Proceedings of a Conference on Low Dimensional Topology in Honor of Joan S. Birman's 70th Birthday, 1998, Columbia University, New York, pp.107-121.

[47] D. Foata & D. Zeilberger, *A combinatorial proof of Bass' evaluations of the Ihara-Selberg zeta function for graphs*, Trans. Amer. Math. Soc. 351 (1999), 2257-2274.

[48] V.G. Kac, Infinite Dimensional Lie Algebras, Cambridge University Press, 1983.

[49] J.G. Kemeny, J. Snell, L. Knapp, W. Anthony, Denumerable Markov Chains, Springer-Verlag, 1976.

[50] M. Lindner, Infinite Matrices and Their Finite Sections: an Introduction to the Limit Operator Method, Birkhauser Verlag, 2000.

[51] C. Swartz, Infinite Matrices and the Gliding Hump, World Scientific, 1996

[52] V.S. Sunder, An Invitation to von Neumann Algebras, Springer-Verlag, 1987

[53] A. Micali & F. Zitan, textitOn homogeneous weighted algebras, Communications in Algebra, **35** (2007), 2371-2385.