Lecture Notes in Mathematics

1921

Jianjun Paul Tian

Evolution Algebras and their Applications

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

Springer

Jianjun Paul Tian

Evolution Algebras and their Applications

Springer

Author

Jianjun Paul Tian
Mathematical Biosciences Institute
The Ohio State University
231 West 18th Avenue
Columbus, OH 43210-1292
USA

Contact after August 2007

Mathematics Department
College of William and Mary
P. O. Box 8795
Williamsburg VA 23187-8795
USA

e-mail: jtian@wm.edu

Library of Congress Control Number: 2007933498

Mathematics Subject Classification (2000): 08C92, 17D92, 60J10, 92B05, 05C62, 16G99

ISSN print edition: 0075-8434
ISSN electronic edition: 1617-9692
ISBN 978-3-540-74283-8 Springer Berlin Heidelberg New York
DOI 10.1007/978-3-540-74284-5

Springer is a part of Springer Science+Business Media
springer.com

Typesetting by the author and SPi using a Springer LaTeX macro package

Cover design: *design & production* GmbH, Heidelberg

Printed on acid-free paper SPIN: 12109265 41/SPi 5 4 3 2 1 0

To my parents

Bi-Yuan Tian and Yu-Mei Liu

My father, the only person I know who can operate two abaci
using his left and right hand simultaneously in his business.

Preface

In this book, we introduce a new type of algebra, which we call evolution algebras. These are algebras in which the multiplication tables are of a special type. They are motivated by evolution laws of genetics. We view alleles (or organelles or cells, etc,) as generators of algebras. Therefore we define the multiplication of two "alleles" G_i and G_j by $G_i \cdot G_j = 0$ if $i \neq j$. However, $G_i \cdot G_i$ is viewed as "self-reproduction," so that $G_i \cdot G_i = \sum_j p_{ij} G_j$, where the summation is taken over all generators G_j. Thus, reproduction in genetics is represented by multiplication in algebra. It seems obvious that this type of algebra is nonassociative, but commutative. When the p_{ij}s form Markovian transition probabilities, the properties of algebras are associated with properties of Markov chains. Markov chains allow us to develop an algebra theory at deeper hierarchical levels than standard algebras. After we introduce several new algebraic concepts, particularly algebraic persistency, algebraic transiency, algebraic periodicity, and their relative versions, we establish hierarchical structures for evolution algebras in Chapter 3. The analysis developed in this book, particularly in Chapter 4, enables us to take a new perspective on Markov process theory and to derive new algebraic properties for Markov chains at the same time. We see that any Markov chain has a dynamical hierarchy and a probabilistic flow that is moving with invariance through this hierarchy. We also see that Markov chains can be classified by the skeleton-shape classification of their evolution algebras. Remarkably, when applied to non-Mendelian genetics, particularly organelle heredity, evolution algebras can explain establishment of homoplasmy from heteroplasmic cell population and the coexistence of mitochondrial triplasmy, and can also predict all possible mechanisms to establish the homoplasmy of cell population. Actually, these mechanisms are hypothetical mechanisms in current mitochondrial disease research. By using evolution algebras, it is easy to identify different genetic patterns from the complexity of the progenies of *Phytophthora infectans* that cause the late blight of potatoes and tomatoes. 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 provide a theoretical framework to unify many phenomena. Among the further research topics related to evolution algebras and other fields, the most significant topic perhaps is to develop a continuous evolution algebra theory for continuous time dynamical systems.

The intended audience of this book includes graduate students and researchers with interest in theoretical biology, genetics, Markov processes, graph theory, and nonassociative algebras and their applications.

Professor Jean-Michel Morel gave me a lot of support and encouragement, which enabled me to take the step to publish my research results as a book. Other editors and staff in LNM made efforts to find reviewers and edit my book. Here, I wish to express my great thanks to them.

I thank Professor Michael T. Clegg for his stimulating problems in coalescent theory. From that point, I began to study genetics and stochastic processes. I am greatly indebted to Professor Xiao-Song Lin, my Ph.D advisor, for his valuable advice and long-time guidance. I am thankful to professors Bai-Lian Larry Li, Michel L. Lapidus, and Barry Arnold for their valuable suggestions. It gives me great pleasure to thank Professors Bun Wong, Yat Sun Poon, Shizhong Xu, Keh-Shin Lii, Peter March, Dennis Pearl, Raymond L. Orbach, Murray Bremner, Yuan Lou, and Yang Kuang for their encouragement. I also thank Professor C. William Birky Jr. for his explanation of non-Mendelian genetics through e-mails. I acknowledge Professor Winfried Just for his suggestions of writing style of the book and a formula in Chapter 3. I am grateful to my current mentor, Professor Avner Friedman, for his detailed and cherished suggestions on the research in this book and my other research directions. I thank three reviewers for their suggestions and constructive comments.

Last, but not the least, I thank Dr. Shannon L. LaDeau for her help on English of the book. I also thank my wife, Yanjun Sophia Li, for her support and love. I acknowledge the support from the National Science Foundation upon agreement No. 0112050.

Mathematical Biosciences Institute, Ohio *Jianjun Paul Tian*
April, 2007

Contents

1

Introduction

While 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, the reproduction of bacteria involved by bacteriophages, asexual reproduction or generally non-Mendelian inheritance, and Markov chains. Therefore, we defined it as a type of new algebra — the evolution algebra. Evolution algebras are nonassociative and non-power-associative Banach algebras. Indeed, they are natural examples of nonassociative complete normed algebras arising from science. It turns out that these algebras have many unique properties, and also have connections with other fields of mathematics, including graph theory (particularly, random graphs and networks), group theory, Markov processes, dynamical systems, knot theory, 3−manifolds, and the study of the Riemann-zeta function (or a version of it called the Ihara-Selberg zeta function). One of the unusual features of evolution algebras is that they possess an evolution operator. This evolution operator reveals the dynamical information of evolution algebras. However, what makes the theory of evolution algebras different from the classical theory of algebras is that in evolution algebras, we can have 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 on 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 semidirect-sum decompositions of a general evolution algebra. Thus, this hierarchical structure demonstrates that an evolution algebra is a mixed algebraic and dynamical subject. The algebraic nature of this hierarchical structure allows us to have a rough skeleton-shape classification of evolution algebras. At the same time, the dynamical 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 algebra induced by a Markov chain, it is easy to see that the Markov chain has a dynamical hierarchy and the probabilistic flow is moving with invariance through this hierarchy, and that all 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 the understanding of non-Mendelian genetics. For instance, once we apply evolution algebra theory to the inheritance of organelle genes, we can predict all possible mechanisms to establish the homoplasmy of cell populations. Actually, these mechanisms are hypothetical mechanisms in current mitochondrial research. Using our algebra theory, it is also easy to understand the coexistence of triplasmy in tissues of sporadic mitochondrial disorder patients. Further more, once the algebraic structure of asexual progenies of *Phytophthora infectans* is obtained, we can make certain important predictions and suggestions to plant pathologists.

In history, mathematicians and geneticists once used nonassociative algebras to study Mendelian genetics. Mendel [30] first exploited symbols that are quite algebraically suggestive to express his genetic laws. In fact, it was later termed "Mendelian algebras" by several other authors. In the 1920s and 1930s, general genetic algebras were introduced. Apparently, Serebrowsky [31] was the first to give an algebraic interpretation of the sign "×", which indicated sexual reproduction, and to give a mathematical formulation of Mendel's laws. Glivenkov [32] introduced the so-called Mendelian algebras for diploid populations with one locus or two unlinked loci. Independently, Kostitzin [33] also introduced a "symbolic multiplication" to express Mendel's laws. The systematic study of algebras occurring in genetics can be attributed to I. M. H. Etherington. In his series of papers [34], he succeeded in giving a precise mathematical formulation of Mendel's laws in terms of nonassociative algebras. Besides Etherington, fundamental contributions have been made by Gonshor [35], Schafer [36], Holgate [37, 38], Hench [39], Reiser [40], Abraham [41], Lyubich [47], and Worz-Busekos [46]. It is worth mentioning two unpublished work in the field. One is the Ph.D. thesis of Claude Shannon, the founder of modern information theory, which was submitted in 1940 (The Massachusetts Institute of Technology) [43]. 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 (The Ohio State University) [44] [45]. 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 general 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 nonassociative 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 only are functions of the images under these representations, are new concepts for mathematicians. Until 1980s, the most comprehensive reference in this area was Worz-Busekros's book [46]. More recent results, such as genetic evolution in genetic algebras, can be found in Lyubich's book [47]. A good survey is Reed's article [48].

General genetic algebras are the product of interaction between biology and mathematics. Mendelian genetics introduced a new subject to mathematics: general 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 [49] and Correns [50] first detected that chloroplast inheritance departed from Mendel's rules, and much later, mitochondrial gene inheritance was also identified in the same way, and non-Mendelian inheritance of organelle genes was recognized with two features — uniparental inheritance and vegetative segregation. Now, non-Mendelian genetics is a basic language of molecular geneticists. Logically, we can ask what non-Mendelian genetics offers to mathematics. The answer is "evolution algebras" [24].

The purpose of the present book is to establish the foundation of the framework of evolution algebra theory and to discuss some applications of evolution algebras in stochastic processes and genetics. Obviously, we are just opening a door to a new subject of the mixture of algebras and dynamics and to the many new research topics that are confronting us. To promote further research in this subject, we include many specific research topics and open problems at the end of this book. Now, I would like to briefly introduce the content contained in each chapter of the book.

In Chapter 2, we introduce the motivations behind the study of evolution algebras from the perspective of three different sciences: biology, physics, and mathematics. We observe phenomena of uniparental inheritance and the reproduction of bacteria involved by bacteriophages; we also analyze the neutral Wright-Fisher model for a haploid population in population genetics. We study motions of particles in a space and discrete flows in a discrete space, and we also observe reactions among particles in general physics. We mention some research in knot theory where negative probabilities are involved. We analyze and view a Markov chain as a discrete time dynamical system. All these phenomena suggest a common and intrinsic algebraic structure, which we define in chapter 3 as evolution algebras.

In Chapter 3, evolution algebras are defined; their basic properties are investigated and the principal theorem about evolution algebras — the hierarchical structure theorem — is established. We define evolution algebras in terms of generators and defining relations. Because the defining relations are unique for an evolution algebra, the generator set can serve as a basis for an evolution algebra. This property gives some advantage in studying evolution algebras. The basic algebraic properties of evolution algebras, such as nonassociativity and nonpower-associativity are studied. Various algebraic concepts in evolution algebras are also investigated, such as evolution subalgebras, the associative multiplication algebra of an evolution algebra, the centroid of an evolution algebra and, the derived Lie algebra of an evolution algebra. 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. From the viewpoint of dynamical systems, we introduce an evolution operator for an evolution algebra that is actually a special right (left) multiplication operator. This evolution operator reveals the dynamical information of an evolution algebra. To describe the evolution flow quantitatively, we introduce a norm for an evolution algebra. Under this norm, an evolution algebra becomes a Banach algebra. As we have mentioned above, what makes the evolution algebra theory different from the classical algebra theory is that in evolution algebras we can have 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 dynamical flow as it displays in the hierarchy of an evolution algebra. The remarkable property of an evolution algebra is its hierarchical structure, which gives a picture of a dynamical process when one takes multiplication in an evolution algebra as time-step in a discrete-time dynamical system. Algebraically, this hierarchy is a sequence of semidirect-sum decompositions of a general evolution algebra. It depends upon the "relative" concepts of algebraic persistency and algebraic transiency. By "relative" concepts, we mean that concepts of higher level algebraic persistency and algebraic transiency are defined over the space generated by transient generators in the previous level. The difference between algebraic persistency and algebraic transiency suggests a sequence of the semidirect-sum decompositions, or suggests a direction of the evolution from the viewpoint of dynamical systems. This hierarchical structure demonstrates that an evolution algebra is a mixed subject of algebras and dynamics. We also obtain the structure theorem for a simple evolution algebra. We give a way to reduce a "big" evolution algebra to a "small" one that still has the same hierarchy as that of the original algebra. We call it the reducibility. This reducibility gives a rough classification, the skeleton-shape classification, of all evolution algebras.

To demonstrate the importance and the applicability of the abstract subject — evolution algebras — we study a type of evolution algebra that corresponds to or is determined by a Markov chain in Chapter 4. We see that any general Markov chain has a dynamical hierarchy and the probabilistic flow is moving with invariance through this hierarchy, and that all Markov chains can be classified by the skeleton-shape classification of their evolution algebras. When a Markov chain is viewed as a dynamical system, there should be a certain mechanism behind the Markov chain. We view this mechanism as a "reproduction process." But it is a very special case of reproduction process. Each state can just "cross" with itself, and different states cannot cross, or they cross to produce nothing. We introduce a multiplication for this reproduction process. Thus an evolution algebra is defined by using transition probabilities of a Markov chain as structural constants. In evolution algebras, the Chapman-Kolmogorov equations can be simply viewed as a composition of evolution operators or the principal power of a special element. By using evolution algebras, one can see algebraic properties of Markov chains. 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. Generally, 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 determined by a Markov chain is finite, algebraic concepts (algebraic persistency and algebraic transiency) and analytic concepts (probabilistic persistency and probabilistic transiency) are equivalent. We also study the spectrum theory of the evolution algebra M_X determined by a Markov chain X. Although the dynamical behavior of an evolution algebra is embodied by various powers of its elements, the evolution operator seems to represent a "total" principal power. From the algebraic viewpoint, we study the spectrum of evolution operators. Particularly, the evolution operator is studied at the $0th$ level in the hierarchy of an evolution algebra. For example, for a finite dimension evolution algebra the geometric multiplicity of the eigenvalue 1 of the evolution operator is equal to the number of the $0th$ simple evolution subalgebras. The spectrum structure at higher level is an interesting further research topic. Another possible spectrum theory could be the study of plenary powers. Actually, we have already defined the plenary power for a matrix. It could give a way to study this possible spectrum theory. Any general Markov chain has a dynamical hierarchy, which can be obtained from its corresponding evolution algebra. We give a description of probability flows on its hierarchy. We also give the sojourn times during each simple evolution subalgebra at each level on the hierarchy. By using the skeleton-shape classification of evolution algebras, we can reduce a bigger Markov chain to a smaller one that still possesses the same dynamical behavior as the original chain does. We have also obtained a new skeleton-shape classification theorem for general Markov chains. Thus, from the evolution algebra theory, algebraic properties about general Markov chains are revealed. In the last section of this chapter, we discuss examples and applications, and show algebraic versions of Markov chains, evolution algebras, also have advantages in computation of Markov processes.

We begin to apply evolution algebra theory to biology in Chapter 5. We first introduce the basic biology of non-Mendelian genetics including organelle population genetics and *Phytophthora infectans* population genetics.

We then give a general algebraic formulation of non-Mendelian inheritance. To understand a puzzling feature of organelle heredity, that is that heteroplasmic cells eventually disappear and the homoplasmic progenies are observed, we construct relevant evolution algebras. We then can predict all possible mechanisms to establish the homoplasmy of cell populations, which actually are hypothetical mechanisms in current mitochondrial research [55]. Theoretically, we can discuss any number of mitochondrial mutations and study their genetic dynamics by using evolution algebras. Remarkably, experimental biologists have observed the coexistence of the triplasmy (partial duplication of mt-DNAs, deletion of mt-DNAs, and wild-type mt-DNAs) in tissues of patients with sporadic mitochondrial disorders. While doctors and biologists cultured cell lines to study the dynamical relations among these mutants of mitochondria, our algebra model could be used to predict the outcomes of their cell line cultures. We show that concepts of algebraic transiency and algebraic persistency catch the essences of biological transitory and biological stability. Moreover, we could predict some transition phases of mutations that are difficult to observe in experiments. We also study another type of uniparental inheritance about *Phytophthora infectans* that cause late blight of potatoes and tomatoes. After constructing several relevant evolution algebras for the progeny populations of *Phytophthora infectans*, we can see different genetically dynamical patterns from the complexity of the progenies of *Phytophthora infectans*. We then predict the existence of intermediate transient races and the periodicity of reproduction of biological stable races. Practically, we can help farmers to prevent spread of late blight disease. Theoretically, we can use evolution algebras to provide information on *Phytophthora infectans* reproduction rates for plant pathologists.

As we mentioned above, evolution algebras have many connections with other fields of mathematics. Using evolution algebras it is expected that we will be able to see problems in many mathematical fields from a new perspective. We have already finished some of the basic study. Most of the research will be very interesting and promising both in theory and in application. To promote better understanding and further research in evolution algebras, in Chapter 6, we list some of the related results we have obtained and put forward further research topics and open problems. For example, we obtain a theorem of classification of directed graphs. We also post a series of open problems about evolution algebras and graph theory. Because evolution algebras hold the intrinsic and coherent relation with graph theory, we will be able to analyze graphs algebraically. The purpose of this is that we try to establish a brand new theory "algebraic graph theory" to reach the goal of Gian-Carlo Rota — "Combinatorics needs fewer theorems and more theory" [29]. On the other hand, it is also expected that graph theory can be used as a tool to study nonassociative algebras. Some research topics in evolution algebras and group theory, knot theory, and Ihara-Selberg zeta function, which we post as further research topics, are also very interesting. Perhaps, the most significant topic is to develop a continuous evolution algebra theory for continuous time dynamical systems. It is also important to use evolution algebras to develop algebraic statistical physics models. In this direction, the big picture in our mind is to describe the general interaction of particles. This means any two generators can multiply and do not vanish when they are different. This involves an operation, multiplication, of three-dimensional matrices. Some preliminary results have already been obtained in this direction. We are also interested in questions such as how evolution algebras reflect properties of a 3-manifold where a particle moves when the recording time period is taken as an infinite sequence, and what new results about the 3-manifold can be obtained by the sequence of evolution algebras, etc.

We give a list of background literature in the last section, though the directly related literature is sparse.