Advances in Mathematics 315 (2017) 285–323

Contents lists available at ScienceDirect

Advances in Mathematics

www.elsevier.com/locate/aim

Local equations for equivariant evolutionary models

Marta Casanellas [a], Jesús Fernández-Sánchez [a,*],
Mateusz Michałek [b]

[a] Dept. Matemàtiques, Universitat Politècnica de Catalunya, Av. Diagonal, 647, 08028 Barcelona, Spain
[b] Polish Academy of Sciences, ul. Śniadeckich 8, 00956 Warsaw, Poland

ARTICLE INFO

Article history:
Received 18 January 2016
Received in revised form 5 May 2017
Accepted 15 May 2017
Available online 13 June 2017
Communicated by Andreas Dress

MSC:
92D15
14H10
60J20

Keywords:
Evolutionary model
Complete intersection
Phylogenetic variety
Phylogenetic tree
Representation theory

ABSTRACT

Phylogenetic varieties related to equivariant substitution models have been studied largely in the last years. One of the main objectives has been finding a set of generators of the ideal of these varieties, but this has not yet been achieved in some cases (for example, for the general Markov model this involves the open "salmon conjecture", see [2]) and it is not clear how to use all generators in practice. Motivated by applications in biology, we tackle the problem from another point of view. The elements of the ideal that could be useful for applications in phylogenetics only need to describe the variety around certain *points of no evolution* (see [13]). We produce a collection of explicit equations that describe the variety on a Zariski open neighborhood of these points (see Theorem 5.4). Namely, for any tree T on any number of leaves (and any degrees at the interior nodes) and for any equivariant model on any set of states κ, we compute the codimension of the corresponding phylogenetic variety. We prove that this variety is smooth at general points of no evolution and, if a mild technical condition is satisfied ("d-claw tree hypothesis"), we

---

[*] Corresponding author.
E-mail addresses: marta.casanellas@upc.edu (M. Casanellas), jesus.fernandez.sanchez@upc.edu (J. Fernández-Sánchez), mateusz.michalek@mis.mpg.de (M. Michałek).

http://dx.doi.org/10.1016/j.aim.2017.05.003

provide an algorithm to produce a complete intersection that describes the variety around these points.

1. Introduction

In the recent years there has been a huge amount of work done on phylogenetic varieties – we advise the reader to consult e.g. [4,7,12,28,40,45] and references therein. These algebraic varieties contain the set of joint distributions at the leaves of a tree evolving under a Markov model of molecular evolution. From the biological point of view, these varieties are interesting because they provide new tools of non-parametric inference of phylogenetic trees. At present, the algebraic/geometric framework of phylogenetic varieties has allowed proving the identifiability of parameters of certain evolutionary models widely used by biologists [8,3], proposing new methods of model selection [35], and producing new phylogenetic reconstruction methods [29,21,15].

From the mathematical point of view, there has been a great effort in finding a whole description of the ideal of these phylogenetic varieties [6,45,26,24]. Still, for some models, many questions remain open for trees on an arbitrary number of leaves n. Indeed, if one is interested in using these algebraic tools with real data, one would need a small set of generators of the ideal (rather than a description of the whole ideal); it would also be desirable to know the degree at which the ideal is generated [2,25,41,45]; and as the codimension of the variety is exponential in n, it is necessary to distinguish between generators that only account for the underlying evolutionary model, those that account for the tree topology, and those that could be useful for inferring the numerical parameters (see [7] for a good introduction to this topic).

For instance, the authors of [45] and [7] raised the question whether knowing a complete intersection containing the phylogenetic variety and of the same dimension would be enough. Eminently, for biological applications it is only relevant to know the description of the variety around the points that make sense *biologically* speaking. If these points are smooth, then a complete intersection can define the variety on a neighborhood of these points.

This is the approach that was considered in [13,33,42] for the particular model of Kimura 3-parameter and is the same goal that we pursued in [18] for abelian group-based models. In the present paper we address this problem for the more general class of *G-equivariant models*, which contains more general algebraic models of interest to biologists (for example the strand-symmetric and the general Markov models). We give an explicit algorithm to construct a complete intersection that describes the variety on a dense open subset around a generic *point of no evolution*. Points of no evolution represent molecular sequences that remain invariant from the common ancestor to the leaves of the tree. Points in the phylogenetic variety that arise from biologically meaningful