

## University of Tasmania Open Access Repository

### Cover sheet

**Title**

Entanglement, invariants, and phylogenetics

**Author**

Sumner, JG

**Bibliographic citation**

Sumner, JG (2006). Entanglement, invariants, and phylogenetics. University Of Tasmania. Thesis.  
<https://doi.org/10.25959/23212172.v1>

Is published in:

**Copyright information**

This version of work is made accessible in the repository with the permission of the copyright holder/s under the following,

**Licence.**

If you believe that this work infringes copyright, please email details to: [oa.repository@utas.edu.au](mailto:oa.repository@utas.edu.au)

Downloaded from University of Tasmania Open Access Repository

Please do not remove this coversheet as it contains citation and copyright information.

University of Tasmania Open Access Repository

Library and Cultural Collections

University of Tasmania

Private Bag 3

Hobart, TAS 7005 Australia

E [oa.repository@utas.edu.au](mailto:oa.repository@utas.edu.au)

CRICOS Provider Code 00586B | ABN 30 764 374 782

[utas.edu.au](http://utas.edu.au)

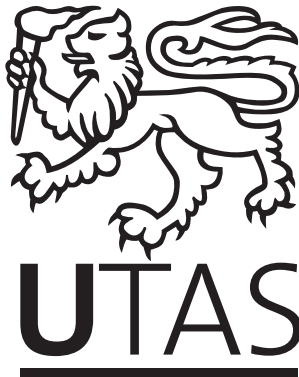
# ENTANGLEMENT, INVARIANTS, AND PHYLOGENETICS

by

Jeremy G Sumner, B.Sc. Hons (Tas)

Submitted in fulfilment of the requirements  
for the Degree of Doctor of Philosophy

School of Mathematics and Physics  
University of Tasmania  
December, 2006



I declare that this thesis contains no material which has been accepted for a degree or diploma by the University or any other institution, except by way of background information and duly acknowledged in the thesis, and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due acknowledgement is made in the text of the thesis.

Signed: \_\_\_\_\_  
Jeremy G Sumner

Date: \_\_\_\_\_

This thesis may be made available for loan and limited copying in accordance with the *Copyright Act 1968*.

Signed: \_\_\_\_\_  
Jeremy G Sumner

Date: \_\_\_\_\_

The following people contributed to the publication of work undertaken as part of this thesis.

*Entanglement invariants and phylogenetic branching* [59].  
Jeremy G Sumner (75%), Peter D Jarvis (25%).

*Using the tangle: a consistent construction of phylogenetic distance matrices* [60].  
Jeremy G Sumner (80%), Peter D Jarvis (20%).

**We the undersigned agree with the above stated proportion of work undertaken for each of the above published (or submitted) peer-reviewed manuscripts contributing to this thesis.**

Signed: \_\_\_\_\_  
Peter D Jarvis  
Supervisor  
School of Mathematics and Physics  
University of Tasmania

Date: \_\_\_\_\_

Signed: \_\_\_\_\_  
Larry Forbes  
Head of School  
School of Mathematics and Physics  
University of Tasmania

Date: \_\_\_\_\_

# ABSTRACT

This thesis develops and expands upon known techniques of mathematical physics relevant to the analysis of the popular Markov model of phylogenetic trees required in biology to reconstruct the evolutionary relationships of taxonomic units from biomolecular sequence data.

The techniques of mathematical physics are plethora and have been developed for some time. The Markov model of phylogenetics and its analysis is a relatively new technique where most progress to date has been achieved by using discrete mathematics. This thesis takes a group theoretical approach to the problem by beginning with a remarkable mathematical parallel to the process of scattering in particle physics. This is shown to equate to branching events in the evolutionary history of molecular units. The major technical result of this thesis is the derivation of existence proofs and computational techniques for calculating polynomial group invariant functions on a multi-linear space where the group action is that relevant to a Markovian time evolution. The practical results of this thesis are an extended analysis of the use of invariant functions in distance based methods and the presentation of a new reconstruction technique for quartet trees which is consistent with the most general Markov model of sequence evolution.

# ACKNOWLEDGEMENTS

First and foremost my thanks go to my supervisor Peter Jarvis. Not only for having the insight to take on this novel work and his outstanding knowledge of mathematical physics, but also for being a true friend and good bloke.

These people have all played their own special role in bringing this thesis to fruition: Michael Sumner, Robert Delbourgo, Patrick McLean; William Joyce and the Physics department of the University of Canterbury; Mike Steel and the organisers of the New Zealand phylogenetics meeting; Rex Lau, Lars Jermin, Michael Charleston and SUBIT; Alexei Drummond; Simon Wotherspoon (for giving me such a hard time), Malgorzata O'Reilly, Jim Bashford, Giuseppe Cimo, Stuart Morgan, Isamu Imahori and Graham Legg; Mum, Dad and Kate; Keith, Tim, Sarah, Wazza and Beans.

A special mention for my high school maths teacher Mr. Rush, who used to laugh when I continually interrupted his classes with: "That's all very well, Mr. Rush, but how is this going to help me lay bricks?"

*Here, as it draws to its last Halt, if anywhere, might both Gentlemen take joy of a brief Holiday from Reason. Yet, “Too busy,” Mason insists, and “Far too cheerful for thah’,” supposes Dixon.*

**Mason and Dixon**

Thomas Pynchon



# TABLE OF CONTENTS

<b>TABLE OF CONTENTS</b>	<b>i</b>
<b>LIST OF TABLES</b>	<b>iv</b>
<b>LIST OF FIGURES</b>	<b>v</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Mathematical background</b>	<b>5</b>
2.1 Group representations . . . . .	5
2.1.1 Group characters . . . . .	7
2.1.2 Tensor product . . . . .	8
2.1.3 Group action on a tensor product space . . . . .	10
2.2 Irreducible representations of the general linear group . . . . .	11
2.2.1 Partitions . . . . .	11
2.2.2 The Schur functions . . . . .	12
2.2.3 Group characters of $GL(n)$ . . . . .	13
2.2.4 The Schur/Weyl duality . . . . .	14
2.2.5 More representations . . . . .	16
2.2.6 One-dimensional representations . . . . .	17
2.3 Invariant theory . . . . .	18
2.3.1 Invariants as irreducible representations . . . . .	19
2.3.2 Using Schur functions to count invariants . . . . .	20
2.4 Invariants of the general linear group . . . . .	21
2.4.1 Invariants of $GL(n)$ on $V^{\otimes m}$ . . . . .	22
2.4.2 Invariants of $\times^m GL(n)$ on $V^{\otimes m}$ . . . . .	24
2.5 Closing remarks . . . . .	27

<b>3</b>	<b>Entanglement and phylogenetics</b>	<b>28</b>
3.1	Quantum mechanics . . . . .	29
3.1.1	Spin $\frac{1}{2}$ and entanglement . . . . .	30
3.1.2	Orbit classes and invariants . . . . .	33
3.1.3	Two qubits and the concurrence . . . . .	34
3.1.4	Three qubits and the tangle . . . . .	35
3.2	Stochastic evolution of biomolecular units . . . . .	37
3.3	Phylogenetic trees . . . . .	38
3.4	Tensor presentation . . . . .	39
3.5	Entanglement and phylogenetics . . . . .	42
3.5.1	Two qubits . . . . .	42
3.5.2	Three qubits . . . . .	43
3.5.3	Phylogenetic relation . . . . .	44
3.6	Closing remarks . . . . .	45
<b>4</b>	<b>Using the tangle</b>	<b>46</b>
4.0.1	Stochastic distance . . . . .	48
4.0.2	Observability of the stochastic distance . . . . .	49
4.1	Pairwise distance measures . . . . .	49
4.1.1	The log det formula . . . . .	50
4.1.2	The tangle . . . . .	52
4.1.3	Star topology . . . . .	54
4.1.4	Summary . . . . .	54
4.2	Generalized pulley principle . . . . .	55
4.2.1	Interpretation . . . . .	59
4.3	The quartet case . . . . .	60
4.4	Closing remarks . . . . .	63
<b>5</b>	<b>Markov invariants</b>	<b>64</b>
5.1	The Markov semigroup . . . . .	64
5.1.1	Invariant functions of the Markov semigroup . . . . .	65
5.2	Alternative computation of invariants of the general linear group	67
5.2.1	Action of $GL(n)$ on $V^{\otimes m}$ . . . . .	68
5.2.2	Examples . . . . .	69

5.2.3	Action of $\times^m GL(n)$ on $V^{\otimes m}$ . . . . .	71
5.3	Computation of the Markov invariants . . . . .	71
5.3.1	Markov invariants of $\mathcal{M}(n)$ on $V^{\otimes m}$ . . . . .	72
5.3.2	Examples . . . . .	72
5.4	Markov invariants of $\times^m \mathcal{M}(n)$ on $V^{\otimes m}$ . . . . .	75
5.4.1	The stochastic invariant . . . . .	76
5.4.2	The $n=2$ case . . . . .	76
5.4.3	The $n=3$ case . . . . .	77
5.4.4	The $n=4$ case . . . . .	78
5.5	What happens on a phylogenetic tree? . . . . .	79
5.5.1	The stangle . . . . .	79
5.5.2	The squangles . . . . .	80
5.6	Review of important invariants . . . . .	81
5.7	Closing remarks . . . . .	81
<b>6</b>	<b>CONCLUSION</b>	<b>83</b>
<b>A</b>	<b>Bias correction of invariant functions</b>	<b>85</b>
A.1	Multinomial distribution . . . . .	85
A.2	Generating function . . . . .	85
A.3	Expectations of polynomials . . . . .	86
A.4	Bias correction . . . . .	87
	<b>BIBLIOGRAPHY</b>	<b>89</b>

# LIST OF TABLES

5.1	Occurrences of $\{d\}$ in $*^m\{k+s, k^{n-1}\}$ with $nk+s=d$ . . . . .	75
5.2	Invariant functions satisfying $f \circ g = \det(g)^k f$ . . . . .	82

# LIST OF FIGURES

2.1	Semi-standard tableaux . . . . .	13
3.1	Phylogenetic tree of four taxa . . . . .	39
3.2	Phylogenetic tree with two leaves . . . . .	42
3.3	Phylogenetic tree with three leaves . . . . .	44
4.1	Phylogenetic tree of two taxa . . . . .	50
4.2	Phylogenetic tree of three taxa . . . . .	52
4.3	Using the generalized pulley principle . . . . .	60
4.4	Four taxa tree with alternative roots . . . . .	61
4.5	Three taxon subtrees . . . . .	62
5.1	Three alternative quartet trees . . . . .	81

# CHAPTER 1

## Introduction

The rationale of this thesis is taken from a remarkable analogy between the stochastic models used to infer phylogenetic relationships in mathematical biology and the structure of multiparticle quantum physics. There is a direct relationship between Feynman diagrams that describe the interactions of sub-atomic particles and phylogenetic trees that graphically represent the evolutionary relationship between taxonomic units. A Feynman diagram gives the graphical representation of creation and annihilation events of particle interactions. A taxonomic unit may be any biomolecular unit such as a gene, an amino acid or base pair, and the time evolution of these molecular units is modelled stochastically under a Markov assumption. Techniques which reconstruct the evolutionary history of molecular units from present observations are based on these models. Given the correct framework, these Markov models and the formalism of multiparticle quantum mechanics can be put into a mathematical correspondence. This is a very useful observation because phylogenetics is a relatively new mathematical problem (for example see the classic paper by Felsenstein [19]) whereas the mathematics of particle physics has been studied for over a century. (For an outstanding introduction to the history of theoretical particle physics see [47], and for a comprehensive introduction to mathematical physics see [61].) Given that there is a mathematical connection between the two problems it would certainly be unfortunate to see results that have been obtained in physics re-derived independently in the context of phylogenetics. This thesis looks at a particular aspect of quantum systems known as *entanglement* and shows that measures of entanglement can be utilized to improve the reconstruction of phylogenetic relationships.

We will need to be clear that the probabilities associated with quantum systems and those of phylogenetic models arise in quite a different scientific way. Quantum mechanics is a probabilistic theory because the theoretical predictions give the correct statistical behaviour regarding the outcomes of particular experiments. The theoretical predictions can be used to infer (incredibly accurately) the distribution of results for many repetitions of the same experiment. (For a popular discussion of the amazing accuracy of quantum theory see Feynman's discussion of the magnetic moment on the electron as predicted from quantum electrodynamics [22].) Since quantum theory is (and should be) seen

as a *theory* of nature there has been argument for many decades on how to interpret this probabilistic aspect of quantum theory. This argument raises quite profound scientific and philosophical issues which, thankfully, we will not be concerned with in this thesis. Models of phylogenetics are exactly that – *models*, and should not be seen as being theories of nature. No one would argue that the time evolution of molecular units follow the Markov model of phylogenetics in detail, but rather that these models are the best (tractable) approximation that give us recourse to establishing properties of phylogenetic history. Primarily the points of interest are the branching structure of the evolutionary history and also the evolutionary distance (or time) between branching events.

After we have made the mathematical analogy between quantum theory and the Markov model of phylogenetics, we will concentrate on only a small part of what can be done using techniques known in mathematical physics. We will focus on the study of entanglement invariants and their generalization to the phylogenetic case [59, 60]. There is potential for concentrating on other techniques such as Lie algebra symmetries [6] and the analysis of the path integral formulation [31, 32], but these techniques will not be explored here. The distance based technique has been used in phylogenetics as a tree building algorithm following the discovery that it is possible to calculate a distance from the observed sequences that is consistent with the Markov model. This distance function is a well defined mathematical object known as a group invariant function and is used in quantum physics to quantify and test for the phenomenon of entanglement. Entanglement is a general property that can exist in many different physical systems and the invariant function used as a distance measure in phylogenetics is used to quantify entanglement for only the most elementary case. Hence, it seems astute to investigate what the next most complicated types of entanglement correspond to in phylogenetics.

## Theoretical outcomes of the thesis

We present a group representation theoretic analysis of the Markov model of phylogenetic trees. Specifically this formalism is used to construct all the one-dimensional representations of the (appropriately defined) Markov semigroup. These one-dimensional representations occur as polynomials in the (discrete) probability distributions predicted from the Markov model which we coin *Markov invariants*. We establish the connection between these one-dimensional representations and that of *phylogenetic invariants* [11, 15, 20, 55] and pairwise distance measures [25, 40]. This representation theoretical approach touches upon existing techniques and can be incorporated into known algorithms to give novel results and insights to the problem of phylogenetic reconstruction. The main theoretical outcome of the thesis is this use of representation theory. We will also develop the theory of invariants of the general linear group on a tensor product space and show how to infer existence of these invariants in different cases. We develop a procedure for computing the explicit form of these invariant functions, firstly developed for the general linear group and then generalized to the Markov semigroup.

## Practical outcomes of the thesis

We study a group invariant function, well known in quantum physics as the *tangle*, in the context of phylogenetics. The tangle is used in physics to give a measure of the amount of entanglement between three qubits. Qubits are two state objects in quantum physics and correspond in phylogenetics to a probability distribution on two states. In phylogenetics the classic example is to use the DNA as a state space and hence the case of four state objects is of interest. To this end we have generalized the tangle to the case of three and four character states. This is a new result that to the best of the author's knowledge was previously unknown. Having successfully generalized the tangle we investigate how the tangle can be used to construct improved phylogenetic distance matrices. Additionally we study a set of Markov invariants which exist for the case of phylogenetic quartet tree. In the case of the evolution of four taxa there are three possible historical evolutionary relationships. We show that these Markov invariants can be used to distinguish these three cases under the assumption of the most general Markov model. It is expected that the use of the tangle to construct distance matrices and using the Markov invariants to distinguish the three possible quartets will lead to improvements of the reconstruction of phylogenetic relationships from observed biomolecular data.

## Structure of the thesis

Chapter 2 begins by introducing the mathematical material needed to understand the results presented in this thesis. This includes a short introduction to group representation theory, group characters and tensor product; a presentation of the Schur/Weyl duality and the Schur functions; a definition of group invariant functions and their relation to one-dimensional representations. The chapter ends with several relevant examples of invariants of the general linear group.

Chapter 3 begins with a light speed introduction to the formalism of quantum mechanics, the concept of entanglement and mathematical analysis thereof using group invariant functions. The Markov model of phylogenetic trees is then developed in its usual presentation, followed by a change of formalism which makes apparent the analogy between phylogenetic trees and multiparticle quantum systems. The chapter ends with a detailed analysis of the mathematical analysis of the invariant functions when evaluated upon a phylogenetic tree.

Chapter 4 gives a review of phylogenetic distance measures and shows how the tangle invariant function used to analyse three qubit entanglement can be generalized to the phylogenetic case and used to improve popular distance measures. This is done by defining the branch lengths of a phylogenetic tree, reviewing the standard measure known as the log det and then using the tangle invariant to give a consistent distance measure for the case of quartets.

Chapter 5 returns to the mathematical detail of Chapter 2 and derives in-



variant functions that are more closely relevant to the Markov model of a phylogenetic tree. This is done by first defining the Markov semigroup. The invariant functions of the general linear group are rederived using a technique which is generalized to derive the Markov invariants. Finally we examine the structure of the Markov invariants on a phylogenetic tree. In particular we concentrate on the quartet case where there exists four Markov invariants which can be used to distinguish between the three possible quartet trees.

## CHAPTER 2

# Mathematical background

In this chapter we will present the requisite mathematical background for developing the results presented in this thesis. It will be assumed that the reader is familiar with elementary concepts of algebra, most importantly the theory of groups and finite dimensional vector spaces (for example see [28]) and the theory of Lie groups and the classical groups (see [42]). The presentation will be brief and the reader interested in proofs is referred to the relevant literature as the discussion progresses. Our aim is to show how representation theory of groups – most notably the Schur/Weyl duality – can be used to count and construct the group invariant functions on a multi-linear (tensor product) space. We will develop some explicit invariants for the general linear group using a method which is known intuitively to many mathematical physicists and we formalize the technique.

## 2.1 Group representations

Throughout this thesis we will be interested in the vector spaces  $\mathbb{C}^n$  and  $\mathbb{R}^n$ . Almost all of the results presented will be equally valid whether one considers the complex or real space. Hence, we will simply refer to the vector space  $V$ , making the distinction between the real and complex case only when confusion may arise. For proofs of theorems that will be presented and further discussion of group representation theory the reader is referred to the excellent texts [27, 35, 42].

**Definition 2.1.1.** A group *representation*  $\rho$  on the vector space  $V$  is a homomorphism from a group  $G$  to the set of invertible, linear transformations  $GL(V)$ . The image element of  $g \in G$  is denoted by  $\rho(g)$  and the *dimension* of the representation is taken to be the dimension of the corresponding vector space.

A simple example of a group representation is constructed from the symmetric group on  $n$  elements,  $S_n$ , by taking a given group element  $\sigma \in S_n$  to simply

permute the basis vectors of the  $n$  dimensional vector space  $V$ :

$$\rho(\sigma)e_i := e_{\sigma i}.$$

It is clear that we have  $\rho(\sigma\sigma') = \rho(\sigma)\rho(\sigma')$  so that  $\rho$  is indeed a homomorphism from  $S_n$  to  $GL(V)$ .

We will often be interested in the case where the abstract group is a matrix group such as the general linear group  $GL(V)$  which is, of course, defined by its *action* on the vector space  $V$ . To avoid confusion, we will refer to this representation as the *defining* representation. To increase confusion we will write elements of the defining representation simply as  $g$ .

Given a matrix group  $G$ , there is always a one-dimensional representation defined by the determinant function:

$$\det : G \rightarrow \mathbb{C}^*,$$

where  $\mathbb{C}^* \cong \mathbb{C} \setminus \{0\}$  is the group of multiplications of non-zero complex numbers. The multiplicative property of the determinant:

$$\det(g_1 g_2) = \det(g_1) \det(g_2),$$

ensures that the determinant function defines a group homomorphism.

**Definition 2.1.2.** A subspace  $U \subseteq V$  is *invariant* under the group representation  $\rho$  if for all  $u \in U$  it follows that  $\rho(g)u \in U$  for all  $g \in G$ .

The notion of invariant subspaces allows us to break a given representation into its essential parts. That is, we can simplify the representation by considering its action upon the invariant subspaces alone.

**Definition 2.1.3.** A representation is *reducible* if there exists a non-trivial invariant subspace  $U$ . An *irreducible* representation is one which has no non-trivial invariant subspaces. A representation is *decomposable* if there exist non-trivial invariant subspaces  $U$  and  $W$  such that  $V \cong U \oplus W$ , and *indecomposable* otherwise. A representation is completely reducible if whenever there exists a non-trivial invariant subspace  $U$ , then there exists a second non-trivial invariant subspace  $W$  such that  $V \cong U \oplus W$ .

The matrix interpretation of a completely reducible representation is that there exists a basis where the matrix representation of each group element takes on a block-diagonal form. We will be exclusively interested in *integral* representations of the general linear group and its subgroups. Integral representations are those in which the entries of the representation matrix are polynomials in the matrix entries of  $GL(V)$  with respect to a particular basis. The integral representations of  $GL(V)$  are completely reducible [35].

**Definition 2.1.4.** The representations  $\rho_1$  and  $\rho_2$  are said to be *equivalent* if there exists an invertible linear transformation  $S$  on  $V$  such that

$$S\rho_1(g)S^{-1} = \rho_2(g)$$

for all  $g \in G$ .

From these considerations we can conclude that a given integral representation of the general linear group can be decomposed as

$$\rho = \bigoplus_a \rho_a,$$

where each  $\rho_a$  is an irreducible representation.

### 2.1.1 Group characters

**Definition 2.1.5.** The *character* of a representation  $\rho$  is defined as the trace function:

$$\chi(g) = \text{tr}(\rho(g)).$$

It follows immediately that the character is unaffected by similarity transformations:

$$\text{tr}(S\rho(g)S^{-1}) = \text{tr}(\rho(g)S^{-1}S) = \text{tr}(\rho(g)),$$

and is hence the same for equivalent representations.

The problem of classifying irreducible representations reduces to identifying the characters. Although the following result is valid only for finite groups, we will see that understanding the representation theory of  $S_n$  (a finite group) is crucial to constructing the irreducible representations of  $GL(V)$  (an infinite group).

**2.1.6.** *For a finite group, the number of non-equivalent irreducible representations of a group  $G$  is equal to the number of conjugacy classes of  $G$ .*

For example the conjugacy classes of the symmetric group can be found by considering the cycle notation which presents an element of  $S_n$  as a product of disjoint cycles. The lengths of these cycles adds to  $n$  and hence we get the well known result that the conjugacy classes of  $S_n$  are labelled by the partitions of  $n$ . (We will discuss partitions in more detail in the next section.) To illustrate this, consider that any element of the symmetric group can be written in the following form:

$$\sigma = (i_1 i_2 \dots i_{\alpha_1})(j_1 j_2 \dots j_{\alpha_2}) \dots (l_1 l_2 \dots l_{\alpha_p}).$$

This element belongs to the conjugacy class which is specified by the partition  $\{\alpha_1, \alpha_2, \dots, \alpha_p\}$  where  $\alpha_1 + \alpha_2 + \dots + \alpha_p = n$ . The fundamental result follows:

**2.1.7.** *The irreducible representations of the symmetric group  $S_n$  can be labelled by the partitions of  $n$ .*

For example we consider the representation on the  $n$ -dimensional vector space  $V$  of the symmetric group  $S_n$  defined, as above, by

$$\rho(\sigma)e_i = e_{\sigma i}.$$

Introducing the change of basis

$$\begin{aligned} z_0 &= \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i, \\ z_a &= \frac{1}{\sqrt{a(a+1)}} \sum_{i=1}^a (e_i - \sqrt{a} e_{a+1}), \quad a = 1, 2, \dots, n-1. \end{aligned} \quad (2.1)$$

It is clear that  $z_0$  spans a one-dimensional invariant subspace

$$\rho(\sigma)z_0 = \frac{1}{\sqrt{n}} \sum_{i=1}^n e_{\sigma i} = z_0;$$

and we have

$$\rho(\sigma)z_a = \frac{1}{\sqrt{a(a+1)}} \sum_{i=1}^a (e_{\sigma i} - \sqrt{a} e_{\sigma(a+1)}),$$

which itself belongs to the span of  $\{z_1, z_2, \dots, z_{n-1}\}$  which is consequently a complementary invariant space. To prove this consider the standard inner product:

$$(e_i, e_j) := \delta_{ij}.$$

and show that

$$(\rho(\sigma)z_a, z_0) = 0, \quad \forall \sigma \in S_n.$$

The representation of the symmetric group on the subspace  $z_0$  corresponds to the partition of  $n$  consisting of a single element:  $\{n\}$ .

Another one-dimensional representation of the symmetric group can be constructed by taking the sign of the permutation

$$\text{sgn}(\sigma) = \pm 1,$$

with the representation space  $\mathbb{C}$ . This representation corresponds to the partition  $\{1, 1, \dots, 1\}$  with  $1 + 1 + \dots + 1 = n$ .

### 2.1.2 Tensor product

The *dual* of the vector space,  $V$ , is denoted as  $V^*$  and defined to be the set of linear functionals  $\{f : V \rightarrow \mathbb{C}\}$ :

$$\begin{aligned} f(cv) &= cf(v), \\ f(v + v') &= f(v) + f(v'), \end{aligned}$$

for all  $c \in \mathbb{C}$  and  $v, v' \in V$ . Of course  $V^*$  itself forms a vector space and we use the basis  $\xi_1, \xi_2, \dots, \xi_n$  such that  $\xi_i(e_j) = \delta_{ij}$ . Since  $V$  and  $V^*$  are complex vector spaces of identical dimension they must be isomorphic and we define the linear functional  $\bar{v}$  as

$$\bar{v}(u) = \sum_{i=1}^n v_i^* u_i,$$

so that

$$\bar{v} = \sum_{i=1}^n v_i^* \xi_i.$$

With these definitions in hand we consider *bi-linear* functionals on the ordered product of two vector spaces  $V_1$  and  $V_2$  with bases  $\{e_i^{(1)}\}$  and  $\{e_j^{(2)}\}$  respectively. Such functionals map  $V_1 \times V_2$  to  $\mathbb{C}$  and satisfy

$$\begin{aligned} f(cv_1, v_2) &= cf(v_1, v_2) = f(v_1, cv_2), \\ f(v_1 + v'_1, v_2) &= f(v_1, v_2) + f(v'_1, v_2), \\ f(v_1, v_2 + v'_2) &= f(v_1, v_2) + f(v_1, v'_2), \end{aligned}$$

for all  $c \in \mathbb{C}$ ,  $v_1, v'_1 \in V_1$  and  $v_2, v'_2 \in V_2$ . Again this set of functionals forms a vector space which we denote as  $(V_1 \otimes V_2)^*$  with basis given by the set of functionals  $\xi_i^{(1)} \otimes \xi_j^{(2)}$  defined as

$$\xi_i^{(1)} \otimes \xi_j^{(2)}(e_k^{(1)}, e_l^{(2)}) := \xi_i(e_k^{(1)})\xi_j(e_l^{(2)}).$$

From which it follows that the bi-linear functional  $f$  can be written as

$$f = \sum_{i,j} f_{ij} \xi_i^{(1)} \otimes \xi_j^{(2)},$$

where  $f_{ij} = f(e_i^{(1)}, e_j^{(2)})$ . From this we can induce the definition of the *tensor product* of  $V_1$  and  $V_2$  to be the vector space  $V_1 \otimes V_2$ . A given element  $\psi \in V_1 \otimes V_2$  is referred to as a tensor and can be expressed uniquely in the form

$$\psi = \sum_{i,j} \psi_{ij} e_i^{(1)} \otimes e_j^{(2)}.$$

This process can be iterated to the tensor product of multiple vector spaces  $\mathcal{H} = V_1 \otimes V_2 \otimes \dots \otimes V_m$  where a given element  $\psi \in \mathcal{H}$  can be expressed as

$$\psi = \sum_{i_1, i_2, \dots, i_m} \psi_{i_1 i_2 \dots i_m} e_{i_1}^{(1)} \otimes e_{i_2}^{(2)} \otimes \dots \otimes e_{i_m}^{(m)}.$$

The tensor product space satisfies the axioms of a vector space with addition and scalar multiplication defined in the obvious way:

$$\begin{aligned} c \cdot \psi &= \sum_{i_1, i_2, \dots, i_m} c \psi_{i_1 i_2 \dots i_m} e_{i_1}^{(1)} \otimes e_{i_2}^{(2)} \otimes \dots \otimes e_{i_m}^{(m)}, \\ \psi + \varphi &= \sum_{i_1, i_2, \dots, i_m} (\psi_{i_1 i_2 \dots i_m} + \varphi_{i_1 i_2 \dots i_m}) e_{i_1}^{(1)} \otimes e_{i_2}^{(2)} \otimes \dots \otimes e_{i_m}^{(m)}. \end{aligned}$$

When one is taking the tensor product of a single vector space we use the notation

$$V^{\otimes m} := V \otimes V \otimes \dots \otimes V.$$

Again,  $\mathcal{H} := V^{\otimes m}$  must be isomorphic to  $\mathcal{H}^* \cong (V^*)^{\otimes m}$  and we define

$$\bar{\psi} = \sum_{i_1, i_2, \dots, i_m} \psi_{i_1 i_2 \dots i_m}^* \xi_{i_1} \otimes \xi_{i_2} \otimes \dots \otimes \xi_{i_m},$$

so that

$$\bar{\psi}(\varphi) = \sum_{i_1, i_2, \dots, i_m} \psi_{i_1 i_2 \dots i_m}^* \varphi_{i_1 i_2 \dots i_m}.$$

### 2.1.3 Group action on a tensor product space

Given a set of representations of a group

$$\rho_a : G \rightarrow GL(V_a), \quad a = 1, 2, \dots, m,$$

it is possible to construct a new representation  $\rho$  by taking the tensor product

$$\mathcal{H} = V_1 \otimes V_2 \otimes \dots \otimes V_m$$

and define the *tensor product representation* on the vector space  $\mathcal{H}$  to act as:

$$\begin{aligned} \rho(g)\psi &:= \rho_1(g) \otimes \rho_2(g) \otimes \dots \otimes \rho_m(g)\psi, \\ &= \sum_{i_1, i_2, \dots, i_m} \psi_{i_1 i_2 \dots i_m} \rho_1(g)e_{i_1}^{(1)} \otimes \rho_2(g)e_{i_2}^{(2)} \otimes \dots \otimes \rho_m(g)e_{i_m}^{(m)}. \end{aligned}$$

In contrast to this we consider another important case which occurs when we have the direct (cartesian) product of  $m$  groups:

$$G = G_1 \times G_2 \times \dots \times G_m,$$

with representations  $\rho_1, \rho_2, \dots, \rho_m$  and associated representation spaces

$$V_1, V_2, \dots, V_m.$$

It is again possible to define a representation  $\bar{\rho}$  on  $\mathcal{H}$  as

$$\bar{\rho}(g)\psi = \bar{\rho}(g_1 \times g_2 \times \dots \times g_m)\psi = \rho_1(g_1) \otimes \rho_2(g_2) \otimes \dots \otimes \rho_m(g_m)\psi.$$

<sup>1</sup> For future use we define the notation

$$\begin{aligned} \times^m G &:= G \times G \times \dots \times G \\ \otimes^m g &:= g \otimes g \otimes \dots \otimes g. \end{aligned}$$

Presently we will recall the character theory of the general linear group to enable us to decompose such representations into their irreducible parts.

---

<sup>1</sup>Interestingly, in quantum physics the appropriate description of a multi-particle system is given by taking the tensor product of different representations of a single group, such as the orthogonal or Lorentz groups, where the choice of each representation is fixed by the individual particle types. Whereas in the case of stochastic models of phylogenetics the reverse is the case; the system is described by taking the group action on the tensor product space as the direct product of the defining representation of the Markov semigroup.

## 2.2 Irreducible representations of the general linear group

It is well known from group representation theory that the finite-dimensional irreducible representations of the general linear and the symmetric group can be put into a correspondence. This result is known as the Schur/Weyl duality. As we saw above, the irreducible representations of the symmetric group on  $n$  elements can be labelled by the partitions of  $n$ . Additionally, there exist algorithms for explicitly constructing these irreducible representations once a partition has been specified. Here we will show how the irreducible representations of the general linear group on  $V$  occur as subspaces of the tensor product space  $V^{\otimes m}$ . These projections are constructed using operators known as *Young's operators* which are computed from the partitions of  $m$ .

### 2.2.1 Partitions

A finite sequence of positive integers

$$\lambda = \{\lambda_1, \lambda_2, \dots\}$$

with  $\lambda_1 \geq \lambda_2 \geq \dots$ , is an (ordered) *partition* of the integer  $n$  if the *weight* of the partition,

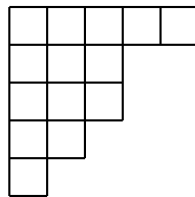
$$|\lambda| := \lambda_1 + \lambda_2 + \dots,$$

satisfies  $|\lambda| = n$ .

It is usual to use a notation which indicates the number of times each integer occurs as a part:

$$\lambda = \{\dots, r^{m_r}, \dots, 2^{m_2}, 1^{m_1}\}$$

so that  $m_i$  of the parts of  $\lambda$  are equal to  $i$ . It is useful to represent a given partition as a *Ferrers diagram* by drawing a row of squares for each part of the partition, and placing these rows upon each other sequentially such that the rows decrease in length down the page. For example the partition  $\lambda = \{5, 3^2, 2, 1\}$  is represented by:





**Definition 2.2.1.** A *Young tableau*,  $T$ , of shape  $\lambda$  with  $|\lambda| = n$  is an assignment of the integers  $1, 2, \dots, n$  to a Ferrers diagram such that the rows and columns are strictly increasing. A *semi-standard tableau*,  $T'$ , requires that only the rows need to be increasing.

For example, the canonical Young tableau of shape  $\{5, 3^2, 2, 1\}$  is:

1	2	3	4	5
6	7	8		
9	10	11		
12	13			
14				

while a semi-standard tableau of the same shape is:

1	1	1	2	3
3	4	4		
4	5	6		
6	6			
7				

**Definition 2.2.2.** The ring of *symmetric functions*,  $\Lambda_n = \mathbb{Z}[x_1, \dots, x_n]^{S_n}$ , is the set of polynomials in  $n$  independent variables  $x_1, \dots, x_n$  which are invariant under the representation of  $S_n$  defined by permutations of the variables.

That is,  $f$  is a symmetric function if and only if:

$$f(x_1, x_2, \dots, x_n) = f(x_{\sigma 1}, x_{\sigma 2}, \dots, x_{\sigma n}), \quad \forall \sigma \in S_n.$$

It is clear that  $\Lambda_n$  is a graded ring:

$$\Lambda_n = \bigoplus_{d \geq 0} \Lambda_n^d$$

where  $\Lambda_n^d \subset \Lambda_n$  consists of the homogeneous symmetric polynomials of degree  $d$ . Various bases exist for the ring of symmetric functions (see [41]). The basis which will be of use to us is given by the Schur functions.

### 2.2.2 The Schur functions

For a given partition  $\lambda$  define the monomial  $x^\lambda = x_1^{\lambda_1} x_2^{\lambda_2} \dots x_n^{\lambda_n}$ . Consider the polynomial which is obtained by anti-symmetrizing:

$$a_\lambda = a_\lambda(x_1, \dots, x_n) = \sum_{\sigma \in S_n} \text{sgn}(\sigma) \sigma(x^\lambda),$$

1 1 2	1 1 3	1 2 2	1 2 3	1 3 2	1 3 3	2 2 3	2 3 3
----------	----------	----------	----------	----------	----------	----------	----------

Figure 2.1: Semi-standard tableaux

where

$$\sigma(x^\lambda) := x_{\sigma 1}^{\lambda_1} x_{\sigma 2}^{\lambda_2} \dots x_{\sigma n}^{\lambda_n}.$$

By considering the partition  $\delta = \{n-1, n-2, \dots, 1\}$  it follows that

$$a_\delta = \prod_{1 \leq i < j \leq n} (x_i - x_j),$$

which is called the *Vandermonde determinant*. The Schur functions are then defined as the quotient

$$s_\lambda = s_\lambda(x_1, x_2, \dots, x_n) = a_{\lambda+\delta} / a_\delta,$$

which is clearly symmetric. A more intuitive and constructive way of defining the Schur functions is to take:

$$s_\lambda = \sum_{T'} x^{T'},$$

where the summation is over all semi-standard  $\lambda$  tableaux  $T'$ . For example, for  $\lambda = \{2, 1\}$  the semi-standard tableaux are displayed in Figure 2.1. In this case each tableau corresponds to a monomial  $x^T$  to give

$$s_{21}(x_1, x_2, x_3) = x_1^2 x_2 + x_1^2 x_3 + x_1 x_2^2 + 2x_1 x_2 x_3 + x_1 x_3^2 + x_2^2 x_3 + x_2 x_3^2$$

which is easily seen to be a symmetric polynomial.

### 2.2.3 Group characters of $GL(n)$

For a given matrix  $g \in GL(n)$  it is possible to use the Jordan decomposition to put it in upper triangular form and hence the character is simply the sum of the eigenvalues:

$$\chi(g) = \text{tr}(g) = x_1 + x_2 + \dots + x_n.$$

This corresponds to the Schur function

$$s_{\{1\}}(x_1, \dots, x_n) = x_1 + x_2 + \dots + x_n.$$

By considering the tensor product representation of  $GL(V)$  on  $V \otimes V$  we have

$$V^{\{2\}} := \{\psi^{(s)} | \psi_{i_1 i_2}^{(s)} = \psi_{i_2 i_1}\}, \quad V^{\{1^2\}} := \{\psi^{(a)} | \psi_{i_1 i_2}^{(a)} = -\psi_{i_2 i_1}\},$$

as irreducible subspaces known as the *symmetric* and *anti-symmetric* tensors with dimensions  $\frac{1}{2}n(n+1)$  and  $\frac{1}{2}n(n-1)$  respectively. We have

$$\psi = \frac{1}{2}(\psi^{(s)} + \psi^{(a)}),$$

where

$$\psi_{i_1 i_2}^{(s)} = \frac{1}{2}(\psi_{i_1 i_2} + \psi_{i_2 i_1}), \quad \psi_{i_1 i_2}^{(a)} = \frac{1}{2}(\psi_{i_1 i_2} - \psi_{i_2 i_1}),$$

so that the decomposition of  $V \otimes V$  under the action of  $GL(n)$  into irreducible subspaces is given by

$$V \otimes V = V^{\{2\}} \oplus V^{\{1^2\}}.$$

Now suppose we take the group element  $g \in GL(n)$ . It follows from an elementary calculation that the character of this group element on the representation  $V^{\otimes 2}$  is simply the product:

$$\chi(g \otimes g) = (x_1 + x_2 + \dots + x_n)(x_1 + x_2 + \dots + x_n).$$

In terms of the Schur functions it follows that we have the decomposition

$$\begin{aligned} \chi(g \otimes g) &= s_{\{1\}}(x)s_{\{1\}}(x), \\ &= s_{\{2\}}(x) + s_{\{1^2\}}(x) \end{aligned}$$

where

$$\begin{aligned} s_{\{2\}}(x) &= (x_1^2 + x_1x_2 + x_1x_3 + \dots + x_2^2 + x_2x_3 + \dots + x_n^2), \\ s_{\{1^2\}}(x) &= (x_1x_2 + x_1x_3 + \dots + x_2x_3 + \dots + x_{n-1}x_n). \end{aligned}$$

Thus we see that the decomposition of the tensor product representation into irreducible parts can be inferred by using the Schur functions as a basis for the ring of symmetric functions. This is the archetypal example from physics and leads to the full Schur/Weyl duality which allows us to classify the irreducible representations of  $GL(n)$  (and its subgroups) by simply using the character formulas and the Schur functions.

### 2.2.4 The Schur/Weyl duality

In this section we will construct the Schur/Weyl duality which states that the irreducible representations of the general linear group and that of the symmetric group can be put into correspondence.

**2.2.3.** *If  $V$  decomposes into the direct sum  $V = U \oplus W$  where  $U$  and  $W$  are invariant subspaces under the group representation  $\rho$ , then the projection operator  $P$ , defined by  $PV \cong U$ , satisfies*

$$P\rho(g) = \rho(g)P, \quad \forall g \in G, \tag{2.2}$$

*and similarly for the orthogonal projection  $(1 - P)$ . Conversely, if  $P$  is a projection operator satisfying (2.2) then the subspace it projects to is invariant under  $\rho$ .*

Consider the representation of the symmetric group on  $V^{\otimes m}$  defined by

$$\sigma(e_{i_1} \otimes e_{i_2} \otimes \dots \otimes e_{i_m}) := e_{i_{\sigma 1}} \otimes e_{i_{\sigma 2}} \otimes \dots \otimes e_{i_{\sigma m}}.$$

It should be clear that the action of any such element of the symmetric group will commute with the tensor product representation of  $GL(n)$ . In addition to this the *algebra* generated from this action will commute with  $GL(n)$  and hence can be used to construct projection operators which satisfy (2.2). Presently we will discuss how to construct such projection operators such that the corresponding invariant subspaces are in fact irreducible.

Consider a Young tableau with shape  $\lambda$  and  $|\lambda| = m$ . Consider the permutations  $p$  which interchange the integers in the same row, and, conversely, permutations  $q$  which interchange numbers in the same column. In the algebra of the symmetric group action defined above, consider the quantities

$$P = \sum_p p,$$

$$Q = \sum_q \text{sgn}(q)q.$$

The Young operator corresponding to the standard tableau  $T$  is then defined to be

$$Y = QP,$$

and we have the fundamental result:

**2.2.4.** *For a given partition  $\lambda$ ,  $Y$  projects onto an irreducible subspace of  $V^{\otimes m}$  under the tensor product representation of  $GL(n)$ . Young tableaux of the same shape label equivalent representations.*

Now suppose  $Y_\lambda$  is the Young operator corresponding to the partition  $\lambda$ . We define the subspace

$$V^\lambda := Y_\lambda V^{\otimes m}.$$

It is possible to prove that the group character of the tensor product representation of the general linear group on the subspace  $V^\lambda$  is none other than the Schur function  $s_\lambda(x_1, x_2, \dots, x_n)$ .

For example we consider the standard tableau:

$$\begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 & \\ \hline \end{array}.$$

With corresponding Young operator given by

$$P = e + (13),$$

$$Q = e - (12),$$

$$Y = e + (13) - (12) - (123).$$

We also note that the dimension of the invariant subspaces are given by setting the characteristic values in the Schur function equal to the identity:

$$\text{dimension of } V^\lambda = s_\lambda(1, 1, \dots, 1).$$

### 2.2.5 More representations

From this construction we can build more representations such as

$$V^\mu \otimes V^\nu,$$

with group character which corresponds to the *outer* product of two Schur functions which is defined as the pointwise product:

$$s_\mu s_\nu(x) := s_\mu(x) s_\nu(x) = \sum_{\lambda} c_{\mu\nu}^\lambda s_\lambda,$$

where  $|\lambda| = |\mu| + |\nu|$  and the  $c_{\mu\nu}^\lambda$  are integer coefficients which can be determined by the Littlewood-Richardson rule [39, 41].

Another way of constructing representations is to consider

$$(V^\mu)^\nu.$$

The group character of this representation is given by another type of multiplication of Schur functions known as the *plethysm* (defined formally in Macdonald [41]). Here we use Young's tableaux to give a constructive definition. Recall that we have

$$s_\mu(x) = \sum_{T'} x^{T'}$$

which is a summation of monomials in  $x_1, x_2, \dots, x_n$ . If there are  $m$  such monomials in  $s_\mu(x)$  and these are denoted by  $y_i$ ,  $1 \leq i \leq m$ , then the plethysm is given by

$$s_\lambda[s_\mu](x) = s_\lambda(y) = \sum_{T'} y^{T'}.$$

The plethysm  $s_\lambda[s_\mu]$  can be interpreted as giving the character of the representation  $(V^\mu)^\lambda$ . That is we take  $V^\mu$  as the defining representation and symmetrize this representation with  $\lambda$ .

Finally the *inner* product of two Schur functions is defined as

$$s_\mu(x) * s_\nu(x) = \sum_{\lambda} \gamma_{\mu\nu}^\lambda s_\lambda(x),$$

where  $|\mu| = |\nu| = |\lambda| = n$  and the  $\gamma_{\mu\nu}^\lambda$  are the integer multiplicities of the  $\lambda$  representation of  $S_n$  occurring in the decomposition of the tensor product of the  $\mu$  and  $\nu$  representations of  $S_n$ . The inner product comes into play if we wish to compute the character of  $GL(n) \times GL(n')$  on  $V^\lambda \otimes V^{\lambda'}$  with  $|\lambda| = m$ ,  $|\lambda'| = m'$ . The character of this representation is  $s_\lambda(x) s_{\lambda'}(y)$  where  $x_1, x_2, \dots, x_n$  and  $y_1, y_2, \dots, y_{n'}$  are the eigenvalues of the relevant group elements in  $GL(n)$  and  $GL(n')$  respectively. The decomposition of characters is given by the formula

$$s_\lambda(x) s_{\lambda'}(y) = \sum_{\rho} \gamma_{\lambda\lambda'}^\rho s_\rho(xy),$$

where  $(xy) = (x_1y_1, x_1y_2, \dots, x_2y_1, \dots, x_ny_n)$  [41].

We will often write the Schur function  $s_\lambda$  simply as  $\{\lambda\}$  and the plethysm will sometimes be written as

$$s_\lambda[s_\mu] = \{\mu\}^{\otimes \{\lambda\}}.$$

In practice we compute Schur multiplications by using the group theory software **Schur** [63]. For further discussion of Schur functions and their various multiplications see [4, 10, 17, 16, 39].

As an example consider the defining representation of  $GL(n)$  on the tensor product  $V^{\otimes m}$ . That is

$$\psi \rightarrow g \otimes g \otimes \dots \otimes g\psi := \otimes^m g\psi,$$

for  $\psi \in V^{\otimes m}$ ,  $g \in GL(n)$ . The character of this representation is given by the pointwise product of  $m$  copies of  $s_{\{1\}}(x)$  and can be decomposed into irreducible characters by using the Littlewood-Richardson coefficients  $c_{\mu\nu}^\lambda$ . In the case where  $m = 4$ , **Schur** gives

$$\begin{aligned} s_{\{1\}}(x)s_{\{1\}}(x)s_{\{1\}}(x)s_{\{1\}}(x) = \\ s_{\{4\}}(x) + 3s_{\{31\}}(x) + 2s_{\{2^2\}}(x) + 3s_{\{21^2\}}(x) + s_{\{1^4\}}(x). \end{aligned}$$

This tells us that under the action of  $GL(n)$  the tensor product  $V^{\otimes 4}$  decomposes into irreducible subspaces:

$$V^{\otimes 4} = V^{\{4\}} + 3V^{\{31\}} + 2V^{\{2^2\}} + 3V^{\{21^2\}} + V^{\{1^4\}},$$

where the multiplicities account for the number of legal standard tableaux for each partition.

### 2.2.6 One-dimensional representations

Recall that the dimension of the irreducible representation  $\lambda$  is given by  $s_\lambda(1, 1, \dots, 1)$ . It follows that the one-dimensional representations occur when there is only a single semi-standard tableau with shape  $\lambda$ . In the case when  $V$  is  $n$ -dimensional it should be clear that the one-dimensional representations occur when we have  $\lambda = \{k^n\}$  for some  $k$ .

Consider the character of  $GL(n)$  on  $V^{\{k^n\}}$ :

$$\begin{aligned} s_{\{k^n\}}(x) &= (x_1x_2 \dots x_n)^k, \\ &= \det(g)^k. \end{aligned}$$

Thus for any  $\eta \in V^{\{k^n\}}$  we have

$$\eta \mapsto \det(g)^k \eta,$$

under the  $\{k^n\}$  representation of  $GL(n)$ .

## 2.3 Invariant theory

Given the defining representation of a group  $G$  on a vector space  $V$ , it is possible to define a representation which acts on the vector space of functions  $f : V \rightarrow \mathbb{C}$  as

$$gf := f \circ g^{-1}. \quad (2.3)$$

(It is necessary to take the inverse of the group element to ensure that the induced representation satisfies the properties of a group homomorphism.) An *invariant* with *weight*  $k \in \mathbb{N}$  is then defined as any function which satisfies

$$g^{-1}f = f \circ g = \det(g)^k f. \quad (2.4)$$

We will be exclusively interested in the case where  $f$  is a polynomial in the dual vector space  $V^*$  with basis elements  $\{\xi_1, \xi_2, \dots, \xi_n\}$ . In order to generate polynomials in this space multiplication is defined pointwise:

$$\xi_a \xi_b(x) := \xi_a(x) \xi_b(x), \quad 1 \leq a, b \leq n.$$

The full set of polynomials generated from this construction is denoted as  $\mathbb{C}[V]$ . A *homogeneous* polynomial satisfies

$$f \circ c1 = c^d f, \quad c \in \mathbb{C},$$

for some positive integer  $d$  which is referred as the *degree*. From elementary considerations it follows that  $\mathbb{C}[V]$  has the structure of a graded algebra over the degree:

$$\mathbb{C}[V] = \bigoplus_d \mathbb{C}[V]_d,$$

where  $\mathbb{C}[V]_d$  is the set of homogeneous polynomials of degree  $d$ .

By counting the degree of the various algebraic quantities we see that  $d = nk$ , and we denote

$$\mathbb{C}[V]_d^G = \{f \in \mathbb{C}[V]_d \mid f \circ g = \det(g)^k f, \forall g \in G\}.$$

Of course we have already studied invariant functions on a finite group! The symmetric functions are none other than the set  $\mathbb{C}[V]^{S_n}$  with  $k = 0$ . Another example comes from the classical groups which are defined by imposing invariant functions. For example the orthogonal group  $O(n)$  acting on  $\mathbb{R}^n$  can be considered to be defined by the invariant function

$$r^2(x) = \sum_{i=1}^n x_i^2.$$

Consider the tensor product space  $\mathbb{C}^2 \otimes \mathbb{C}^2$  with associated group action  $GL(2) \times GL(2)$ . The following relation holds for any  $\psi = \sum_{i,j} \psi_{ij} e_i \otimes e_j \in \mathbb{C}^2 \otimes \mathbb{C}^2$ :

$$(\psi'_{11} \psi'_{22} - \psi'_{12} \psi'_{21}) = \det g (\psi_{11} \psi_{22} - \psi_{12} \psi_{21}),$$

where  $\psi' := g_1 \otimes g_2 \psi$  and  $\det g = \det g_1 \det g_2$ . So  $(\psi_{11} \psi_{22} - \psi_{12} \psi_{21})$  is an invariant.

### 2.3.1 Invariants as irreducible representations

In this section we will show that the group invariant polynomials  $\mathbb{C}[V^{\otimes m}]_d^G$  occur exactly as the one-dimensional representations  $V^{\{k^n\}}$  in the decomposition of  $(V^{\otimes m})^{\{d\}}$  with  $md = kn$ . As a first step consider a vector space  $U$ . We establish the vector space isomorphism:

$$U^{\{d\}} \cong \mathbb{C}[U]_d.$$

This follows by observing that if  $U$  has basis  $\{u_i\}$ , then  $U^{\{d\}}$  consists of all tensors of the form

$$\psi = \sum_{i_1, i_2, \dots, i_d} \psi_{i_1 i_2 \dots i_d} u_{i_1} \otimes u_{i_2} \otimes \dots \otimes u_{i_d},$$

where  $\psi_{i_1 i_2 \dots i_d}$  is invariant under permutations of indices. Now if  $U^*$  has basis  $\{\zeta_i\}$ , consider an arbitrary element of  $\mathbb{C}[U]_d$ :

$$f = \sum_{i_1, i_2, \dots, i_d} f_{i_1 i_2 \dots i_d} \zeta_{i_1} \zeta_{i_2} \dots \zeta_{i_d}.$$

Clearly  $f_{i_1 i_2 \dots i_d}$  is also invariant under permutations of indices. This identification establishes the isomorphism. We define the canonical isomorphism  $\omega : U^{\{d\}} \rightarrow \mathbb{C}[U]_d$  as

$$\omega(\psi) = \sum_{i_1, i_2, \dots, i_d} \psi_{i_1 i_2 \dots i_d} \zeta_{i_1} \zeta_{i_2} \dots \zeta_{i_d}, \quad (2.5)$$

with inverse

$$\omega^{-1}(f) = \sum_{i_1, i_2, \dots, i_d} f_{i_1 i_2 \dots i_d} u_{i_1} \otimes u_{i_2} \otimes \dots \otimes u_{i_d}.$$

By explicit computation

$$\omega(\otimes^d g^t \psi) = \omega(\psi) \circ g = g^{-1} \omega(\psi), \quad (2.6)$$

and

$$\omega^{-1}(g^{-1} f) = \omega^{-1}(f \circ g) = \otimes^d g^t \omega^{-1}(f) \quad (2.7)$$

for all  $g \in GL(U)$ ,  $f \in \mathbb{C}[U]_d$  and  $\psi \in U^{\{d\}}$ .

From these considerations we generalize to the case where  $U = V^{\otimes m}$  and establish the main result of this section:

**Theorem 2.3.1.** *Consider integers  $m, d, k, n$  with  $md = kn$  and label the occurrences of  $V^{\{k^n\}}$  in the decomposition of  $(V^{\otimes m})^{\{d\}}$  by an integer  $a$ . It follows that*

$$\mathbb{C}[V^{\otimes m}]_d^{GL(n)} \cong \oplus_a V_a^{\{k^n\}}.$$



*Proof.* Suppose  $f \in \mathbb{C}[V^{\otimes m}]_d^{GL(n)}$ . We have

$$\begin{aligned}\omega^{-1}(g^{-1}f) &= \omega^{-1}(\det(g)^k f) = \det(g)^k \omega^{-1}(f) \\ &= \otimes^d g^t \omega^{-1}(f),\end{aligned}$$

and hence the representation space  $\text{span}[\omega^{-1}(f)]$  provides a one-dimensional representation of  $GL(n)$ . Conversely, suppose that  $\text{span}[\psi]$  with  $\psi \in (V^{\otimes m})^{\{d\}}$  provides a one-dimensional representation of  $GL(n)$  such that

$$\otimes^d g\psi = \det(g)^k \psi.$$

Noting that  $\det(g) = \det(g^t)$  it follows that

$$\begin{aligned}\omega(\otimes^d g\psi) &= \omega(\det(g)^k \psi) = \det(g)^k \omega(\psi) \\ &= \omega(\otimes^d g^t \psi) = g^{-1} \omega(\psi),\end{aligned}$$

so we can conclude that  $\omega(\psi) \in \mathbb{C}[V^{\otimes m}]_d^{GL(n)}$ .  $\square$

### 2.3.2 Using Schur functions to count invariants

By the preceding theorems we conclude the following:

**Theorem 2.3.2.** *The number of invariants in  $\mathbb{C}[V^{\otimes m}]_d^{GL(n)}$  of weight  $k$  is equal to the number of occurrences of  $\{k^n\}$  in the decomposition of  $(\times^m \{1\})^{\otimes \{d\}}$ .*

We now consider the character of  $\times^m GL(n)$  on  $(V^{\otimes m})^\lambda$ :

$$\begin{aligned}s_\lambda(x^{(1)}x^{(2)} \dots x^{(m)}) \\ = \sum_{\substack{\mu_1, \dots, \mu_m, \\ \nu_1, \dots, \nu_{m-1}}} \gamma_{\mu_1 \nu_1}^\lambda \gamma_{\mu_2 \nu_2}^{\nu_1} \dots \gamma_{\mu_{m-1} \nu_{m-1}}^{\nu_{m-2}} s_{\mu_1}(x^{(1)}) s_{\mu_2}(x^{(2)}) \dots s_{\mu_m}(x^{(m)}),\end{aligned} \quad (2.8)$$

where  $(x^{(1)}x^{(2)} \dots x^{(m)}) = (x_{i_1}^{(1)}x_{i_2}^{(2)} \dots x_{i_m}^{(m)})_{1 \leq i_a \leq n}$ . Now each term in (2.8) is an irreducible character

$$s_{\sigma_1}(x^{(1)}) s_{\sigma_2}(x^{(2)}) \dots s_{\sigma_m}(x^{(m)}),$$

with  $|\sigma_i| = |\lambda|$  and multiplicity

$$q(\sigma_1, \sigma_2, \dots, \sigma_m; \lambda) := \sum_{\nu_1, \dots, \nu_{m-2}} \gamma_{\sigma_1 \nu_1}^\lambda \gamma_{\sigma_2 \nu_2}^{\nu_1} \dots \gamma_{\sigma_{m-1} \sigma_m}^{\nu_{m-2}}.$$

From the definition of the inner product

$$q(\sigma_1, \sigma_2, \dots, \sigma_m; \lambda) = \{\text{multiplicity of } \lambda \text{ in } \sigma_1 * \sigma_2 \dots * \sigma_m\}.$$

The dimension of each of the irreducible representations (2.9) is equal to the product of the dimensions of each component irreducible representations. To identify invariant functions we are led to the following theorem:

**Theorem 2.3.3.** *The number of weight  $k$  invariants in  $\mathbb{C}[V^{\otimes m}]_d^{\times^m GL(n)}$  is equal to the number of occurrences of the Schur function  $\{d\}$  in  $*^m \{k^n\}$ .*

## 2.4 Invariants of the general linear group

We have established that any one-dimensional representation of  $GL(n)$  occurs as a partition of the form  $\{k^n\}$ . This is because the columns of the partitions correspond to the anti-symmetrization process of Young operators and it is clear that if we anti-symmetrize  $n$  elements  $n$  times then there will only be a single independent element remaining. Presently we will present a generic scheme which allows us to generate the exact polynomial form of these representations.

Consider the definition of the determinant of a matrix  $g$ :

$$\det(g) = \sum_{\sigma \in S_n} \text{sgn}(\sigma) g_{1\sigma(1)} g_{2\sigma(2)} \cdots g_{n\sigma(n)}.$$

By defining the (anti-symmetric) Levi-Civita tensor

$$\epsilon := \sum_{i_1, i_2, \dots, i_n} \epsilon_{i_1 i_2 \dots i_n} e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_n},$$

where  $\epsilon_{\sigma(1)\sigma(2)\dots\sigma(n)} := \text{sgn}(\sigma)$ , it follows that the determinant can be expressed as

$$\det(g) = \frac{1}{n!} \sum_{\substack{i_1, i_2, \dots, i_n, \\ j_1, j_2, \dots, j_n}} g_{i_1 j_1} g_{i_2 j_2} \cdots g_{i_n j_n} \epsilon_{i_1 i_2 \dots i_n} \epsilon_{j_1 j_2 \dots j_n}. \quad (2.9)$$

Presently we will show that

$$\epsilon' := g \otimes g \otimes \cdots \otimes g \epsilon = \det(g) \epsilon,$$

for all matrices  $g$ . In components we have

$$\epsilon'_{i_1 i_2 \dots i_n} = \sum_{j_1, j_2, \dots, j_n} g_{i_1 j_1} g_{i_2 j_2} \cdots g_{i_n j_n} \epsilon_{j_1 j_2 \dots j_n},$$

and it is clear that  $\epsilon'_{i_1 i_2 \dots i_n}$  is completely anti-symmetric under interchange of indices and hence must be proportional to  $\epsilon_{i_1 i_2 \dots i_n}$ . Finally we use (2.9) to conclude that

$$\epsilon' = \det(g) \epsilon.$$

**Theorem 2.4.1.** *Consider a function  $f : V^{\otimes n} \times V^{\otimes m} \rightarrow \mathbb{C}$  which satisfies the conditions:*

1. *For fixed  $\chi \in V^{\otimes n}$  we have  $f \in \mathbb{C}[V^{\otimes m}]_d$ . That is  $f(\chi, c\psi) = c^d f(\chi, \psi)$ .*
2. *For fixed  $\psi \in V^{\otimes m}$  we have  $f \in \mathbb{C}[V^{\otimes n}]_k$ . That is  $f(c\chi, \psi) = c^k f(\chi, \psi)$ .*
3.  *$f(\chi, \otimes^m g\psi) = f(\otimes^n g^t \chi, \psi)$ .*

The function  $f_\epsilon : V^{\otimes m} \rightarrow \mathbb{C}$  given by

$$f_\epsilon(\psi) := f(\epsilon, \psi)$$

then satisfies  $f_\epsilon(\otimes^m g\psi) = \det(g)^k f_\epsilon(\psi)$ .

*Proof.* We have

$$\begin{aligned} f_\epsilon(\otimes^m g\psi) &= f(\epsilon, \otimes^m g\psi) = f(\otimes^n g^t \epsilon, \psi) = f(\det(g)\epsilon, \psi) = \det(g)^k f(\epsilon, \psi) \\ &= \det(g)^k f_\epsilon(\psi). \end{aligned}$$

□

This theorem gives us some idea of how to explicitly construct invariants for the general linear group. The rest of this chapter will be devoted to the illustration of several examples.

### 2.4.1 Invariants of $GL(n)$ on $V^{\otimes m}$

For this case the number of invariants of  $GL(n)$  on  $V^{\otimes m}$  is given by the multiplicity of  $\{k^n\}$  in  $(\times^m \{1\})^{\otimes \{d\}}$  with  $nk = md$ . Here we will consider  $m = 2$  and the cases  $n = 2, 3, 4$ .

#### The case of $GL(2)$

In the case that  $n = m = 2$ , the possible degrees of the invariants are

$$d = 1, 2, 3, 4, \dots$$

and using **Schur** we find

$$\begin{aligned} (\{1\} \times \{1\})^{\otimes \{1\}} &\supset \{1^2\}, \\ (\{1\} \times \{1\})^{\otimes \{2\}} &\supset 2\{2^2\}, \\ (\{1\} \times \{1\})^{\otimes \{3\}} &\supset 2\{3^2\}, \\ (\{1\} \times \{1\})^{\otimes \{4\}} &\supset 3\{4^2\}, \\ (\{1\} \times \{1\})^{\otimes \{5\}} &\supset 3\{5^2\}, \\ (\{1\} \times \{1\})^{\otimes \{6\}} &\supset 4\{6^2\}. \end{aligned}$$

At each degree the correct number of invariants can be built from

$$\begin{aligned} f_1(\psi) &:= \sum_{i_1, i_2} \psi_{i_1 i_2} \epsilon_{i_1 i_2} = (\psi_{12} - \psi_{21}), \\ f_2(\psi) &:= \sum_{i_1, i_2, j_1, j_2} \psi_{i_1 i_2} \psi_{j_1 j_2} \epsilon_{i_1 j_1} \epsilon_{i_2 j_2} = (\psi_{11} \psi_{22} - \psi_{12} \psi_{21}), \end{aligned} \tag{2.10}$$

and are non-zero, algebraically independent, and by inspection satisfy (2.4.1).

**The case of  $GL(3)$** 

In the case that  $n = 3, m = 2$ , the possible degrees of invariants are

$$d = 3, 6, 9, 12, \dots$$

Computing plethysms in **Schur** gives

$$\begin{aligned} (\{1\} \times \{1\})^{\otimes\{3\}} &\supset 2\{2^3\}, \\ (\{1\} \times \{1\})^{\otimes\{6\}} &\supset 3\{4^3\}, \\ (\{1\} \times \{1\})^{\otimes\{9\}} &\supset 4\{6^3\}. \end{aligned}$$

At each degree the correct number of invariants can be built from the two  $d = 3$  invariants:

$$\begin{aligned} f_1(\psi) &:= \sum_{i_1, i_2, j_1, j_2, k_1, k_2} \psi_{i_1 i_2} \psi_{j_1 j_2} \psi_{k_1 k_2} \epsilon_{i_1 j_1 j_2} \epsilon_{i_2 j_2 k_2} \\ &= -\psi_{13} \psi_{22} \psi_{31} + \psi_{12} \psi_{23} \psi_{31} + \psi_{13} \psi_{21} \psi_{32} \\ &\quad - \psi_{11} \psi_{23} \psi_{32} - \psi_{12} \psi_{21} \psi_{33} + \psi_{11} \psi_{22} \psi_{33}, \\ f_2(\psi) &:= \sum_{i_1, i_2, j_1, j_2, k_1, k_2} \psi_{i_1 i_2} \psi_{j_1 j_2} \psi_{k_1 k_2} \epsilon_{i_1 i_2 j_1} \epsilon_{j_2 k_1 k_2} \\ &= \psi_{13}^2 \psi_{22} - \psi_{12} \psi_{13} \psi_{23} - \psi_{13} \psi_{21} \psi_{23} + \psi_{11} \psi_{23}^2 \\ &\quad - 2\psi_{13} \psi_{22} \psi_{31} + 3\psi_{12} \psi_{23} \psi_{31} - \psi_{21} \psi_{23} \psi_{31} + \psi_{22} \psi_{31}^2 \\ &\quad - \psi_{12} \psi_{13} \psi_{32} + 3\psi_{13} \psi_{21} \psi_{32} - 2\psi_{11} \psi_{23} \psi_{32} - \psi_{12} \psi_{31} \psi_{32} \\ &\quad - \psi_{21} \psi_{31} \psi_{32} + \psi_{11} \psi_{32}^2 + \psi_{12}^2 \psi_{33} \\ &\quad - 2\psi_{12} \psi_{21} \psi_{33} + \psi_{21}^2 \psi_{33}, \end{aligned} \tag{2.11}$$

which are non-zero, linearly independent and satisfy (2.4.1).

**The case of  $GL(4)$** 

In the case that  $n = 4, m = 2$ , the possible degrees of the invariants are

$$d = 2, 4, 6, \dots$$

and **Schur** gives

$$\begin{aligned} (\{1\} \times \{1\})^{\otimes\{2\}} &\supset \{1^4\}, \\ (\{1\} \times \{1\})^{\otimes\{4\}} &\supset 3\{2^4\}, \\ (\{1\} \times \{1\})^{\otimes\{6\}} &\supset 3\{3^4\}, \\ (\{1\} \times \{1\})^{\otimes\{8\}} &\supset 6\{4^4\}. \end{aligned}$$

The correct number of invariants can be constructed from three invariants of degree  $d = 2, 4, 4$  respectively:

$$\begin{aligned} f_1(\psi) &:= \sum_{i_1, i_2, j_1, j_2} \psi_{i_1 i_2} \psi_{j_1 j_2} \epsilon_{i_1 i_2 j_1 j_2}, \\ f_2(\psi) &:= \sum_{i_1, i_2, j_1, j_2, k_1, k_2, l_1, l_2} \psi_{i_1 i_2} \psi_{j_1 j_2} \psi_{k_1 k_2} \psi_{l_1 l_2} \epsilon_{i_1 j_1 k_1 l_1} \epsilon_{i_2 j_2 k_2 l_2}, \\ f_3(\psi) &:= \sum_{i_1, i_2, j_1, j_2, k_1, k_2, l_1, l_2} \psi_{i_1 i_2} \psi_{j_1 j_2} \psi_{k_1 k_2} \psi_{l_1 l_2} \epsilon_{i_1 i_2 j_1 k_1} \epsilon_{j_2 k_2 l_1 l_2}, \end{aligned} \quad (2.12)$$

which by explicit expansion (either by hand or using a computer algebra package) are non-zero, algebraically independent and satisfy (2.4.1).<sup>2</sup>

### 2.4.2 Invariants of $\times^m GL(n)$ on $V^{\otimes m}$

We consider the existence of invariants  $q : V^{\otimes m} \rightarrow \mathbb{C}$  which take the form

$$q(gx) = \det(g)^k q(x),$$

for all  $g = g_1 \otimes g_2 \otimes \dots \otimes g_m$  with  $g_a \in GL(n)$  for  $1 \leq a \leq m$ . We mimic the construction of the previous section and give sufficient conditions for the existence of such functions.

**Theorem 2.4.2.** *Consider a function  $q : (\times^m V^{\otimes n}) \times V^{\otimes m} \rightarrow \mathbb{C}$  which satisfies the conditions:*

1. *For fixed  $\psi \in V^{\otimes m}$  we have*

$$q(\chi_1, \dots, c\chi_a, \dots, \chi_m; \psi) = c^k q(\chi_1, \dots, \chi_a, \dots, \chi_m; \psi),$$

*for each  $1 \leq a \leq m$ .*

2. *For fixed  $\chi_a \in V^{\otimes n}$ ,  $1 \leq a \leq m$ , we have*

$$q(\chi_1, \dots, \chi_m; c\psi) = c^d q(\chi_1, \dots, \chi_m; \psi).$$

3. *For all  $g = g_1 \otimes g_2 \otimes \dots \otimes g_m$  we have*

$$q(\chi_1, \dots, \chi_m; g\psi) = q(\otimes^n g_1^t \chi_1, \dots, \otimes^n g_m^t \chi_m; \psi).$$

*The function  $q_\epsilon : V^{\otimes m} \rightarrow \mathbb{C}$  given by*

$$q_\epsilon(\psi) := q(\epsilon, \epsilon, \dots, \epsilon; \psi),$$

*satisfies  $q_\epsilon(g\psi) = \det(g)^k f_\epsilon(\psi)$  for all  $g = g_1 \otimes g_2 \otimes \dots \otimes g_m$ .*

---

<sup>2</sup>As the number of indices in these expressions is becoming prohibitively large, we will adopt a convention from now until the end of the thesis that, unless otherwise indicated, any indices that appear after a summation sign are to be summed over appropriate bounds.

*Proof.* We have

$$\begin{aligned}
q_\epsilon(g\psi) &= q(\epsilon, \dots, \epsilon; g\psi) \\
&= q(\otimes^n g_1^t \epsilon, \dots, \otimes^n g_m^t \epsilon; \psi) \\
&= q(\det(g_1)\epsilon, \dots, \det(g_m)\epsilon; \psi) \\
&= \det(g_1)^k \det(g_2)^k \dots \det(g_m)^k q(\epsilon, \dots, \epsilon; \psi) \\
&= \det(g)^k q_\epsilon(\psi).
\end{aligned}$$

□

With these sufficient conditions in mind we will use the Schur functions to ascertain existence of these invariants and give examples of their exact form.

### The case of $k = 1$

From (2.3.3) the existence of such invariants requires that for the  $m$ -fold inner product we have:

$$\{1^n\} * \{1^n\} * \dots * \{1^n\} \ni \{n\}.$$

Now for even  $m$  we have

$$\{1^n\} * \{1^n\} * \dots * \{1^n\} = \{n\}$$

and for odd  $m$

$$\{1^n\} * \{1^n\} * \dots * \{1^n\} = \{1^n\}.$$

So that there exists a single invariant for each even  $m$  and no invariants for odd  $m$ .

For  $m = 2$  and  $n = 2, 3, 4$  these invariants are

$$\begin{aligned}
\det_2(\psi) &= \sum \psi_{i_1 i_2} \psi_{j_1 j_2} \epsilon_{i_1 j_1} \epsilon_{i_2 j_2}, \\
\det_3(\psi) &= \sum \psi_{i_1 i_2} \psi_{j_1 j_2} \psi_{k_1 k_2} \epsilon_{i_1 j_1 k_1} \epsilon_{i_2 j_2 k_2}, \\
\det_4(\psi) &= \sum \psi_{i_1 i_2} \psi_{j_1 j_2} \psi_{k_1 k_2} \psi_{l_1 l_2} \epsilon_{i_1 j_1 k_1 l_1} \epsilon_{i_2 j_2 k_2 l_2},
\end{aligned} \tag{2.13}$$

which can be seen to satisfy (2.4.2) and can be generalized in the obvious manner for any  $n$ . (These polynomials should be distinguished from the determinant of a matrix; although their functional form is identical to that of the determinant, they arise as invariant functions on the linear space  $V \otimes V$ .)

For  $m = 4$  and  $n = 2, 3, 4$  we can define:

$$\begin{aligned}
Q_2(\psi) &= \sum \psi_{i_1 i_2 i_3 i_4} \psi_{j_1 j_2 j_3 j_4} \epsilon_{i_1 j_1} \epsilon_{i_2 j_2} \epsilon_{i_3 j_3} \epsilon_{i_4 j_4}, \\
Q_3(\psi) &= \sum \psi_{i_1 i_2 i_3 i_4} \psi_{j_1 j_2 j_3 j_4} \psi_{k_1 k_2 k_3 k_4} \epsilon_{i_1 j_1 k_1} \epsilon_{i_2 j_2 k_2} \epsilon_{i_3 j_3 k_3} \epsilon_{i_4 j_4 k_4}, \\
Q_4(\psi) &= \sum \psi_{i_1 i_2 i_3 i_4} \psi_{j_1 j_2 j_3 j_4} \psi_{k_1 k_2 k_3 k_4} \psi_{l_1 l_2 l_3 l_4} \epsilon_{i_1 j_1 k_1 l_1} \epsilon_{i_2 j_2 k_2 l_2} \epsilon_{i_3 j_3 k_3 l_3} \epsilon_{i_4 j_4 k_4 l_4},
\end{aligned} \tag{2.14}$$

which can also be seen to satisfy (2.4.2) and can be generalized in the obvious way for arbitrary  $n$ . We refer to these invariants as *quangles*.

**The case of  $\times^m GL(2)$  and  $k = 2$** 

For  $m = 2, 3, 4$  **Schur** shows that

$$\begin{aligned}\{2^2\} * \{2^2\} &\ni \{4\}, \\ \{2^2\} * \{2^2\} * \{2^2\} &\ni \{4\}, \\ \{2^2\} * \{2^2\} * \{2^2\} * \{2^2\} &\ni 3\{4\}.\end{aligned}$$

At  $m = 2$  the required invariant is the pointwise product of  $\det_2$  with itself. Whereas at  $m = 3$  we have the *tangle*<sup>3</sup>

$$\mathcal{T}_2(\psi) = \sum \psi_{i_1 i_2 i_3} \psi_{j_1 j_2 j_3} \psi_{k_1 k_2 k_3} \psi_{l_1 l_2 l_3} \epsilon_{i_1 j_1} \epsilon_{i_2 j_2} \epsilon_{k_1 l_1} \epsilon_{k_2 l_2} \epsilon_{i_3 l_3} \epsilon_{j_3 k_3}. \quad (2.15)$$

At  $m = 4$ , the pointwise product of  $Q_2$  with itself forms a  $k = 2$  invariant and we have the additional invariants:

$$\begin{aligned}\mathcal{I}_1 &:= \sum \psi_{i_1 i_2 i_3 i_4} \psi_{j_1 j_2 j_3 j_4} \psi_{k_1 k_2 k_3 k_4} \psi_{l_1 l_2 l_3 l_4} \epsilon_{i_1 j_1} \epsilon_{i_2 j_2} \epsilon_{k_1 l_1} \epsilon_{k_2 l_2} \epsilon_{i_3 k_3} \epsilon_{i_4 k_4} \epsilon_{j_3 l_3} \epsilon_{j_4 l_4}, \\ \mathcal{I}_2 &:= \sum \psi_{i_1 i_2 i_3 i_4} \psi_{j_1 j_2 j_3 j_4} \psi_{k_1 k_2 k_3 k_4} \psi_{l_1 l_2 l_3 l_4} \epsilon_{i_1 j_1} \epsilon_{i_2 l_2} \epsilon_{i_3 l_3} \epsilon_{i_4 k_4} \epsilon_{j_2 k_2} \epsilon_{j_3 k_3} \epsilon_{j_4 l_4} \epsilon_{k_1 l_1},\end{aligned}$$

which satisfy (2.4.2) and can be shown to be non-zero and algebraically independent.

**The case of  $GL(3)^{\times m}$  and  $k = 2$** 

For  $m = 2, 3, 4$  **Schur** shows that

$$\begin{aligned}\{2^3\} * \{2^3\} &\ni \{6\}, \\ \{2^3\} * \{2^3\} * \{2^3\} &\ni \{6\}, \\ \{2^3\} * \{2^3\} * \{2^3\} * \{2^3\} &\ni 4\{6\}.\end{aligned}$$

At  $m = 2$ , the pointwise product of  $\det_3$  with itself forms a  $k = 2$  invariant and at  $m = 3$  the tangle can be generalized to the  $n = 3$  case:

$$\begin{aligned}\mathcal{T}_3(\psi) &= \sum \psi_{i_1 i_2 i_3} \psi_{j_1 j_2 j_3} \psi_{k_1 k_2 k_3} \psi_{l_1 l_2 l_3} \psi_{m_1 m_2 m_3} \psi_{n_1 n_2 n_3} \\ &\quad \cdot \epsilon_{i_1 j_1 k_1} \epsilon_{j_2 k_2 l_2} \epsilon_{k_3 l_3 m_3} \epsilon_{l_1 m_1 n_1} \epsilon_{m_2 n_2 i_2} \epsilon_{n_3 i_3 j_3},\end{aligned} \quad (2.16)$$

which by explicit expansion can be shown to be non-zero.

The invariants at  $m = 4$  remain uninvestigated.

---

<sup>3</sup>The tangle is known and used in physics to analyse multiparticle entanglement in quantum mechanics. This will be reviewed in Chapter 3

**The case of  $\times^m GL(4)$  and  $k = 2$**

For  $m = 2, 3, 4$  **Schur** shows that

$$\begin{aligned}\{2^4\} * \{2^4\} &\supset \{8\}, \\ \{2^4\} * \{2^4\} * \{2^4\} &\supset \{8\}, \\ \{2^4\} * \{2^4\} * \{2^4\} * \{2^4\} &\supset 7\{8\}.\end{aligned}$$

At  $m = 2$  the pointwise product of  $\det_4$  with itself is a  $k = 2$  invariant and at  $m = 3$  the tangle can again be generalized:

$$\begin{aligned}\mathcal{T}_4 = \sum & \psi_{i_1 j_1 k_1} \psi_{i_2 j_2 k_2} \psi_{i_3 j_3 k_3} \psi_{i_4 j_4 k_4} \psi_{i_5 j_5 k_5} \psi_{i_6 j_6 k_6} \psi_{i_7 j_7 k_7} \psi_{i_8 j_8 k_8} \\ & \cdot \epsilon_{i_1 i_2 i_3 i_4} \epsilon_{i_5 i_6 i_7 i_8} \epsilon_{j_1 j_5 j_4 j_8} \epsilon_{j_2 j_6 j_3 j_7} \epsilon_{k_1 k_5 k_2 k_6} \epsilon_{k_3 k_7 k_4 k_8},\end{aligned}\tag{2.17}$$

and shown to be non-zero by explicit expansion.  
The invariants at  $m = 4$  remain uninvestigated.

## 2.5 Closing remarks

In this chapter we have given a review of the use of the character theory to build the irreducible representations of the general linear group. We have demonstrated the concrete connection between the one-dimensional representations and the classical invariants, and have presented theorems that allow us to count these invariants at given degree  $d$  and weight  $k$ .



## CHAPTER 3

# Entanglement and phylogenetics

Stochastic methods that model character distributions in aligned sequences are part of the standard armoury of phylogenetic analysis [19, 21, 44, 51, 54]. The evolutionary relationships are usually represented as a bifurcating tree directed in time. It is remarkable that there is a strong conceptual and mathematical analogy between the construction of phylogenetic trees using stochastic methods, and the process of scattering in particle physics [31]. It is the purpose of the present chapter to show that there is much potential in taking an algebraic, group theoretical approach to the problem where the inherent symmetries of the system can be fully appreciated and utilized.

Entanglement is of considerable interest in physics and there has been much effort to elucidate the nature of this curious physical phenomenon [8, 14, 26, 38, 62]. Entanglement has its origin in the manner in which the state probabilities of a quantum mechanical system must be constructed from the individual state probabilities of its various subsystems. Whenever there are global conserved quantities, such as spin, there exist entangled states where the choice of measurement of one subsystem can affect the measurement outcome of another subsystem no matter how spatially separated the two subsystems are. This curious physical property is represented mathematically by nonseparable tensor states. Remarkably, if the pattern frequencies of phylogenetic analysis are interpreted in a tensor framework it is possible to show that the branching process itself introduces entanglement into the state. In the context of phylogenetics this element of entanglement corresponds to nothing other than that of phylogenetic relation. This is a mathematical curiosity that can be studied using methods from quantum physics. This is a novel way of approaching phylogenetic analysis which has not been explored before.

This chapter will begin by establishing the formalism of quantum mechanics and introducing the concept of entanglement through an elementary example. A short review of the use of group invariant functions to analyse entanglement will be presented. The stochastic model of a phylogenetic tree will then be developed in its standard form, followed by a discussion which establishes a presentation of this model in the form of a group action on a tensor product space as used in quantum mechanics. The invariant functions used to study entanglement will then be examined in the context of phylogenetic trees.

**Note:** Elements of this chapter are extracted from [59].

## 3.1 Quantum mechanics

The formalism of the quantum mechanical description of physical systems amounts to four fundamental postulates.

**Postulate 1.** The mathematical description of any physical system occurs as a *state vector*  $\psi$  in a complex vector space,  $V$ , together with an inner product known as a *Hilbert space*  $\mathcal{H} = (V, \cdot)$ .

For a given physical system it is not *a priori* apparent exactly how the Hilbert space should be chosen. As will be elaborated later, a basic property of quantum mechanics is that it is not possible to determine (in practice or in principle) the exact and complete configuration of a physical system. Thus, the Hilbert space is chosen not to represent all possible configurations of the system, but rather to represent whichever part is observable and under consideration in a given experimental setup. For example the full description of an electron is given by the tensor product of the representation space of the spin,  $\mathbb{C}^2$ , with that of the representation space of spatial position, square integrable functions  $\{f : \mathbb{R}^3 \rightarrow \mathbb{C}\}$ . However, one is often only interested in the spin degrees of freedom of the system and simply ignores the position component of the state vector.

For our purposes it will be enough to consider only the case where  $\mathcal{H}$  is the finite dimensional vector space with inner product given in terms of notation from Chapter 2 as

$$(\psi, \varphi) = \overline{\psi}(\varphi).$$

**Postulate 2.** The dynamical evolution of any physical system is governed by the linear equation

$$i\hbar \frac{\partial \psi(t)}{\partial t} = H(t)\psi(t), \quad (3.1)$$

where  $\hbar$  is Planck's constant and  $H(t)$  is a *Hermitian* operator:

$$(\psi, H\varphi) = (H\psi, \varphi),$$

known as the *Hamiltonian*. Completely equivalently, the dynamical evolution is described by solutions of (3.1):

$$\psi(t_2) = U(t_2, t_1)\psi(t_1),$$

where  $U(t_2, t_1)$  is a unitary operator

$$(U\psi, U\varphi) = (\psi, \varphi).$$

From this postulate it is not apparent how the Hamiltonian should be chosen in any particular case. Historically, Dirac formalized the idea of *classical analogy* where the Hamiltonian is interpreted as the total energy of the system [13]. However, this procedure is limited to systems which have a classical counterpart and the general case is left to the modern quantum physicist.

**Postulate 3.** An *observable* of a physical system is described by an Hermitian operator  $A$  with associated eigenvalues  $\{\alpha_1, \alpha_2, \dots\}$  and eigenspaces defined by the projection operators  $\{P_1, P_2, \dots\}$ . If the state vector before measurement is  $\psi$ , then the probability of the result  $\alpha_i$  is given by

$$\frac{(\psi, P_i \psi)}{(\psi, \psi)},$$

and the state after measurement is  $\psi' = P_i \psi$ .

From this definition it is apparent that  $U$  must be unitary to preserve total probability. We will follow the standard procedure of *normalizing* the state vector:

$$(\psi, \psi) = 1.$$

**Postulate 4.** The state space of a composite of  $m$  quantum systems with individual state spaces  $\mathcal{H}_1, \mathcal{H}_2, \dots, \mathcal{H}_m$  is given by the tensor product:

$$\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \dots \otimes \mathcal{H}_m.$$

From this definition it may seem that the state vector of a composite system should be expressed as the *product state*

$$\psi = \varphi^{(1)} \otimes \varphi^{(2)} \otimes \dots \otimes \varphi^{(m)}, \quad (3.2)$$

where  $\varphi^{(a)} \in \mathcal{H}_a$  is the state vector of each individual system. However, for the general case, there are physical reasons why there must exist states which cannot be written in the form (3.2). We will explore these states and their curious properties in the next section.

### 3.1.1 Spin $\frac{1}{2}$ and entanglement

One way to proceed in the search for the appropriate state space  $\mathcal{H}$  is to study the representation spaces of the irreducible representations of a symmetry group of a physical system. For the case of three dimensional Euclidean space, consider the symmetry group of proper rotations; the special orthogonal group  $SO(3)$ . The irreducible representations of  $SO(3)$  are labelled by the *spin* quantum numbers  $s = \{0, \frac{1}{2}, 1, \frac{3}{2}, 2, \frac{5}{2}, \dots\}$  (see [42]). Here we will study the case  $s = \frac{1}{2}$  where the representation is two-dimensional:  $\mathcal{H} = \mathbb{C}^2$ , and a state vector is referred to as a *qubit*. The physics of the spin of a qubit is captured

by considering an orthonormal basis for  $\mathbb{C}^2$  as  $\{z_+, z_-\}$  and introducing the observable  $S_z$  satisfying

$$S_z z_+ = \frac{\hbar}{2} z_+, \quad S_z z_- = -\frac{\hbar}{2} z_-,$$

so that the states  $\psi^+ := z_+$  and  $\psi^- := z_-$  are eigenvectors of the spin operator. Analogously, we can define the  $x$  basis  $\{x_+, x_-\}$  (or any other orthonormal basis) by rotating the  $z$  basis using the group element of the two-dimensional representation of  $SO(3)$  which corresponds to the appropriate physical rotation. In particular, we have

$$x_+ = \frac{1}{\sqrt{2}}(z_+ + z_-), \quad x_- = \frac{1}{\sqrt{2}}(x_+ - x_-).$$

The measurement operators are then defined as being the projection operators onto the appropriate basis vectors. For instance the projection operators for spin in the  $z$  direction satisfy

$$P_z^+ z_+ = z_+, \quad P_z^+ z_- = 0, \quad P_z^- z_+ = 0, \quad P_z^- z_- = z_-.$$

A generic qubit can be written as

$$\psi = c_1 z_+ + c_2 z_-.$$

Introduce the random variables  $A_z \in \{+1, -1\}$  to correspond to the value of the spin along the  $z$  axis, and we have

$$\mathbb{P}(A_z = 1) = (\psi, P_z^+ \psi) = |c_1|^2,$$

and

$$\mathbb{P}(A_z = -1) = (\psi, P_z^- \psi) = |c_2|^2.$$

Now we turn our attention to composite states of  $m$  qubits where the state space becomes

$$\mathcal{H} = (\mathbb{C}^2)^{\otimes m}.$$

The most general state can be expressed as

$$\psi = \sum \psi_{i_1 i_2 \dots i_m} e_{i_1} \otimes e_{i_2} \otimes \dots \otimes e_{i_m},$$

so that the state is specified by  $2^m$  complex numbers  $\psi_{i_1 i_2 \dots i_m}$ . In the case where  $\psi$  can be expressed in the form of a product state, we have

$$\psi_{i_1 i_2 \dots i_m} = \varphi_{i_1}^{(1)} \varphi_{i_2}^{(2)} \dots \varphi_{i_m}^{(m)},$$

and we see that the state is specified by  $2m$  complex numbers. The difference in these parameter counts between the general state and the product state is the origin of *entanglement*.

To illustrate the simplest example of entanglement consider the case of a spin

zero particle splitting into two spin  $\frac{1}{2}$  qubits labelled as  $A$  and  $B$ . To ensure that the total spin is zero, it must be the case that the total state is

$$\psi = \frac{1}{\sqrt{2}}(z_+ \otimes z_- - z_- \otimes z_+),$$

which ensures that  $(S_z \otimes 1 + 1 \otimes S_z)\psi = 0$ . We introduce the random variables  $A_z$  for particle  $A$  and  $B_z$  for particle  $B$ . For the state,  $\psi$ , the measurement of spins of  $A$  and  $B$  along the  $z$  axis is associated with the probabilities

$$\begin{aligned}\mathbb{P}(A_z = 1, B_z = 1) &= \mathbb{P}(A_z = -1, B_z = -1) = 0, \\ \mathbb{P}(A_z = 1, B_z = -1) &= \mathbb{P}(A_z = -1, B_z = 1) = 1/2,\end{aligned}$$

$$\mathbb{P}(A_z = 1) = \mathbb{P}(A_z = -1) = 1/2,$$

and

$$\mathbb{P}(B_z = 1) = \mathbb{P}(B_z = -1) = 1/2.$$

Now if we consider the same state but with spin measurements taken along the  $x$  axis, it is a simple exercise to show that

$$\psi = \frac{1}{\sqrt{2}}(x_+ \otimes x_- - x_- \otimes x_+).$$

Now if we were to go ahead and compute the various probabilities associated with the observable  $S_x$  we would come to the same probabilities as above. That is, the spins of  $A$  and  $B$  are always opposite to give  $A_x = 1, B_x = -1$  with probability  $\frac{1}{2}$  and  $A_x = -1, B_x = +1$  with probability  $\frac{1}{2}$ . One can go further and show that this is true for *any* orthonormal basis of  $\mathbb{C}^2$ . This implies that no matter which axis the spins are measured along, the outcome at  $A$  is always the negative of the outcome at  $B$ . These probabilities have been amply confirmed by experiment.

A problem arises if one wishes to interpret the probabilities of the formalism of quantum mechanics as representing our ignorance of the full state of the physical system. Such a description of these events would require that at the moment of splitting, each particle actually carries the requisite information as how to respond to a spin measurement on an arbitrary axis, and somehow this information is unobservable or hidden from us. This additional information over and above the state vector was historically coined the *hidden variables*. However, Bell showed that it is actually impossible to specify the required hidden variables [7] and thus it is not possible to interpret the probabilities as simply representing our ignorance of the system. This implies that quantum mechanics requires that the physical world is probabilistic in an intrinsic way. An alternative way out of this predicament is to assume that there is a non-local communication between particles  $A$  and  $B$ , which ensures that spins are opposite along any axis. However, at the moment of measurement,  $A$  and  $B$  could be separated by a very large distance! Thus the entanglement leads us to the dilemma of having to accept one of the following:

- Quantum systems have an essentially non-local property.
- The probabilities in quantum mechanics do not just indicate our ignorance of the configuration of a physical system, but are an essential part of physical reality.

Einstein was unhappy with both options, and never made his peace with the quantum theory that he was so instrumental in constructing. This is because the first violates the spirit, if not the detail of special relativity grossly, and the second implies that Einstein's contention that "God does not play dice" cannot be true.

Recall that the conditional probability that the random variable  $A = x$  given that  $B = y$  is defined to be

$$\mathbb{P}(A = x|B = y) := \frac{\mathbb{P}(A = x, B = y)}{\mathbb{P}(B = y)}.$$

The random variables  $A$  and  $B$  are said to be *stochastically independent* [18] if and only if

$$\mathbb{P}(A = x, B = y) = \mathbb{P}(A = x)\mathbb{P}(B = y),$$

from which it would follow that

$$\mathbb{P}(A = x|B = y) = \mathbb{P}(A = x),$$

which motivates the definition. (This notion of stochastic independence can be extended to multiple random variables. For details see Feller [18].)

In quantum mechanics, stochastic independence is implied if the state is a product  $\psi = \varphi^{(1)} \otimes \varphi^{(2)}$ . For if the state is a product state, we have

$$\begin{aligned} \mathbb{P}(A = i, B = j) &= \overline{\varphi^{(1)}}(P_i \varphi^{(1)}) \overline{\varphi^{(2)}}(P_j \varphi^{(2)}) \\ &:= \mathbb{P}(A = i) \mathbb{P}(B = j). \end{aligned}$$

In what follows we will equate entanglement with this notion of stochastic dependence.

### 3.1.2 Orbit classes and invariants

We have seen that a quantum system exhibits entanglement if the state vector cannot be written as a product. Mathematically one would like to partition the set of entangled state vectors into equivalence classes which capture the essential property of entanglement. A systematic approach to the classification problem is to study the orbit classes of the tensor product space under a group action which is designed to preserve the essential non-local properties of entanglement. The orbit of an element  $\psi \in \mathcal{H}$  under the group action  $G$  is defined as the set of elements  $\{\psi' = g\psi \text{ for some } g \in G\}$ .

In quantum physics the appropriate group action is known to be the set of

SLOCC operators, (Stochastic Local Operations with Classical Communication) [14, 26, 38, 43, 45]. Mathematically SLOCC operators correspond to the ability to transform the individual parts of the tensor product space  $\mathcal{H} \cong \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \dots \otimes \mathcal{H}_m$  with arbitrary invertible, linear operations. These operators are expressed by group elements of the form

$$g = g_1 \otimes g_2 \otimes \dots \otimes g_m,$$

where  $m$  is the number of individual spaces making up the tensor product, and  $g_i \in GL(\mathcal{H}_i)$ .

The task is to identify the orbit classes of a given tensor product space under the general set of SLOCC operators. A powerful tool in this analysis is the construction of the invariant functions  $\mathbb{C}[\mathcal{H}]^G$ . By definition these invariants are relatively constant up to the determinant upon each orbit class of  $\mathcal{H}$ . It can be shown that there exists (under the action of the general linear group at least), a *finite* set of elements which generate the full set of invariants on a given linear space. It can also be shown that the set of orbit classes of a given linear space can be completely classified given a full set of invariants on that space [46].

In what follows we study the orbit class problem for the state space of two qubits and then that of three qubits.

### 3.1.3 Two qubits and the concurrence

Using the notation of Chapter 2, the concurrence is defined using (2.13):

$$\mathcal{C} = \det_2,$$

so that

$$\mathcal{C}(\psi) = \sum \psi_{i_1 i_2} \psi_{j_1 j_2} \epsilon_{i_1 j_1} \epsilon_{i_2 j_2}.$$

We wish to construct the orbit classes of  $\mathcal{H} = \mathbb{C}^2 \otimes \mathbb{C}^2$  under the group action  $GL(\mathbb{C}^2) \times GL(\mathbb{C}^2)$ . Any state  $\psi \in \mathcal{H}$  can be expressed using the four parameters  $\psi_{i_1 i_2}$  which in turn can be arranged as a matrix  $M = [\psi_{i_1 i_2}]$ . Under the group transformation

$$\psi \rightarrow g_1 \otimes g_2 \psi,$$

the corresponding matrix transformation is

$$M \rightarrow g_1 M g_2^t.$$

Hence we can answer the orbit class problem by taking a canonical  $2 \times 2$  matrix  $X$  and considering the set of matrices  $\{M = AXB; A, B \in GL(\mathbb{C}^2)\}$ .

**Theorem 3.1.1.** *The vector space  $V \otimes V$  where  $V \equiv \mathbb{C}^2$  has three orbits under the group action  $GL(2) \times GL(2)$ . Under the identification  $M = [\psi_{i_1 i_2}]$  for all  $\psi$  the orbits are characterized by the following canonical forms:*

- (i) Null-orbit  $X = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ ,
- (ii) Separable-orbit  $Y = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ ,
- (iii) Entangled-orbit  $Z = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ .

The separable and entangled-orbits can be distinguished by the determinant function.

*Proof.* (i) The null-orbit has only one member, the null vector; it is of course unchanged by the group action.

(ii) We are required to show that the set of  $2 \times 2$  matrices  $\mathcal{M} = \{S : S = AYB; A, B \in GL(V)\}$  is all matrices such that  $\det(S) = 0$ . We begin by taking a general member of  $\mathcal{M}$ ,  $S = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$  with  $ad - bc = 0$ . Clearly the matrices

$$S' := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} S = \begin{pmatrix} c & d \\ a & b \end{pmatrix}, \quad S'' := S \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} b & a \\ d & c \end{pmatrix}, \quad \text{and}$$

$$S''' := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} S \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} d & c \\ b & a \end{pmatrix}$$

also belong to  $\mathcal{M}$ . So without loss of generality we can take  $a \neq 0$  and it is an easy computation to show that

$$S = \begin{pmatrix} 1 & 0 \\ c/a & 1 \end{pmatrix} Y \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix},$$

so that  $\mathcal{M}$  is the set of  $2 \times 2$  matrices with vanishing determinant.

(iii) Clearly any  $2 \times 2$  matrix  $N$  with non-zero determinant can be written as  $N = AZB$  where  $A, B \in GL(\mathbb{C}^2)$ .  $\square$

**Corollary 3.1.2.** *The orbits of  $\mathcal{H} = \mathbb{C}^2 \otimes \mathbb{C}^2$  under  $SL(\mathbb{C}^2) \times SL(\mathbb{C}^2)$  are labelled by the determinant function  $\det[\phi(h)]$ .*

For further discussion see [8, 14, 38].

### 3.1.4 Three qubits and the tangle

It is known that there are six orbit classes of  $\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2$  under the action  $GL(\mathbb{C}^2) \times GL(\mathbb{C}^2) \times GL(\mathbb{C}^2)$ . These orbits classes can be distinguished by functions of the concurrence and another relative invariant known as the tangle



[14, 26].

We begin by defining three partial concurrence operations as

$$\begin{aligned}\mathcal{C}_1(\psi) &= \sum \psi_{ijk} \psi_{lmn} \epsilon_{jm} \epsilon_{kn} e_i \otimes e_l, \\ \mathcal{C}_2(\psi) &= \sum \psi_{ijk} \psi_{lmn} \epsilon_{il} \epsilon_{kn} e_j \otimes e_m, \\ \mathcal{C}_3(\psi) &= \sum \psi_{ijk} \psi_{lmn} \epsilon_{il} \epsilon_{jm} e_k \otimes e_n.\end{aligned}\tag{3.3}$$

From these definitions it is easy to see that

$$\begin{aligned}\mathcal{C}_1(\psi') &:= \mathcal{C}_1(g_1 \otimes g_2 \otimes g_3 \psi) \\ &= [\det(g_2) \det(g_3)] g_1 \otimes g_1 \mathcal{C}_1(\psi),\end{aligned}$$

with similar expressions for  $\mathcal{C}_2$  and  $\mathcal{C}_3$ .

The tangle is an invariant satisfying

$$\mathcal{T}(\psi) = [\det(g_1) \det(g_2) \det(g_3)]^2 \mathcal{T}(\psi),$$

and from (2.15) can be written in the form

$$\mathcal{T}(\psi) = \sum \psi_{a_1 a_2 a_3} \psi_{b_1 b_2 b_3} \psi_{c_1 c_2 c_3} \psi_{d_1 d_2 d_3} \epsilon_{a_1 b_1} \epsilon_{a_2 b_2} \epsilon_{c_1 d_1} \epsilon_{c_2 d_2} \epsilon_{b_3 c_3} \epsilon_{a_3 d_3}.$$

The six orbit classes are described by the completely disentangled states

$$\psi = \varphi^{(1)} \otimes \varphi^{(2)} \otimes \varphi^{(3)}, \quad \varphi^{(a)} \in \mathbb{C}^2;$$

the partially entangled states which form three orbit classes characterized by the separability of the canonical tensors

$$\begin{aligned}\psi_p^{(1)} &= \sum \varphi_i^{(1)} \varphi_{jk}^{(23)} e_i \otimes e_j \otimes e_k, \\ \psi_p^{(2)} &= \sum \varphi_{ik}^{(13)} \varphi_j^{(2)} e_i \otimes e_j \otimes e_k, \\ \psi_p^{(3)} &= \sum \varphi_{ij}^{(12)} \varphi_k^{(3)} e_i \otimes e_j \otimes e_k;\end{aligned}$$

the completely entangled states equivalent to the *GHZ* state

$$\psi_{ghz} = \frac{1}{\sqrt{2}}(e_0 \otimes e_0 \otimes e_0 + e_1 \otimes e_1 \otimes e_1);$$

and the completely entangled states equivalent to the *W* state

$$\psi_w = \frac{1}{\sqrt{3}}(e_0 \otimes e_0 \otimes e_1 + e_0 \otimes e_1 \otimes e_0 + e_1 \otimes e_0 \otimes e_0).$$

The tangle and the concurrence and its partial counterparts can be used to fully distinguish these orbit classes. For the completely disentangled tensors we have

$$\mathcal{C}_a(\psi) = 0, \quad \mathcal{T}(\psi) = 0,$$

for all  $a = 1, 2, 3$ . Whereas for the first partially entangled state we have

$$\mathcal{C}_1(\psi_p^{(1)}) \neq 0, \quad \mathcal{C}_2(\psi_p^{(1)}) = \mathcal{C}_3(\psi) = 0, \quad \mathcal{T}(\psi_p^{(1)}) = 0,$$

and similar relations for the remaining two partially entangled states. States in the  $GHZ$  orbit satisfy

$$\mathcal{C}_a(\psi_{ghz}) \neq 0, \quad \mathcal{T}(\psi_{ghz}) \neq 0$$

for all  $a = 1, 2, 3$ . Whereas states on the  $W$  orbit satisfy

$$\mathcal{C}_a(\psi_w) \neq 0, \quad \mathcal{T}(\psi_w) = 0$$

for all  $a = 1, 2, 3$ .

Notice that the  $GHZ$  and  $W$  orbits characterize different classes of three qubit entanglement. In the  $GHZ$  orbit each qubit is entangled with the other two qubits *and* the three qubits are entangled as a triplet. In the  $W$  orbit the qubits are entangled as pairs but are *not* entangled as a triplet.

## 3.2 Stochastic evolution of biomolecular units

It is standard to model sequence evolution as a stochastic process. A discrete set  $\mathcal{K}$  is associated with biomolecular units which we refer to as *bases* and define  $n := |\mathcal{K}|$ . For example, in the case of DNA sequences made up of the four nucleotides adenine, cytosine, guanine, thymine, we have  $\mathcal{K} = \{A, G, C, T\}$  and  $n = 4$ . The instance of a particular base in the sequence is equated with the time dependent random variable  $X(t) \in \mathcal{K}$  and the stochastic time evolution is modelled as a continuous time Markov chain (CTMC) so that

$$\frac{d}{dt} \mathbb{P}(X(t) = i) = \sum_j q_{ij}(t) \mathbb{P}(X(t) = j), \quad i, j \in \mathcal{K}. \quad (3.4)$$

The  $q_{ij}(t)$  are called *rate parameters* and must satisfy the relations

$$q_{ij}(t) \geq 0, \quad \forall i \neq j; \quad q_{ii}(t) = - \sum_{j \neq i} q_{ji}(t). \quad (3.5)$$

Define  $Q(t) = [q_{ij}(t)]_{(i,j \in \mathcal{K})}$  as the *rate matrix* associated with the Markov chain. The Markov chain is called *homogeneous* if the rate matrix is time independent. The results presented in this thesis are equally valid for inhomogeneous models where the rate matrix is time dependent and so we allow for this generality throughout. It is also common to impose further symmetries upon the rate matrix such as the Jukes Cantor and Kimura 3ST models [44]. However, the results presented here are again valid for any rate matrix satisfying (3.5), and hence no restriction upon the rate parameters is made. This

model is referred to as the *general Markov model* [1].

For notational simplicity we will write  $\pi_i(t) := \mathbb{P}(X(t) = i)$  and, given an initial distribution  $\pi_i(0)$ , write solutions of (3.4) as

$$\pi_i(t) = \sum_{j \in \mathcal{K}} m_{ij}(t, s) \pi_j(s), \quad 0 \leq s < t;$$

where  $m_{ij}(t, s) := \mathbb{P}(X(t) = i | X(s) = j)$  are the *transition probabilities* of the chain. We define the matrix  $M(t, s) = [m_{ij}(t, s)]_{(i, j \in \mathcal{K})}$  such that in the homogeneous case the transition probabilities only depend on the difference  $(t - s)$  and can be represented in terms of the rate matrix as

$$M(t, s) = M(t - s, 0) = e^{Q[(t-s)]} := \sum_{n=0}^{\infty} \frac{Q^n [(t-s)]^n}{n!}.$$

In the inhomogeneous case there are several representations available for the matrix of transition probabilities (for details see [29, 50]). The representation that is of most use to us here is the time-ordered product:

$$M(t, s) = \mathbb{T} \exp \int_s^t Q(u) du \quad (3.6)$$

(see for example [30] for the definition of the time-ordering operator  $\mathbb{T}$ .) For sufficiently small  $\delta t$ , we can write this in the approximate form

$$\begin{aligned} M(t, s) &\simeq M(t, t - \delta t) \dots M(s + 2\delta t, s + \delta t) M(s + \delta t, s) \\ &= e^{Q(t-\delta t)\delta t} e^{Q(t-2\delta t)\delta t} \dots e^{Q(s+\delta t)\delta t} e^{Q(s)\delta t}. \end{aligned}$$

From these solutions it is clear that

$$\det[M(t, s)] = \exp \int_s^t \text{tr}[Q(u)] du. \quad (3.7)$$

A more fundamental way to define the transition matrices of a CTMC is to impose the *backward* and *forward* Kolmogorov equations [29]:

$$\begin{aligned} \frac{\partial M(t, s)}{\partial s} &= -M(t, s)Q(s), \\ \frac{\partial M(t, s)}{\partial t} &= Q(t)M(t, s). \end{aligned} \quad (3.8)$$

### 3.3 Phylogenetic trees

The remaining task is to model the case of phylogenetically related molecular sequences evolving under a stochastic process. Effectively the model consists of multiple copies of the random variable  $X(t)$  taken as a generalization (via a tree structure) of a cartesian product and then modelled collectively as a

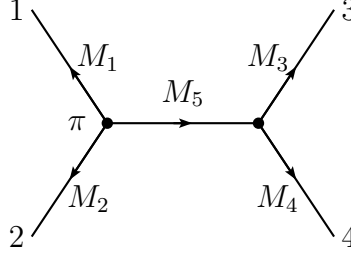


Figure 3.1: Phylogenetic tree of four taxa

CTMC. The reader is referred to [53] for a more extended discussion of the model. Here we keep the presentation to a minimum while allowing for the introduction of some essential notation and concepts.

A tree,  $T$ , is a connected graph without cycles and consists of a set of vertices,  $V$ , and edges,  $E$ . Vertices of degree one are called *leaves* and we partition the set of vertices as  $V = L \cup N$  where  $L$  is the set of leaves and  $N$  is the set of internal vertices. We work with *orientated* trees, which are defined by directing each edge of  $T$  away from a distinguished vertex,  $\pi$ , known as the *root* of the tree. Consequently, a given edge lying between vertices  $u$  and  $v$  is specified as an ordered pair  $e = (u, v)$ , where  $u$  lies on the (unique) path between  $v$  and  $\pi$ . The general Markov model of a phylogenetic tree is then made by assigning a set of random variables  $\{X_s, s \in V\}$  to the vertices of the tree; these random variables are assumed to be conditionally independent and individually satisfy the properties of a CTMC. Taking a distribution at the root of the tree,  $\{\mathbb{P}(X_\pi = i) := \pi_i, i \in \mathcal{K}\}$ , completes the specification of the phylogenetic tree. The interpretation of a phylogenetic tree is that the probability distribution at each leaf is associated with the observed sequence of a single taxon and the joint probability distribution across a number of leaves is associated with the aligned sequences of the same number of molecular sequences.

For example in Figure 3.1 we present the tree consisting of four leaves which has probability distribution

$$p_{i_1 i_2 i_3 i_4} = \sum_{j,k} m_{i_1 j}^{(1)} m_{i_2 j}^{(2)} m_{i_3 k}^{(3)} m_{i_4 k}^{(4)} m_{k j}^{(5)} \pi_j,$$

where

$$p_{i_1 i_2 i_3 i_4} := \mathbb{P}(X_1 = i_1, X_2 = i_2, X_3 = i_3, X_4 = i_4),$$

and we refer to these quantities as *pattern probabilities*.

### 3.4 Tensor presentation

Setting  $\mathbb{P}(X(t) = i) = p_i(t)$ , we introduce the  $n$ -dimensional vector space  $V$  with preferred basis  $\{e_1, e_2, \dots, e_n\}$  and associate the probabilities uniquely with the vector

$$p(t) = p_1(t)e_1 + p_2(t)e_2 + \dots + p_n(t)e_n.$$

The time evolution of this vector is then governed by equation (3.4) written in operator form as

$$\frac{d}{dt}p(t) = Q(t)p(t).$$

The solution of this equation is written as

$$p(t) = M(t, s)p(s).$$

The probabilities can be recovered by taking the inner product

$$p_i(t) = (e_i, p(t)),$$

and defining

$$\theta = \sum_{i=1}^n e_i, \tag{3.9}$$

we have

$$(\theta, p(t)) = 1, \quad \forall t.$$

In analogy we label the joint probabilities as

$$p_{i_1 i_2 \dots i_m}(t) := \mathbb{P}(X_1 = i_1, X_2 = i_2, \dots, X_m = i_m; t),$$

and by introducing the tensor product space  $V^{\otimes m}$  we associate these probabilities with the unique tensor

$$P(t) := \sum p_{i_1 i_2 \dots i_m}(t) e_{i_1} \otimes e_{i_2} \otimes \dots \otimes e_{i_m}.$$

Again the probabilities are recovered from the inner product:

$$p_{i_1 i_2 \dots i_m}(t) = (e_{i_1} \otimes e_{i_2} \otimes \dots \otimes e_{i_m}, P(t)),$$

and we define  $\Omega = \sum e_{i_1} \otimes e_{i_2} \otimes \dots \otimes e_{i_m}$  so that

$$(\Omega, P(t)) = 1, \quad \forall t.$$

We now introduce the branching events into this formalism.

Consider a vertex on a phylogenetic tree where the stochastic evolution of a single random variable branches into that of two random variables. The corresponding mathematical operation is a mapping  $V \rightarrow V \otimes V$ . In order to formalize this we introduce the *branching* operator  $\delta : V \rightarrow V \otimes V$ . The most general action of a (linear) operator  $\delta$  upon the basis elements of  $V$  can be expressed as

$$\delta e_i = \sum_{j,k} \Gamma_i^{jk} e_j \otimes e_k, \tag{3.10}$$

where  $\Gamma_i^{jk}$  are an arbitrary set of coefficients set by the assumption of conditional independence across branches of the tree.

To this end it is only necessary to consider initial probability distributions of the form

$$\begin{aligned}\pi^{(\gamma)} &= \sum_i \delta_i^\gamma e_i, \\ \gamma &= 1, 2, \dots, n.\end{aligned}$$

Directly subsequent to the branching event the two leaf state is given by

$$\begin{aligned}P^{(\gamma)} &= \delta \pi^{(\gamma)}, \\ &= \sum_{i,j,k} \delta_i^\gamma \Gamma_i^{jk} e_j \otimes e_k.\end{aligned}$$

We implement the conditional independence upon the branches by setting

$$\begin{aligned}\mathbb{P}(X_1 = i_1, X_2 = i_2, t = t' | X_1 = X_2 = \gamma, t = 0) \\ = \mathbb{P}(X_1 = i_1, t = t' | X_1 = \gamma, t = 0) \mathbb{P}(X_2 = i_2, t = t' | X_2 = \gamma, t = 0).\end{aligned}\quad (3.11)$$

Using the tensor formalism the transition probabilities can be expressed as

$$\begin{aligned}\mathbb{P}(X_1 = i_1, t = t' | X_1 = \gamma, t = 0) &= \sum_{k_1} m_{i_1 k_1}^{(1)}(t') \delta_{k_1}^\gamma, \\ \mathbb{P}(X_2 = i_2, t = t' | X_2 = \gamma, t = 0) &= \sum_{k_2} m_{i_2 k_2}^{(2)}(t') \delta_{k_2}^\gamma, \\ \mathbb{P}(X_1 = i_1, X_2 = i_2, t = t' | X_1 = X_2 = \gamma, t = 0) \\ &= \sum_{k_1, k_2, k_3} m_{i_1 k_1}^{(1)}(t') m_{i_2 k_2}^{(2)}(t') \delta_{k_3}^\gamma \Gamma_{k_3}^{k_1 k_2}.\end{aligned}$$

Implementing (3.11) leads to the requirement that

$$\Gamma_\gamma^{k_1 k_2} = \delta_{k_1}^\gamma \delta_{k_2}^\gamma,$$

and the basis dependent definition of the branching operator

$$\delta e_i = e_i \otimes e_i.$$

From this construction we can express the phylogenetic tree Figure 3.1 as

$$P = (1 \otimes 1 \otimes M_3 \otimes M_4) 1 \otimes 1 \otimes \delta(M_1 \otimes M_2 \otimes M_5) 1 \otimes \delta \cdot \delta \pi,$$

which can also be written in the more convenient form

$$P = (M_1 \otimes M_2 \otimes M_3 \otimes M_4) 1 \otimes 1 \otimes \delta(1 \otimes 1 \otimes M_5) 1 \otimes \delta \cdot \delta \pi.$$

This form can be generalized so that any phylogenetic tree can be expressed in the form

$$P := M_1 \otimes M_2 \otimes \dots \otimes M_m \tilde{P}, \quad (3.12)$$

with  $M_a \in GL(n)$ ,  $1 \leq a \leq m$ , and  $\tilde{P}$  is found by taking  $P$  and setting the Markov operators on the leaf edges,  $M_1, M_2, \dots, M_m$ , all equal to the identity operator. This representation will be of importance to us as we consider invariant theory in terms of phylogenetics.

### 3.5 Entanglement and phylogenetics

In this final section we will study the properties of a phylogenetic tensor evaluated on invariant functions of the general linear group. Recalling (3.7), we see that in all reasonable cases the determinant of the transition matrices of a phylogenetic tree is non-zero. This implies that the transition matrices are elements of  $GL(n)$ . Thus in the case of a phylogenetic tensor of the form (3.12), an invariant will take the form

$$f(P) = \prod_{a=1}^m \det(M_a)^k f(\tilde{P}).$$

Presently we study the case where  $|\mathcal{K}| = 2$  and the phylogenetic tensor occurs in the tensor product space relevant to two qubits and three qubits respectively.

#### 3.5.1 Two qubits

For the case of two qubits the most general phylogenetic tensor is given by

$$P = (M_1 \otimes M_2)\delta\pi, \quad (3.13)$$

which corresponds to the tree of Figure 3.2. Following (3.12) we have

$$\tilde{P} = \delta\pi.$$

As will be discussed in detail in Chapter 4, the concurrence can be used to establish the magnitude of divergence between a pair of sequences derived from a single branching event. The concurrence of the phylogenetic state (3.13) is given by

$$\mathcal{C}(P) = \det[M_1] \det[M_2] \mathcal{C}(\delta\pi).$$

Explicitly we have

$$\begin{aligned} \mathcal{C}(\tilde{P}) &= \sum \delta_{i_1 i_2} \pi_{i_2} \delta_{i_3 i_4} \pi_{i_4} \epsilon_{i_1 i_3} \epsilon_{i_2 i_4} \\ &= \pi_1 \pi_2, \end{aligned}$$

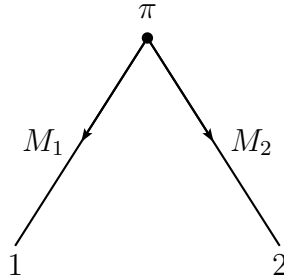


Figure 3.2: Phylogenetic tree with two leaves

and find that

$$\mathcal{C}(P) = \det[M_1] \det[M_2] \pi_1 \pi_2.$$

Assuming that the determinants of the Markov operators are non-zero we see that the phylogenetic tensor is on the entangled orbit.

In comparison, if there is no stochastic dependence between the random variables the phylogenetic state can be expressed as

$$P = p_1 \otimes p_2,$$

which is a product state, such that the random variables  $X_1$  and  $X_2$  are stochastically independent, and the concurrence vanishes. Thus the non-vanishing of the concurrence can be used as a test of stochastic dependence between any two molecular sequences. In Chapter 4 we will show that the determinants of the Markov operators tend to zero as  $t$  tends to infinity and we conclude that the phylogenetic (3.13) state tends to a product state after an infinite amount of divergence. This is what one would expect as the case of infinite divergence should correspond exactly to the case of stochastic independence.

### 3.5.2 Three qubits

In this section we study the phylogenetic state

$$P = (M_1 \otimes M_2 \otimes M_3) 1 \otimes \delta(1 \otimes M_4) \delta\pi, \quad (3.14)$$

which corresponds to the tree Figure 3.3. Again following (3.12) we have

$$\tilde{P} = 1 \otimes \delta(1 \otimes M_4) \delta\pi.$$

We now determine which orbit the phylogenetic state (3.14) lies in. By the general properties of the tangle we find that

$$\mathcal{T}(P) = \prod_{i=1}^3 (\det M_i)^2 \mathcal{T}(\tilde{P}),$$

and by explicit computation

$$\mathcal{T}(\tilde{P}) = (\det M_4)^2 (\pi_1 \pi_2)^2,$$

to conclude that

$$\mathcal{T}(P) = (\det M_1 \det M_2 \det M_3 \det M_4)^2 (\pi_1 \pi_2)^2.$$

From this we can conclude that the phylogenetic state (3.14) lies on the  $GHZ$  orbit and the evaluation of the tangle upon three aligned sequence can be used as a test of triplet stochastic dependence.



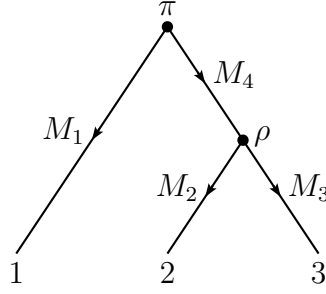


Figure 3.3: Phylogenetic tree with three leaves

### 3.5.3 Phylogenetic relation

Referring to (3.7), we see that for continuous time Markov chains the determinants of the transition matrices satisfy:

$$\begin{aligned} 0 < \det M(s, t) &\leq 1, \quad \forall 0 \leq t < \infty, \\ \lim_{t \rightarrow \infty} \det M(t) &= 0. \end{aligned} \tag{3.15}$$

Above we have seen that for phylogenetic data of three aligned sequences derived from a tree the tangle polynomial is non-zero, and for two aligned sequences derived from a tree the concurrence is also non-zero. But taking (3.15) into account we see that, if any one of the branches of a phylogenetic tree is extended to infinite length this will induce the vanishing of these invariant functions which implies that the corresponding part of the phylogenetic tensor decouples from the overall state to form a partial product state. Thus the case of no stochastic dependence directly corresponds to entanglement of the tensor state and stochastic dependence can be tested for using invariant functions. Introducing independent time parameters for each external branch, we can express the phylogenetic tree (3.3) as

$$P(t_1, t_2, t_3) := [M_1(0, t_1) \otimes M_2(0, t_2) \otimes M_3(0, t_3)] 1 \otimes \delta[1 \otimes M_4] \delta \pi.$$

Now, as we have seen, the tangle polynomial will satisfy

$$\begin{aligned} \lim_{t_a \rightarrow \infty} \mathcal{T}(P(t_1, t_2, t_3)) &= 0, \\ \forall a &= 1, 2, 3. \end{aligned}$$

For the concurrence we have

$$\lim_{t_a \rightarrow \infty} \mathcal{C}_b(P(t_1, t_2, t_3)) = 0, \tag{3.16}$$

if and only if  $a = b$ . From these observations we can conclude that we have the limit:

$$\lim_{t_1 \rightarrow \infty} p_{i_1 i_2 i_3}(t_1, t_2, t_3) = p_{i_1}^{(1)} p_{i_2 i_3}^{(23)}(t_2, t_3),$$

and similar for  $t_2, t_3$ . The phylogenetic state decouples into a partial product state after an infinite amount of stochastic divergence. This is what one would expect, as the branch lengths of the tree become so large that it is impossible to observe the branching event which relates to leaves.

From these observations we define a *phylogenetic relation* to exist whenever the relevant phylogenetic tensor cannot be written as a product state.

## 3.6 Closing remarks

In this chapter we have established the mathematical connection between the notion of entanglement and that of phylogenetic relation. We showed that simple group invariant functions used to quantify entanglement can be utilized in the phylogenetic case. We focused on the invariant function known as the tangle, but considered only the case of two character states. In the next chapter we will study the properties of the tangle in the case of three and four character states.

## CHAPTER 4

# Using the tangle

The distance based approach to phylogenetic reconstruction using the neighbor joining algorithm is a commonly used technique [23, 37, 49, 52]. Under the assumptions of a Markov model of sequence evolution, the phylogenetic relationship is uniquely reconstructible from (suitably defined) pairwise distances [54]. The approach relies crucially upon the calculation of distance matrices from aligned sequence data which give a measure of the pairwise evolutionary distance between the extant taxa under consideration. As far as tree building algorithms are concerned it is required that the distances are strictly *linearly* related to the sum of the (theoretical) edge lengths of the phylogenetic tree, and that the parameters of the linear relation do not vary across the tree. It is essential to the analysis that the measure of distance chosen has both biological and statistical as well as mathematical significance. If one assumes the standard Markov model, the edge lengths of a phylogenetic tree can be taken mathematically to be a quantity that we refer to as the *stochastic distance*. (For mathematical discussion of this quantity see Goodman [24] who refers to the stochastic distance as *intrinsic time*, and see also Barry and Hartigan [5] who gave a biological interpretation.) Under the assumptions of a general Markov model the log det formula is commonly used to obtain pairwise distances. Further, if one may assume a stationary process then the log det formula can be modified to give an estimate of the actual stochastic distance [40]. (That is, the constants of the linear relation are set by the stationarity assumption.)

Distance based methods and, consequently, the log det formula are often used in favour of other methods (such as maximum likelihood) in cases where there has been significant compositional heterogeneity during the evolutionary history. The theoretical basis which motivates this usage was presented by Steel [56] and is discussed in Lockhart, Steel, Hendy and Penny [40] and Gu and Li [25]. More recently, Jermini, Ho, Ababneh, Robinson and Larkum published a simulation study which confirms that the log det outperforms other techniques in this case [33]. Lockhart *et al.* showed that by using the assumption that the base composition remains close to constant, the log det formula can be modified to give an estimate of the actual stochastic distance. However, as will be shown, in both its original and modified form the log det formula includes

an approximation crucially dependent upon the compositional heterogeneity remaining minimal. The effectiveness of the log det formula to correctly reconstruct the phylogenetic history when there has been significant compositional heterogeneity is thus brought into question. Hence there is a contradictory state of affairs between the theoretical basis of the log det and the circumstances under which it is implemented. In this chapter we will generalize the log det formula in such a way that this dependence upon base composition is truly absent.

A disadvantage of the log det formula is that it uses only *pairwise* sequence data and is blind to the fact that extra information regarding pairwise distances can be obtained from the sequence data of additional taxa. Felsenstein [21] mentions that it is surprising that distance techniques work at all given that they ignore the extra information in higher order alignments. This chapter details exactly how the log det formula can be improved upon by taking functions of aligned sequence data for *three* taxa at a time. It may seem counter-intuitive that consideration of a third taxon can impart information regarding the evolutionary distance between two taxa, but it is the case that by considering a third taxon the log det formula can be refined. This result depends crucially upon the fact that, as is somewhat trivially the case for two taxa, there is only one possible (unrooted) tree topology relating three taxa. (For discussion of what a tree topology is see [44], Chapter 5.) It is possible to refine the log det formula by considering the respective distance to an arbitrary third taxon. The reader should note that the use of triplet sequence data to the problem of reconstruction of the Markov model was also considered in [12] and [48]. The approach discussed in the present chapter is original in the sense that triplets of the aligned sequences are being used explicitly in a distance method, and follows on from the theoretical discussions of [59].

A complication arises regarding the total stochastic distance between leaves and the placement of the root of a phylogenetic tree. It turns out that if we define phylogenetic trees of identical topology to be equivalent if they give the identical probability distributions then we find that the total stochastic distance between leaves is not, in general, left unchanged as we move the root of the tree. The so defined equivalence class provides a generalization of Felsenstein's *pulley principle* [19] and was first presented in Steel, Szekely and Hendy [57]. The fact that the stochastic distance is not left unchanged is a surprising result and has important implications regarding the interpretation of the edge lengths of phylogenetic trees defined under the Markov model. In particular this result implies that the log det technique is an inconsistent estimator of pairwise distances on phylogenetic trees. It is the purpose of this chapter to present a new estimator that is consistent in the case of phylogenetic quartets. We are motivated to present this construction of quartet distance matrices by the interest in phylogenetic reconstruction of large trees from the correct determination of the set of  $\binom{n}{4}$  quartets [9, 58].

This chapter will begin by formally defining the stochastic distance. We will then examine how the general linear group invariants, the *det* (2.13) and the *tangle* (2.17), can be used to estimate the stochastic distance between any two taxa on a phylogenetic tree. As a consequence of this discussion we will examine a generalized pulley principle and finish by showing that by including

the tangle in the analysis we can arrive at a consistent estimator.

**Note:** This chapter follows closely the text of [60].

### 4.0.1 Stochastic distance

In this chapter we will be interested in the assignment of edge lengths to phylogenetic trees. To this end we consider the rate of change of base changes at time  $s$ :<sup>1</sup>

$$\lambda(s) := \sum_{i \in \mathcal{K}} \frac{\partial \mathbb{P}(X(t) = i | X(s) \neq i)}{\partial t} \Big|_{t=s}.$$

By considering (3.5) and (3.8) this quantity can be explicitly expressed using the rate parameters:

$$\begin{aligned} \lambda(s) &= - \sum_i q_{ii}(s), \\ &= -\text{tr} Q(s). \end{aligned}$$

From these considerations we define the *stochastic distance* to be given by the expression

$$\omega(s, t) := \int_s^t \lambda(u) du.$$

By considering the time-ordered product representation (3.6) and the Jacobi identity  $\det e^X = e^{\text{tr} X}$ , we find that the stochastic distance can be directly related to the transition probabilities of the Markov chain:

$$\omega(s, t) = -\log \det M(s, t). \quad (4.1)$$

Our assignment of edge lengths will take the Markov matrix associated with each edge and set the edge length equal to the stochastic distance.

The relation (4.1) is known in various guises in both the mathematical and phylogenetic literature [5, 24] and, as will be confirmed in the next section, is the basis of the log det formula. It should also be noted that (4.1) will remain positive and finite because  $\omega(s, s) = 0$ ,  $\lambda(s) \geq 0$  and the integral  $\int_0^T \lambda(t) dt$  is not expected to diverge.<sup>2</sup>

---

<sup>1</sup>It is standard to include a factor of  $n^{-1}$  in this definition. However, this factor clutters the consequent formulae and here we do not include it as it has no consequence to the forgoing discussion and can always be incorporated into the analysis later.

<sup>2</sup>There are two cases where the integral may diverge, but we can safely exclude these possibilities as follows. i.  $\lambda(t)$  may be a badly behaved function. We can reject this possibility outright in phylogenetics as there is every reason to expect the rate parameters to change smoothly with time. ii.  $T \rightarrow \infty$ . We can safely ignore this possibility as we will be assuming that the divergence times of the Markov chain are sufficiently small such that the phylogenetic historical signal is still obtainable.

### 4.0.2 Observability of the stochastic distance

An interesting consideration (which at first sight is at odds with our aims) is that given a single random process modelled as a CTMC there is simply no way of inferring the value of the stochastic distance from an observed distribution without making restrictive assumptions about the process and the initial distribution. This is best illustrated by considering a stationary CTMC for which the rate-parameters are time-independent and given an initial distribution  $\pi_i(0)$  satisfy

$$\sum_j q_{ij} \pi_j(0) = 0, \quad \forall i.$$

Now, although the consequent distribution is time-independent,  $\pi_i(t) = \pi_i(0)$ , and hence carries zero informative value in comparison to the initial distribution, the stochastic distance itself increases linearly with time

$$\omega(0, t) = - \left( \sum_i q_{ii} \right) t.$$

From this observation it is clear that in the general case if all we have access to is the final distribution, there is no way we can estimate the stochastic distance unless we make some additional assumptions about the stochastic process. The remarkable fact is that in the case of phylogenetics it is possible to estimate the stochastic distance from the observed distribution. (As we will show in Section 4.1, this is true even for the case where the underlying chains are stationary!)

## 4.1 Pairwise distance measures

In this section we will derive and discuss a standard approach to the construction of distance matrices. (For an excellent perspective of the various measures of phylogenetic pairwise distance see [3].) A distance matrix,  $\phi = [\phi_{ab}]_{(a,b) \in L}$ , is constructed from the aligned sequence data of multiple extant taxa such that each entry gives a suitable estimate of the distance between a given pair of taxa. The mathematical conditions on the  $\phi_{ab}$  are the standard conditions of a distance function as well as the four point condition [54] (which is required for the distance measure to be consistent with the tree structure):

$$\begin{aligned} \phi_{ab} &\geq 0, \\ \phi_{ab} &= 0 \text{ iff } a = b, \\ \phi_{ab} &= \phi_{ba}, \\ \phi_{ab} + \phi_{cd} &\leq \max\{\phi_{ac} + \phi_{bd}, \phi_{ad} + \phi_{bc}\}; \quad \forall a, b, c, d \in L. \end{aligned} \tag{4.2}$$

There are no further conditions required upon  $\phi$  for it to give a unique tree reconstruction [54]. However it is of course desirable for the distance measure

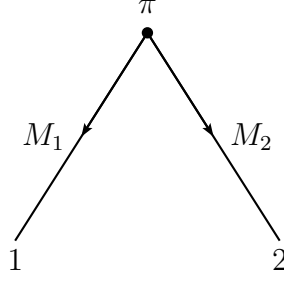


Figure 4.1: Phylogenetic tree of two taxa

to have a well defined biological interpretation. To this end, for a given edge  $e$ , we define the *edge length*,  $\omega_e$ , which we set to be the stochastic distance (4.1) taken from the Markov model:

$$\omega_e = -\log \det M_e.$$

It is then apparent that any significant estimate of pairwise distance must statistically be expected to converge to a value which is linearly related to the sum of the stochastic distances lying on the (unique) path between the two taxa under consideration. It should be clear that such a measure will satisfy the relations (4.2). It is crucial to the performance of the distance measure under a tree building algorithm that the parameters of the linear relation are expected to be *constant* for all pairs of taxa. That is, given the unique path between leaf  $a$  and  $b$ ,  $P(T; a, b)$ , we are demanding that statistically we have the following convergence:

$$\phi_{ab} \rightarrow \alpha \omega(a, b) + \beta,$$

where

$$\omega(a, b) := \sum_{e \in P(T; a, b)} \omega_e,$$

and  $\alpha$  and  $\beta$  are expected to be independent of  $a$  and  $b$ . As we will see, the  $\log \det$  formula does not satisfy this property for the most general models.

#### 4.1.1 The log det formula

In Figure 4.1 we consider the two taxa phylogenetic tree, with pattern probabilities given by

$$p_{i_1 i_2} = \sum_j m_{i_1 j}^{(1)} m_{i_2 j}^{(2)} \pi_j. \quad (4.3)$$

By considering the matrices defined as

$$\begin{aligned} P^{(1,2)} &:= [p_{ij}]_{(i,j) \in \mathcal{K}}, \\ D_\pi &:= [\text{diag}(\pi_i)]_{i \in \mathcal{K}}; \end{aligned}$$

it is easy to show that (4.3) is equivalent to

$$P^{(1,2)} = M_1 D_\pi M_2^t.$$

Taking the determinant of this expression and considering (4.1) yields

$$\begin{aligned} \det P^{(1,2)} &= \det M_1 \det M_2 \det D_\pi \\ &= e^{-(\omega_1 + \omega_2)} \prod_i \pi_i. \end{aligned} \quad (4.4)$$

This expression can be generalized to the case of any two taxa from a given phylogenetic tree:

$$\det P^{(a,b)} = e^{-\omega(a,b)} \prod_i \pi_i^{(a,b)}, \quad (4.5)$$

where  $\pi_i^{(a,b)}$  is the distribution at the most recent ancestral vertex between taxa  $a$  and  $b$  determined by the meeting point of the two paths traced backwards along the phylogenetic tree from leaf  $a$  and  $b$ .

Now  $\omega(a,b)$  is theoretically equal to the total stochastic distance between each of  $a$  and  $b$  and their most recent ancestral vertex and hence it is clear that  $-\log \det P^{(a,b)}$  will be linearly related to this quantity. In the original formulation of the log det, a distance measure between two taxa was defined as

$$\begin{aligned} d_{ab} &:= -\log \det P^{(a,b)} \\ &= \omega(a,b) - \sum_i \log[\pi_i^{(a,b)}], \end{aligned} \quad (4.6)$$

and shown to satisfy the conditions (4.2) [56]. From this relation it seems that one can take  $\alpha = 1$  and  $\beta = -\sum_i \log[\pi_i^{(a,b)}]$  and evaluate (4.6) on the observed pattern frequencies for each pair of taxa to calculate a well defined distance matrix from a set of aligned sequence data (as was presented in [40]). This procedure depends crucially upon the shifting term  $\beta = \sum_i \log[\pi_i^{(a,b)}]$  being independent of  $a$  and  $b$ . However, this is only true in special circumstances such as star phylogeny or if the base composition is constant (the stationary model). In the general case, one is led to a different shifting term depending on the topology of the tree (this was noted in Sumner and Jarvis [59] and we reproduce the result here). Consider the phylogenetic tree of three taxa given in Figure 4.2 with pattern probabilities given by

$$p_{i_1 i_2 i_3} = \sum_{j,k} m_{i_1 j}^{(1)} m_{i_2 k}^{(2)} m_{i_3 k}^{(3)} m_{k j}^{(4)} \pi_j.$$

By calculating (4.6) for the three possible pairs of taxa we find that

$$\begin{aligned} d_{12} &= (\omega_1 + \omega_4 + \omega_2) - \sum_i \log \pi_i, \\ d_{13} &= (\omega_1 + \omega_4 + \omega_3) - \sum_i \log \pi_i, \\ d_{23} &= (\omega_1 + \omega_3) - \sum_i \log \rho_i, \end{aligned}$$



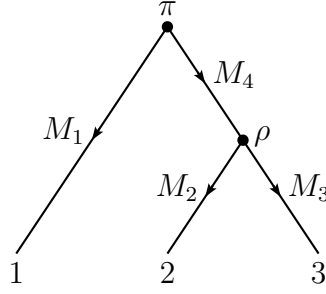


Figure 4.2: Phylogenetic tree of three taxa

from which it is explicitly clear that the shifting term is *not* constant across this phylogenetic tree. The shifting term is dependent on the base composition at the most recent ancestral node of the two taxa and from the above example it is clear that this depends on the topology of the tree and is not always simply the root of the tree. This means that (4.6) does not produce distance matrices whose entries are linearly related to the edge length of the tree because the entries of the matrix will depend essentially upon the topology of the tree.

It is, however, possible to obtain an estimate of the total stochastic distance between any two taxa by modifying the log det formula. The ancestral base composition is approximated by using the harmonic mean

$$\prod_i \pi_i^{(a,b)} \approx \left[ \prod_{i_1, i_2} \pi_{i_1}^{(a)} \pi_{i_2}^{(b)} \right]^{\frac{1}{2}}, \quad (4.7)$$

where  $\pi_k^{(a,b)}$  is the closest common ancestral base composition between taxa  $a$  and  $b$  and  $\pi_i^{(a)} := \mathbb{P}(X_a(\tau_a) = i)$  (and similarly for  $b$ ). One is then led to the formula

$$d'_{ab} := -\log \det P^{(a,b)} + \frac{1}{2} \sum_{i_1, i_2} (\log \pi_{i_1}^{(a)} + \log \pi_{i_2}^{(b)}), \quad \forall a, b \in L. \quad (4.8)$$

where  $d'_{ab}$  is then an estimator of the total stochastic distance between taxa  $a$  and  $b$ . (This form of the log det formula was presented in [40] and [54]).

In the case of a stationary base composition model the additional assumption is made that

$$\sum_j m_{ij}^{(e)} \pi_j = \pi_i; \quad \forall e \in E.$$

In this case we have

$$\pi_i^{(a,b)} = \pi_i^{(a)} = \pi_i^{(b)}, \quad \forall a, b \in L,$$

and it is clear that the harmonic mean approximation becomes an exact relation and the log det formula is expected to converge exactly to the total stochastic distance between the two taxa.

### 4.1.2 The tangle

In this section we will show how the log det formula can be generalized to obtain, for the most general Markov models, an unbiased estimate of the

distance matrix. The basis of the technique is the existence of a measure analogous to (4.4) which is valid for *triplets*.

Sumner and Jarvis [59] presented a polynomial function  $\mathcal{T}$  which is known in quantum physics as the tangle and can be evaluated on phylogenetic data sets of three aligned sequences in the case of  $n = 2$ . Evaluated on the pattern probabilities of any phylogenetic tree of three taxa,  $\{a, b, c\}$ , the tangle takes on the theoretical value

$$\mathcal{T}(a, b, c) = e^{-2\omega(a, b, c)} \left( \prod_{i \in \mathcal{K}} \pi_i \right)^2, \quad (4.9)$$

where

$$\omega(a, b, c) := \sum_{e \in T} \omega_e,$$

$\pi$  is the common ancestral root of the three taxa and this relation holds independently of the particular tree topology which relates  $\{a, b, c\}$ . This independence upon the topology is a very nice property and is crucial to the practical use of the tangle as a distance measure. The similarity between (4.9) and (4.5) should be noted.

In this chapter we report generalized tangles, which are polynomials which satisfy (4.9) for the cases of  $n = 3, 4$  in addition to the  $n = 2$  case which was presented in [59]. It is possible to infer the existence of the tangles and derive their polynomial form from group theoretical considerations. Here we give forms using the completely antisymmetric (Levi-Civita) tensor,  $\epsilon$ , which has components  $\epsilon_{i_1 i_2 \dots i_n}$  and satisfies  $\epsilon_{12 \dots n} = 1$ . For the cases of  $n = 2, 3, 4$  the tangles are given by<sup>3</sup>

$$\mathcal{T}_2 = \frac{1}{2!} \sum_1^2 p_{i_1 i_2 i_3} p_{j_1 j_2 j_3} p_{k_1 k_2 k_3} p_{l_1 l_2 l_3} \epsilon_{i_1 j_1} \epsilon_{i_2 j_2} \epsilon_{k_1 l_1} \epsilon_{k_2 l_2} \epsilon_{i_3 l_3} \epsilon_{j_3 k_3},$$

$$\mathcal{T}_3 = \frac{1}{3!} \sum_1^3 p_{i_1 i_2 i_3} p_{j_1 j_2 j_3} p_{k_1 k_2 k_3} p_{l_1 l_2 l_3} p_{m_1 m_2 m_3} p_{n_1 n_2 n_3} \\ \cdot \epsilon_{i_1 j_1 k_1} \epsilon_{j_2 k_2 l_2} \epsilon_{k_3 l_3 m_3} \epsilon_{l_1 m_1 n_1} \epsilon_{m_2 n_2 i_2} \epsilon_{n_3 i_3 j_3},$$

$$\mathcal{T}_4 = \frac{1}{4!} \sum_1^4 p_{i_1 j_1 k_1} p_{i_2 j_2 k_2} p_{i_3 j_3 k_3} p_{i_4 j_4 k_4} p_{i_5 j_5 k_5} p_{i_6 j_6 k_6} p_{i_7 j_7 k_7} p_{i_8 j_8 k_8} \\ \cdot \epsilon_{i_1 i_2 i_3 i_4} \epsilon_{i_5 i_6 i_7 i_8} \epsilon_{j_1 j_5 j_4 j_8} \epsilon_{j_2 j_6 j_3 j_7} \epsilon_{k_1 k_5 k_2 k_6} \epsilon_{k_3 k_7 k_4 k_8};$$

respectively, (where the summation is over every index). The expression (4.9) can be proved by studying the group theoretical properties of the tangle (see [59]) and by explicitly expanding the above forms. For the tangle on two characters we find

$$\begin{aligned} \mathcal{T}_2 = & -p_{122}^2 p_{211}^2 + 2p_{121} p_{122} p_{211} p_{212} - p_{121}^2 p_{212}^2 + 2p_{112} p_{122} p_{211} p_{221} + \\ & 2p_{112} p_{121} p_{212} p_{221} - 4p_{111} p_{122} p_{212} p_{221} - p_{112}^2 p_{221}^2 - 4p_{112} p_{121} p_{211} p_{222} + \\ & 2p_{111} p_{122} p_{211} p_{222} + 2p_{111} p_{121} p_{212} p_{222} + 2p_{111} p_{112} p_{221} p_{222} - p_{111}^2 p_{222}^2. \end{aligned}$$

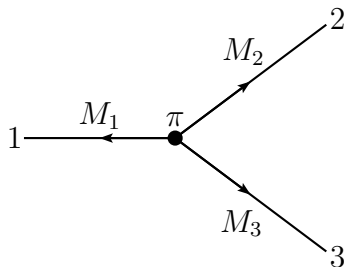
Substantial computer power is required to explicitly compute  $\mathcal{T}_3$  and  $\mathcal{T}_4$ . These polynomials have 1152 and 431424 terms, respectively.

---

<sup>3</sup>This expression for  $\mathcal{T}_2$  corrects for the erroneous expression presented in [59].

### 4.1.3 Star topology

Consider the phylogenetic tree relating three taxa with a star topology:



with pattern probabilities given by the formula

$$p_{i_1 i_2 i_3} = \sum_j m_{i_1 j}^{(1)} m_{i_2 j}^{(2)} m_{i_3 j}^{(3)} \pi_j.$$

Here we will use the fact that the root of this tree is also the common ancestral root of any pair of the three taxa. (This is not the case in general if we allow for a general rooting of the tree and/or more than three taxa. The complications arising in these cases will be dealt with in the next section.)

Considering the formulae (4.9) and (4.4) we are led to introduce the novel distance matrix,  $\Delta$ , with the pairwise distance between  $\{a, b\}$  given by

$$\Delta_{ab}^{(c)} := -\log \mathcal{T}(a, b, c) + \log \det P^{(a,c)} + \log \det P^{(b,c)}, \quad a, b, c \in L. \quad (4.10)$$

From (4.4) and (4.9) it follows that

$$\Delta_{ab}^{(c)} = \omega(a, b),$$

such that our new formula will directly give the stochastic distance between the two taxa. There is no need to make the harmonic mean approximation and this distance measure is mathematically and biologically meaningful. This is the main result of this chapter: given a set of aligned sequence data, the tangle formula (4.10) can be used to compute the *exact* pairwise edge lengths for any triplet. As mentioned above, the explicit polynomial form of the tangle has been computed for the cases of two, three and four bases and it is our intent that (4.10) will provide a significant improvement over the log det formula in the calculation of pairwise distance matrices for these cases.

### 4.1.4 Summary

Considering the stochastic distance to be the correct way to assign edge lengths to branches of a phylogenetic tree, we have reviewed three different ways of obtaining a distance measure between any two taxa  $a$  and  $b$ :

1.  $d_{ab} = -\log \det P^{(a,b)}$
2.  $d'_{ab} = -\log \det P^{(a,b)} + \frac{1}{2} \sum_{i_1, i_2} (\log \pi_{i_1}^{(a)} + \log \pi_{i_2}^{(b)})$
3.  $\Delta_{ab}^{(c)} = -\log \mathcal{T}(a, b, c) + \log \det P^{(a,c)} + \log \det P^{(b,c)}$

where one substitutes the observed pattern frequencies into these expressions. From the previous considerations we found that these three distance measures have the following properties:

1. When  $d_{ab}$  is evaluated on a set of observed pattern frequencies, this estimator satisfies the requirements of a distance function (4.2), but is inconsistent with the general Markov model as the estimate is *not* expected to converge to a value that is linearly related to  $\omega(a, b)$ .
2. When  $d'_{ab}$  is evaluated on a set of observed pattern frequencies, this estimator satisfies the requirements (4.2) and is expected to converge to a value that is linearly related to  $\omega(a, b)$  whenever the compositional heterogeneity is absent. In the heterogeneous case this quantity approximates  $\omega(a, b)$  by using (4.7).
3. When  $\Delta_{ab}^{(c)}$  is evaluated on a set of observed pattern frequencies, this estimator satisfies the requirements of (4.2) and is expected to converge exactly to  $\omega(a, b)$  in all cases.

Thus we see that the tangle formula (4.10) should be a significant improvement as an empirical estimator of  $\omega(a, b)$  upon both forms of the log det formula. However, the formula (4.10) depends on taking an arbitrary third taxon,  $c$ . The question remains as to what to do in the case of constructing pairwise distances for sets of greater than three taxa. The surprising answer to this question will be addressed in the next section where we will bring into question the uniqueness of the theoretical quantity  $\omega(a, b)$ . The discussion has consequences for the interpretation of each of the estimators of pairwise distances that we have discussed.

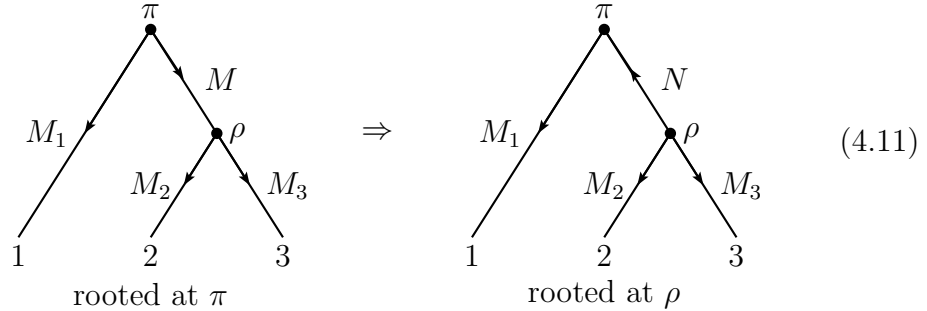
## 4.2 Generalized pulley principle

In this section we generalize the Felsenstein's pulley principle [19]. In its original formulation the pulley principle describes the unrootedness of phylogenetic trees where the underlying Markov model is assumed to be reversible and stationary. Here we show how the pulley principle may be generalized to remain valid under the most general Markov models. Our immediate motivation is to show that (4.10) remains a valid distance measure under the circumstance of a general phylogenetic tree of multiple taxa. Unfortunately this generalization introduces surprising mathematical complications which have consequences

not only for our formula (4.10), but also for the log det technique and any other estimate of the stochastic distance upon a phylogenetic tree. The discussion will lead to the consequence that, for a given tree topology, there are multiple – actually, infinitely many – phylogenetic trees with identical probability distributions. (These phylogenetic trees differ by arbitrary rerootings and consequential redirection of edges.) We will see that the generalized pulley principle shows that as far as inference from the observed pattern frequencies is concerned, there is no theoretical justification behind specifying the root of a phylogenetic tree if the most general Markov model is allowed. Also, we will see that the theoretical value of the stochastic distance is not constant for arbitrary rerootings of a phylogenetic tree. Clearly, if the stochastic distance is not uniquely defined theoretically, then one must be careful in interpreting any formula that gives an estimate thereof from the observed data.

Considering a phylogenetic tree as a directed graph shows that a rerooting involves redirecting an edge (or part thereof). The property required is that the Markov chain on the involved edge is taken to progress as if time has been reversed, and we refer to the new chain as the *time-reversed* chain. This should be compared to the requirement of *reversibility* as defined in the mathematical literature, (for example see [29]). In the case of a stationary and reversible Markov chain the time-reversed chain (as we will define) is identical to the original chain.

By way of example, we take the rooted tree of three taxa (4.7) and redirect the relevant internal edge to give the following rerooting:



Our immediate task is to infer the existence of an appropriate time-reversed Markov chain,  $N$ , such that these two phylogenetic trees give identical probability distributions. If we equate the pattern probabilities of (4.11) and contract all edges except the one we are reversing, we are led to the simple algebraic solution

$$n_{ij} = \frac{m_{ji}\pi_i}{\rho_j}. \quad (4.12)$$

(This solution was presented in [57].) Presently we use this result to give an explicit form in the general case.

Given a CTMC  $X(t)$  with transition probabilities

$$m_{ij}(t, s) := \mathbb{P}(X(t) = i | X(s) = j),$$

we wish to find a second CTMC,  $Y(t)$ , such that, given any  $T \geq 0$ , we have

$$\mathbb{P}(Y(t) = i) = \pi_i(T - t), \quad \forall \quad 0 \leq t \leq T.$$

That is, if the direction of time is reversed, the second CTMC  $Y(t)$  has identical distribution to  $X(t)$ . The uniqueness of  $Y(t)$  is a technical matter which we do not consider, because in the phylogenetic case there are extra restrictions which led to the unique solution (4.12).

Considering again the general case, we write

$$\mathbb{P}(Y(t) = i | Y(s) = j) := n_{ij}(t, s)$$

and use (4.12) to infer the general solution

$$n_{ij}(t, s) = \frac{m_{ji}(T - s, T - t) \pi_i(T - t)}{\pi_j(T - s)}. \quad (4.13)$$

It is trivial to show that these transition probabilities satisfy the requirements of a CTMC:

$$\sum_j n_{ji}(t, s) = 1, \quad \forall i,$$

$$N(t, s)N(s, u) = N(t, u),$$

where  $N(t, s) = [n_{ij}(t, s)]_{(i,j \in \mathcal{K})}$ .

Furthermore, by using (3.8) we find that the rate parameters of the time-reversed chain can be expressed as

$$\begin{aligned} f_{ij}(s) &:= \left. \frac{\partial n_{ij}(t, s)}{\partial t} \right|_{t=s} \\ &= \frac{q_{ji}(T - s) \pi_i(T - s)}{\pi_j(T - s)} - \sum_k \frac{\delta_{ij} q_{ik}(T - s) \pi_k(T - s)}{\pi_j(T - s)} \end{aligned}$$

From which it follows that

$$f_{ij}(s) \geq 0, \quad \forall i \neq j; \quad f_{ii}(s) = - \sum_{j \neq i} f_{ji}(s)$$

which confirms that the  $f_{ij}(s)$  are a valid set of rate parameters for a CTMC (as expected). It should be noted that even in the case where  $X(t)$  is a homogeneous chain it is certainly not the case in general that  $Y(t)$  is also homogeneous. Consider, however, the stationary and reversible case, with the respective conditions:

$$\sum_j q_{ij} \pi_j(0) = 0,$$

$$q_{ij} \pi_j(0) = q_{ji} \pi_i(0),$$

where the stationarity condition ensures that

$$\pi_i(t) = \pi_i(0), \quad \forall t.$$

In this circumstance it follows that

$$f_{ij} = q_{ij},$$

such that  $Y(t) \equiv X(t)$  and is hence also stationary and reversible. This was the basis of Felsenstein's initial formulation of the pulley principle – if one considers only stationary and reversible Markov chains on a phylogenetic tree, any time-reversed chain is identical to the original Markov chain and hence a phylogenetic tree can be arbitrarily rerooted. We have given a continuous time generalization of Felsenstein's result which removes the stationary and reversible restriction.

Equipped with the solution (4.13) it is possible to take any phylogenetic tree and find an alternative tree of identical topology, but rooted in a different place, such that the alternative tree generates an identical probability distribution to that of the original. This is the basis of our generalized pulley principle.

The reader should note that we have proven, under the assumptions of the most general Markov model, that it is not possible to determine the orientation of a phylogenetic tree by only considering the joint probability distribution it generates at the leaves. Thus, any procedure that attempts to determine the root from the observed pattern frequencies must be justified by making additional assumptions about the underlying stochastic process. Chang [12] showed that the tree topology and (up to permutations of rows) the set of transition matrices, are reconstructible from the set of *triples* of the joint distribution at the leaves. This is consistent with our result as Chang explicitly prohibited internal nodes with two incident edges and worked with unrooted/unorientated trees. Baake [2] showed that (up to similarity transformation) the *return-trip* matrices (in our notation  $M(s, t)N(t, s)$ ) are identifiable from the set of *pair-wise* joint distributions at the leaves. Again this is consistent with our result. The curious aspect of the generalized pulley principle is that the stochastic distance is *not* conserved along the edge of the tree where the directedness was reversed. This is easy to show by considering the determinant of (4.13)

$$\det N(t, s) = \det M(T - s, T - t) \prod_i \frac{\pi_i(T - t)}{\pi_i(T - s)} \quad (4.14)$$

Thus the stochastic distance in the reversed time chain is equal to that of the original chain if and only if

$$\prod_i \frac{\pi_i(T - t)}{\pi_i(T - s)} = 1. \quad (4.15)$$

This property of CTMCs and their time-reversed counterparts was observed by Barry and Hartigan [5]. It can be seen that in the stationary case (4.15) will certainly be true. There are other cases where (4.15) may hold but there does

not seem to any biologically sound way to interpret the required condition. In the proceeding discussion we will consider the consequences of the generalized pulley principle upon the interpretation of distance matrices. We see that for a given observed distribution we can use the generalized pulley principle to show that there are multiple edge length assignments using the stochastic distance which are consistent with the Markov model on a phylogenetic tree. These edge length assignments differ from one another as a consequence of (4.14).

### 4.2.1 Interpretation

For illustrative purposes we consider the consequence to the stochastic distance of the rerooting of a phylogenetic tree of two taxa. We consider the phylogenetic trees illustrated in Figure 4.3, and by using the generalized pulley principle define their respective transition matrices so that their probability distributions are identical:

$$n_{ij} = \frac{m_{ji}\pi_i}{\rho_j},$$

$$\rho_i = \sum_j m_{ij}\pi_j.$$

We find in the first case that we have

$$\omega_\pi(1, 2) = -\log \det M_1 - \log \det M - \log \det M_2,$$

and in the second case

$$\omega_\rho(1, 2) = -\log \det M_1 - \log \det N - \log \det M_2.$$

Now in general  $\det M \neq \det N$  and we see that the two possible pairwise distances are not expected to be equal. However, from an empirical perspective it is impossible to distinguish these two possible theoretical scenarios because the probability distributions are identical. Now, because any estimator of the pairwise distance must be inferred from the observed distribution, we conclude that one must be careful to consider exactly what theoretical quantity one is obtaining an estimate of. For the case of the log det formula we find that the quantity it is estimating depends essentially upon the base composition of the observed sequences as follows:

Considering the pairwise distance  $d'_{ab}$  given by (4.8), from the generalized pulley principle we see that this formula will give an estimate of the stochastic distance between  $a$  and  $b$ , where the common ancestral node is placed such that the quantity

$$\chi(a, b) := \prod_i \pi_i^{(a,b)} - \left[ \prod_{i_1, i_2} \pi_{i_1}^{(a)} \pi_{i_2}^{(b)} \right]^{\frac{1}{2}}$$

is minimized. Thus the log det method will be inconsistent in the sense that, if there has been compositional heterogeneity, the pairwise distance it produces



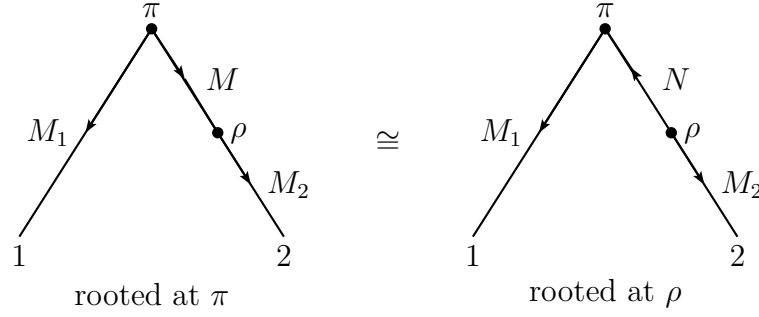


Figure 4.3: Using the generalized pulley principle

will be an estimate for the edge length assignment where  $\chi(a, b)$  is minimized. This may have nothing to do with true placement of the common ancestral vertex and it may even be the case that  $\chi(a, b)$  has multiple minimum points. The situation amounts to the fact that, for a given phylogenetic tree, one is (potentially) using the log det to estimate pairwise distances with a different edge length assignment for each and every pair of taxa. Clearly for the analysis of multiple taxa this could become a significant problem and any alternative approach which removes this inconsistency would be beneficial to the analysis. We see that the consequences of the generalized pulley principle and (4.14) to the interpretation of the Markov model of phylogenetics are quite subtle. The generalized pulley principle is telling us that there is no direct way to distinguish the rootedness (and equivalently the directedness of internal edges) of phylogenetic trees. This is due to the fact that there are (infinitely) many phylogenetic trees of identical topology which generate identical probability distributions, differing only by the assignment of stochastic distance and the associated redirection of internal edges.

### 4.3 The quartet case

In this section we will show that in the case of a phylogenetic tree of four taxa, the tangle can be used to construct consistent quartet distance matrices. These distance matrices will be consistent in the sense that theoretically they are constructed from one topology with *one* edge length assignment. This should be compared to the log det formula which in the general case can be estimating a different edge length assignment for each and every pairwise distance.

For analytic purposes we use the generalized pulley principle to root the four taxon tree in two ways, as illustrated in Figure 4.4. The difference between the two cases is simply in the directedness of the internal edge and the generalized pulley principle allows us to calculate the required transition probabilities so that the two trees generate identical probability distributions. The pattern probabilities for the two cases are given by

$$\begin{aligned}
 p_{i_1 i_2 i_3 i_4} &= \sum_{j,k} m_{i_1 j}^{(1)} m_{i_2 j}^{(2)} m_{i_3 k}^{(3)} m_{i_4 k}^{(4)} m_{k j}^{(5)} \pi_j \\
 &= \sum_{j,k} m_{i_1 k}^{(1)} m_{i_2 k}^{(2)} m_{i_3 j}^{(3)} m_{i_4 j}^{(4)} n_{k j}^{(5)} \rho_j
 \end{aligned} \tag{4.16}$$

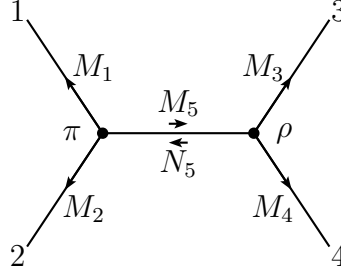


Figure 4.4: Four taxa tree with alternative roots

where to ensure the equality of the two expressions we have

$$n_{ij}^{(5)} = \frac{m_{ji}^{(5)} \pi_i}{\rho_j},$$

and  $\rho_i = \sum_j m_{ij}^{(5)} \pi_j$ .

From these expressions we wish to calculate the theoretical values of the formula (4.10) for each possible group of three taxa. To obtain these values one simply chooses the form of the tree such that after the deletion of a fourth taxon one is left with a three taxon tree of star topology. By sequentially deleting one taxon at a time we are led to the four star topology subtrees illustrated in Figure 4.5 and the corresponding pattern probabilities are given by the expressions

$$\begin{aligned} p_{ijk}^{(123)} &= \sum_{l_1, l_2} m_{il_1}^{(1)} m_{jl_1}^{(2)} m_{kl_2}^{(3)} m_{l_2 l_1}^{(5)} \pi_{l_1}, \\ p_{ijk}^{(124)} &= \sum_{l_1, l_2} m_{il_1}^{(1)} m_{jl_1}^{(2)} m_{kl_2}^{(4)} m_{l_2 l_1}^{(5)} \pi_{l_1}, \\ p_{ijk}^{(134)} &= \sum_{l_1, l_2} m_{il_2}^{(1)} n_{l_2 l_1}^{(5)} m_{jl_1}^{(2)} m_{kl_1}^{(4)} \rho_{l_1}, \\ p_{ijk}^{(234)} &= \sum_{l_1, l_2} m_{il_2}^{(2)} n_{l_2 l_1}^{(5)} m_{jl_1}^{(3)} m_{kl_1}^{(4)} \rho_{l_1}. \end{aligned}$$

From this it is easy to calculate the values simply by considering the results of the previous section:

$$\begin{aligned} \Delta_{12}^{(3)} &= \omega(1, 2), & \Delta_{12}^{(4)} &= \omega(1, 2), \\ \Delta_{13}^{(2)} &= \omega_\pi(1, 3), & \Delta_{13}^{(4)} &= \omega_\rho(1, 3), \\ \Delta_{14}^{(2)} &= \omega_\pi(1, 4), & \Delta_{14}^{(3)} &= \omega_\rho(1, 4), \\ \Delta_{23}^{(1)} &= \omega_\pi(2, 3), & \Delta_{23}^{(4)} &= \omega_\rho(2, 3), \\ \Delta_{24}^{(1)} &= \omega_\pi(2, 4), & \Delta_{24}^{(3)} &= \omega_\rho(2, 4), \\ \Delta_{34}^{(1)} &= \omega(3, 4), & \Delta_{34}^{(2)} &= \omega(3, 4), \end{aligned} \tag{4.17}$$

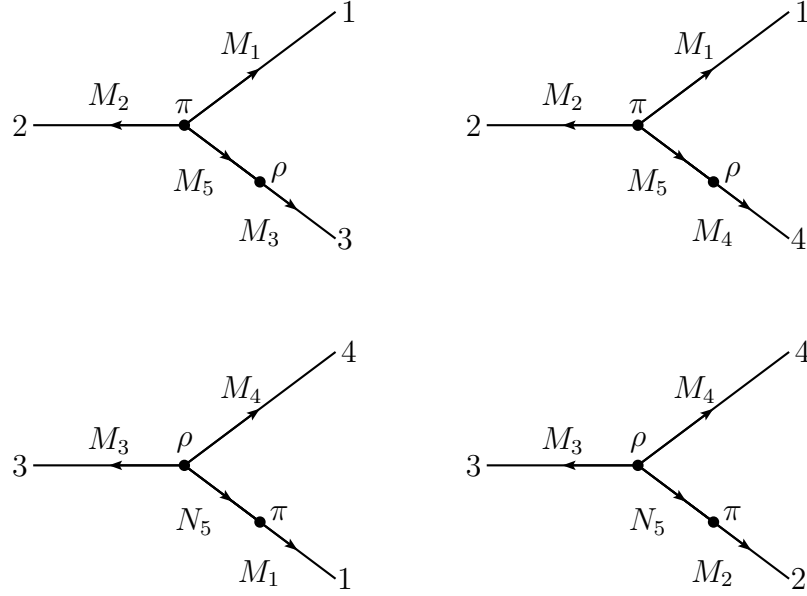


Figure 4.5: Three taxon subtrees

where

$$\omega(a, b) = \omega_a + \omega_b,$$

$$\omega_\pi(a, b) = \omega_a + \omega_m + \omega_b,$$

$$\omega_\rho(a, b) = \omega_a + \omega_n + \omega_b,$$

$$\omega_m = -\log \det M,$$

$$\omega_n = -\log \det N;$$

and we have made use of (4.14) in the form

$$\omega_n = \omega_m - \sum_i (\log \pi_i - \log \rho_i).$$

We see that for any two taxa we have two options for assigning a pairwise distance. In the cases of the pairs (12) and (34) we see that either choice is consistent with the other, whereas in the case of the pair (13), (14), (24) and (34) the two choices lead to an inconsistent assignment of the internal edge length upon the tree. Effectively what is happening here is that for a four taxa tree there are two possible edge length assignments for the internal edge and for a given pair of taxa ( $ab$ ) and third taxa  $c$ , the tangle formula (4.10) is estimating the distance between  $a$  and  $b$  by assigning one of the two possible edge lengths to the internal edge depending on the topology of the tree.

It is possible to eliminate this inconsistency by using either a *max* or *min* criterion in the construction of the distance matrix:

$$\phi_{ab}^{max} := \max\{\Delta_{ab}^{(c)}, \Delta_{ab}^{(c')}\}$$

or

$$\phi_{ab}^{min} := \min\{\Delta_{ab}^{(c)}, \Delta_{ab}^{(c')}\}.$$

By making one of these choices to construct a distance matrix we choose the directedness of the internal edge of the phylogenetic tree (4.3) consistently. This procedure leads to an improvement of consistency upon the log det technique for the construction of quartet phylogenetic distance matrices. It is hoped that this technique can be used fruitfully to improve the reconstruction of phylogenetic quartets, which can be used as a first step in the reconstruction of large phylogenetic trees [9, 58].

## 4.4 Closing remarks

In this chapter we have given a review of the standard assignment of branch weights to phylogenetic trees, reviewed the use of the log det formula as an estimator of pairwise distances and shown how a previously unknown polynomial, the tangle, can be used to construct an improved estimator. We have generalized Felsenstein's *pulley principle* and used this result to show exactly how the distance matrix estimates become inconsistent when applied to the reconstruction problem of multiple taxa. We have shown that the tangle formula along with a *max/min* criterion can be used to remove this inconsistency and construct consistent quartet distance matrices.

## CHAPTER 5

# Markov invariants

In this chapter we will refine the use of invariant theory on phylogenetic trees by defining *Markov invariants* to be invariant functions specific to the general Markov model of sequence evolution. To achieve this we return to the representation theory introduced in Chapter 2 and show how the Schur functions can be used to give a count of the existence of the Markov invariants. A procedure which constructs the explicit polynomial form of these invariants will be developed and we examine, as prompted from Chapter 3, the structure of these invariants once placed on a phylogenetic tree. For the triplet and quartet case we show that there exist Markov invariants which have the additional property of being *phylogenetic invariants* [1, 15, 55]. These previously unobserved invariants can be used to achieve quartet reconstruction under the assumptions of the general Markov model.

### 5.1 The Markov semigroup

In Chapter 3 we considered the transition matrices of a continuous time Markov chain as a subset of the general linear group, and used this property to study the structure of invariant polynomials (used as measures of entanglement in quantum physics) when evaluated on a phylogenetic tree. In this section we will close the gap between the general linear group and the subset consisting of the transition matrices of a CTMC by formally defining the *Markov semigroup*. (For a detailed discussion of the Lie group properties of the Markov semigroup and its relation to the Affine group see [34].)

Recalling the vector  $\theta = \sum e_i$  (3.9), the Markov semigroup on  $n$  elements,  $\mathcal{M}(n)$ , with parameters  $s \leq t < \infty$  is defined relative to  $\theta$  as the subset of  $GL(n)$  which satisfies:

1.  $M(s, s) = 1$ ,
2.  $M(t', t)M(t, s) = M(t', s) \quad \forall s < t < t'$ ,
3.  $(\theta, M(t, s)v) = (\theta, v) \quad \forall v \in V$ .

In general this set does not form a group. Consider the time evolution of a probability vector  $p(t)$ , defined by

$$p(t) = M(t, s)p(s), \quad s \leq t.$$

This time evolution will conserve the total probability

$$\sum p_i(t) = (\theta, p(t)) = (\theta, M(t, s)p(s)) = (\theta, p(s)) = \sum p_i(s) = 1.$$

Defining

$$Q(s) := \frac{\partial M(t, s)}{\partial t} \Big|_{t=s},$$

it follows that in the  $\{e_i\}$  basis, the matrix elements of  $Q(t) = [q_{ij}(t)]$  satisfy

$$q_{ij}(t) \geq 0, \quad \forall i \neq j; \quad q_{ii}(t) = - \sum_{j \neq i} q_{ji}(t),$$

and hence each  $M(s, t)$  is a valid transition matrix for a CTMC.

In Chapter 3 we saw that the Markov model of phylogenetics can be considered in terms of the action  $\times^m GL(n)$  on  $V^{\otimes m}$  (3.12). We refine this to the action of  $\times^m \mathcal{M}(n)$  on  $V^{\otimes m}$  so that any phylogenetic tensor can be written as

$$P = M_1 \otimes M_2 \otimes \dots \otimes M_m \tilde{P},$$

with  $M_a \in \mathcal{M}(n)$ ,  $1 \leq a \leq m$ . Our present task will be to define and derive invariant functions,  $w : V^{\otimes m} \rightarrow \mathbb{C}$ , which satisfy

$$w(P) = \prod_{a=1}^m \det(M_a)^k w(\tilde{P}),$$

for all  $M_a \in \mathcal{M}(n)$ ,  $1 \leq a \leq m$ , and analyse their relevance to the problem of phylogenetic tree reconstruction. (It should be noted that an invariant of the general linear group is certainly an invariant of the Markov semigroup, but the converse is not necessarily true.)

### 5.1.1 Invariant functions of the Markov semigroup

Before considering the more general case of the action  $\times^m \mathcal{M}(n)$  on  $V^{\otimes m}$  given by

$$\psi \rightarrow M_1 \otimes M_2 \otimes \dots \otimes M_m \psi,$$

we will first define invariant functions of the action  $\mathcal{M}(n)$  on  $V^{\otimes m}$  given by

$$\psi \rightarrow \otimes^m M \psi.$$

Given that  $\mathcal{M}(n)$  does not form a group we have to be careful in our definitions of representations and invariant functions. To this end we define the set of functions  $\mathbb{C}[V^{\otimes m}]_d^{\mathcal{M}(n)}$  as the subset  $w \in \mathbb{C}[V^{\otimes m}]_d$  which satisfy

$$\begin{aligned} w \circ \otimes^m M &= \det(M)^k w, \quad \forall M \in \mathcal{M}(n), \\ md &= kn, \end{aligned} \tag{5.1}$$

(where we have carefully not invoked the inverse element  $M^{-1}$ ). Presently we will derive a sufficient condition for the existence of such invariant functions. Consider  $w \in \mathbb{C}[V]_d$  satisfying (5.1). Under the canonical isomorphism  $\omega : \mathbb{C}[V^{\otimes m}]_d \rightarrow (V^{\otimes m})^{\{d\}}$  (2.5) we have

$$w = \omega(\chi),$$

for some  $\chi \in (V^{\otimes m})^{\{d\}}$ . Carefully taking note of the relations (2.6) and (2.7) it follows that

$$\omega(\chi) \circ M^{\otimes m} = \omega(\otimes^{md} M^t \chi).$$

Hence  $w := \omega(\chi)$  will satisfy (5.1) if and only if

$$\otimes^{md} M^t \chi = \det(M)^k \chi, \quad \forall M \in \mathcal{M}(n). \tag{5.2}$$

Consider the tensor  $\phi \in V^{\{k^n\}} \otimes V^{\otimes s}$  expressed as

$$\phi = \eta \otimes (\theta^{\otimes s}),$$

with  $\eta \in V^{\{k^n\}}$ . Recalling (2.2.6) and the definition of the Markov semigroup it follows that  $\phi$  satisfies (5.2):

$$\otimes^{kn+s} M^t \phi = \det(M)^k \phi, \quad \forall M \in \mathcal{M}(n).$$

Consider the decomposition of  $V^{\{k^n\}} \otimes V^{\otimes s}$  into irreducible representation spaces of  $GL(n)$ :

$$V^{\{k^n\}} \otimes V^{\otimes s} = \sum_{|\lambda|=kn+s} h_\lambda V^\lambda,$$

for some unknown multiplicities  $h_\lambda$ . Our present task is to identify the irreducible representation space in which the tensor  $\phi$  is contained. Assume  $\phi \in V^\mu$  with  $|\mu| = kn + s$  and recall that

$$V^\mu = Y_\mu V^{\otimes kn+s},$$

where  $Y_\mu$  is the projection operator satisfying

$$\begin{aligned} Y_\mu^2 &= Y_\mu, \\ Y_\mu Y_{\mu'} &= 0, \quad |\mu'| = |\mu|, \end{aligned}$$

so that  $Y_\mu$  is the unique Young operator satisfying

$$Y_\mu \phi = \phi.$$

Considering the inherent permutation symmetry of  $\phi$ , it is clear that

$$\mu = \{k + s, k^{n-1}\}.$$

From this we conclude that  $\phi \in V^{\{k+s, k^{n-1}\}}$ , and there exists  $\chi \in (V^{\otimes m})^{\{d\}}$  satisfying (5.2) whenever

$$(V^{\otimes m})^{\{d\}} \supset V^{\{k+s, k^{n-1}\}},$$

as an irreducible subspace under  $GL(n)$ .

**Proposition 5.1.1.** *A sufficient condition for the existence of a Markov invariant  $w \in \mathbb{C}[V^{\otimes m}]_d^{\mathcal{M}(n)}$  is that  $(\times^m \{1\})^{\otimes \{d\}} \supset \{k + s, k^{n-1}\}$  for some  $md = nk + s$ .*

In direct analogy to the development of Theorem (2.3.3) we generalize this to the action of  $\times^m \mathcal{M}(n)$  on  $V^{\otimes m}$ :

**Proposition 5.1.2.** *A sufficient condition for the existence of a Markov invariant  $w \in \mathbb{C}[V^{\otimes m}]_d^{\times^m \mathcal{M}(n)}$  is that  $*^m \{k + s, k^{n-1}\} \supset \{d\}$  for some  $d = nk + s$ .*

Using the representation theoretical tools we have developed it does not seem trivial to show that these conditions are also necessary. However we now have at our disposal a tool for inferring the existence of Markov invariants in various cases.

In the next section we will return to the construction of invariants for the general linear group in order to derive a technique allowing us to compute these Markov invariants.

## 5.2 Alternative computation of invariants of the general linear group

The construction of invariants of the general linear group was presented in Chapter 2 using the properties of the Levi-Civita tensor. Unfortunately this construction does not generalize to the case of the Markov semigroup. In this section we show how Young tableaux can be used to construct the invariant functions of  $GL(n)$  directly. In the next section we show how this technique can be generalized to allow for the construction of the Markov invariants.



### 5.2.1 Action of $GL(n)$ on $V^{\otimes m}$

Recall that the number of invariants of weight  $k$  in  $\mathbb{C}[V^{\otimes m}]_d^{GL(n)}$  is equal to the number of occurrences of the partition  $\{k^n\}$  in  $(\times^m \{1\})^{\otimes \{d\}}$  with  $kn = md$ . This gives us a technique for the proof of existence of invariant polynomials, but leaves us with the problem of their explicit construction. Recall Theorem 2.3.2 and we see that our task is to identify the one-dimensional representations of the general linear group in the decomposition of  $(V^{\otimes m})^{\{d\}}$ . Suppose we consider  $U = V^{\otimes m}$  as a  $(nm)$ -dimensional vector space with basis  $u_1, u_2, \dots, u_{nm}$ . As we saw in Chapter 2, if  $U$  has a basis  $u_1, u_2, \dots, u_{nm}$  then any  $\chi \in U^{\{d\}}$  can be constructed from an arbitrary  $\chi \in U^{\otimes d}$  by taking

$$\varphi = Y_{\{d\}}\chi,$$

where the Young operator acts on the  $\{u_{\alpha_1} \otimes u_{\alpha_2} \otimes \dots \otimes u_{\alpha_d}\}$  basis of  $U^{\otimes d}$ ,  $1 \leq \alpha_1, \dots, \alpha_d \leq nm$ . Now we define

$$\phi = Y_{\{k^n\}}\varphi,$$

where the Young operator now acts on the  $\{e_{i_1} \otimes e_{i_2} \otimes \dots \otimes e_{i_{dm}}\}$  basis of  $(V^{\otimes m})^{\{d\}} \cong U^{\{d\}}$ ,  $1 \leq i_1, \dots, i_{dm} \leq n$ . The final step is to construct the single independent component of  $\phi$  using the semi-standard tableau:

1	1	...	1
2	2		2
$\vdots$			
n	n		n

and then map over the invariant ring using  $\omega : (V^{\otimes m})^{\{d\}} \rightarrow \mathbb{C}[V^{\otimes m}]_d$ . The invariant is then

$$f := \omega(\phi) = \omega(Y_{\{k^n\}}\varphi) = \omega(Y_{\{k^n\}}Y_{\{d\}}\chi),$$

which will satisfy

$$f \circ g = \det(g)^k f,$$

for all  $g \in GL(n)$ .

There is no problem with choosing the operator  $Y_{\{d\}}$  as there is only one possible standard tableau:

1	2	...	d
---	---	-----	---

However there does not seem to be any *a priori* way of deciding which standard tableau to use for the symmetrization  $Y_{\{k^n\}}$ . In general there are more standard tableaux than one-dimensional representations. This is not a serious issue since the Young symmetrization procedure needs to be implemented in

an algebraic computation computer package. Our procedure was to make judicious choices of standard tableaux and check for algebraic independence of the resulting invariants until the correct count was achieved. In what follows we will present the results of these computations.

The above outlines the formal procedure. In practice we implement the algorithm as follows. The above is equivalent to computing

$$\Psi_{i_1 \dots i_{md}} := Y_{\{k^n\}} \psi_{i_1 \dots i_m} \psi_{i_{m+1} \dots i_{2m}} \dots \psi_{i_{(m-1)d+1} \dots i_{md}}, \quad (5.3)$$

where

$$(ab) \psi_{i_1 \dots i_m} \dots \psi_{i_a \dots i_b} \dots \psi_{i_{(m-1)d+1} \dots i_{md}} := \psi_{i_1 \dots i_m} \dots \psi_{i_a \dots i_b} \dots \psi_{i_{(m-1)d+1} \dots i_{md}},$$

for any  $1 \leq a, b \leq md$ , defines the meaning of (5.3) and there is no need to symmetrize with  $Y_{\{d\}}$ . (In practice the symmetries inherent in this procedure give us some clue as to how to choose the appropriate standard tableaux for  $\{k^n\}$ .) We then set the indices of  $\Psi_{i_1 \dots i_{md}}$  using the single semi-standard tableaux to get

$$w(\psi) = \Psi_{12 \dots n 12 \dots n \dots 12 \dots n}. \quad (5.4)$$

Now this expression only depends on the choice of standard tableau for  $\{k^n\}$ . In practice we compute (5.4) for different standard tableaux until we have the correct number of independent invariants.

### 5.2.2 Examples

We consider the case  $m = 2$ . We have  $(\{1\} \times \{1\})^{\otimes \{1\}} = \{2\} + \{1^2\} \ni \{1^2\}$ , and hence there is one invariant of degree  $d = 1$ . Of course this invariant can simply be found by symmetrizing  $V^{\otimes 2}$  with the only standard tableau of shape  $\{1^2\}$ :

$$\begin{array}{|c|} \hline 1 \\ \hline 2 \\ \hline \end{array}$$

with corresponding Young operator

$$Y_{\{1^2\}} = (e - (12)).$$

The symmetrized tensor is

$$\Psi_{i_1 i_2} = Y_{\{1^2\}} \psi_{i_1 i_2} = \psi_{i_1 i_2} - \psi_{i_2 i_1}.$$

The invariant is found by inserting index labels from the relevant semi-standard tableau, so that

$$w(\psi) = \Psi_{12} = \psi_{12} - \psi_{21}.$$

For  $d = 2$  the output of **Schur** shows that  $(\{1\} \times \{1\})^{\otimes \{2\}} \ni 2\{2^2\}$ .

There are two Young operators with shape  $\{2^2\}$ :

$$\begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & 4 \\ \hline \end{array} \quad \begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & 4 \\ \hline \end{array} \quad .$$

The invariants are then given by

$$\Psi_{i_1 i_2 i_3 i_4} = Y_{\{2^2\}} \psi_{i_1 i_2} \psi_{i_3 i_4}.$$

For the first tableau we have

$$Y_{\{2^2\}} = (e - (13) - (24) + (13)(24))(e + (12) + (34) + (12)(34)),$$

and find explicitly for the semi-standard tableau corresponding to component  $\Psi_{1212}$ :

$$h_1(\psi) = \psi_{12}^2 + 2\psi_{12}\psi_{21} + \psi_{21}^2 - 4\psi_{11}\psi_{22},$$

and for the second tableau

$$h_2(\psi) = \psi_{12}^2 - \psi_{12}\psi_{21} + \psi_{21}^2 - \psi_{11}\psi_{22}.$$

It is a simple exercise to show that these invariants are linear combinations of the two invariants produced in Chapter 2 (2.10):

$$\begin{aligned} h_1 &= f_1^2 - 4f_2, \\ h_2 &= f_1^2 - f_2. \end{aligned}$$

For the case of  $GL(3)$  on  $V^{\otimes 2}$  **Schur** shows that  $(\{1\} \times \{1\})^{\otimes \{3\}} \supset 2\{2^3\}$ . The invariants are constructed from arbitrary  $\psi \in (V^{\otimes 2})^{\otimes 3}$  as

$$f = \omega(Y_{\{2^3\}} \circ Y_{\{3\}} \psi),$$

with the standard tableaux

$$\begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & 4 \\ \hline 5 & 6 \\ \hline \end{array} \quad \begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 & 4 \\ \hline 5 & 6 \\ \hline \end{array}$$

generating two independent elements:

$$\begin{aligned} h_1(\psi) &= -\psi_{13}^2 \psi_{22} + \psi_{12} \psi_{13} \psi_{23} + \psi_{13} \psi_{21} \psi_{23} - \psi_{11} \psi_{23}^2 - 2\psi_{13} \psi_{22} \psi_{31} \\ &\quad + \psi_{12} \psi_{23} \psi_{31} + \psi_{21} \psi_{23} \psi_{31} - \psi_{22} \psi_{31}^2 + \psi_{12} \psi_{13} \psi_{32} \\ &\quad + \psi_{13} \psi_{21} \psi_{32} - 2\psi_{11} \psi_{23} \psi_{32} + \psi_{12} \psi_{31} \psi_{32} + \psi_{21} \psi_{31} \psi_{32} - \psi_{11} \psi_{32}^2 \\ &\quad - \psi_{12}^2 \psi_{33} - 2\psi_{12} \psi_{21} \psi_{33} - \psi_{21}^2 \psi_{33} + 4\psi_{11} \psi_{22} \psi_{33}, \end{aligned}$$

and

$$\begin{aligned} h_2(\psi) &= \psi_{13}^2 \psi_{22} - \psi_{12} \psi_{13} \psi_{23} - \psi_{13} \psi_{21} \psi_{23} + \psi_{11} \psi_{23}^2 + \psi_{12} \psi_{23} \psi_{31} \\ &\quad - \psi_{21} \psi_{23} \psi_{31} + \psi_{22} \psi_{31}^2 - \psi_{12} \psi_{13} \psi_{32} + \psi_{13} \psi_{21} \psi_{32} - \psi_{12} \psi_{31} \psi_{32} \\ &\quad - \psi_{21} \psi_{31} \psi_{32} + \psi_{11} \psi_{32}^2 + \psi_{12}^2 \psi_{33} + \psi_{21}^2 \psi_{33} - 2\psi_{11} \psi_{22} \psi_{33}. \end{aligned}$$

Again it is possible to show that these invariants are linear combinations of the corresponding invariants produced in Chapter 2 (2.11).

### 5.2.3 Action of $\times^m GL(n)$ on $V^{\otimes m}$

Recalling Theorem 2.3.3, we note that the number of weight  $k$  invariants in  $\mathbb{C}[V^{\otimes m}]_d^{\times^m GL(n)}$  is equal to the number of occurrences of  $\{d\}$  in the decomposition of  $*^m\{k^n\}$ . For even  $m$  we have the identity

$$*^m\{1^n\} = \{n\},$$

and for odd  $m$

$$*^m\{1^n\} = \{1^n\}.$$

Thus we see that for even  $m$  there is a single invariant function of degree  $d=n$  and for odd  $m$  there are none. For even  $m$  the invariant is generated from

$$\Psi_{i_1 \dots i_{nm}} = Y_{\{1^n\}}^{(1)} Y_{\{1^n\}}^{(2)} \dots Y_{\{1^n\}}^{(m)} \psi_{i_1 \dots i_m} \psi_{i_{m+1} \dots i_{2m}} \dots \psi_{i_{(n-1)m} \dots i_{nm}}$$

where each standard tableau  $Y^{(a)}$ ,  $1 \leq a \leq m$ , is

$a$
$m + a$
$(n-1)m + a$

We then set the indices of  $\Psi_{i_1 \dots i_{nm}}$  using the single semi-standard tableau for each Young operator to obtain

$$w(\psi) = \Psi_{11 \dots 122 \dots 2 \dots nm \dots n}.$$

It should be clear that this procedure is completely equivalent to the invariants obtained using the Levi-Civita tensor Chapter 2 (2.14). In the case  $n = 2$ , this procedure generates the determinant invariants (2.13) and for  $n = 4$  the quangles (2.14).

However, as we will now see, we need to use the tableaux technique in order to do the same job for the Markov semigroup.

## 5.3 Computation of the Markov invariants

Here we will generalize the above technique for computing invariants of the general linear group to the case of the Markov semigroup. It should be noted that in the case of the general linear group, the basis in which the calculations are performed is of no consequence as the invariants take on the identical form (up to scaling) in *any* basis. (This is by definition!) However, in the case of the Markov invariants all calculations with Young operators must be performed in the basis  $\{z_0, z_a\}$ , see Chapter 2 (2.1). This is due to the very definition of the Markov semigroup which depends on a particular choice of the vector  $\theta = \sqrt{n}z_0$ . Thus, in the subsequent discussion, it should be remembered that all Markov invariants are presented in the form they take in the  $\{z_0, z_a\}$  basis.

### 5.3.1 Markov invariants of $\mathcal{M}(n)$ on $V^{\otimes m}$

In this section we consider the action of  $\mathcal{M}(n)$  on  $V^{\otimes m}$  given by

$$\psi \rightarrow \otimes^m \psi.$$

Recalling Conjecture 5.1.1, it follows that if

$$(\times^m \{1\})^{\otimes \{d\}} \ni \{k + s, k^{n-1}\}$$

for some  $md = nk + s$  there exists a Markov invariant  $w \in \mathbb{C}[V^{\otimes m}]_d^{\mathcal{M}(n)}$ . (In all that follows it should be noted that the case  $s=0$  reproduces an invariant of the general linear group.) Computing

$$\Psi_{i_1 \dots i_{dm}} := Y_{\{k+s, k^{n-1}\}} \psi_{i_1 \dots i_m} \psi_{i_{m+1} \dots i_{2m}} \dots \psi_{i_{(d-1)m+1} \dots i_{dm}}$$

where the standard tableau of shape  $\{k + s, k^{n-1}\}$  used to define  $Y_{\{k+s, k^{n-1}\}}$  is not fixed, but is chosen judiciously. The final step is to compute  $w(\psi)$  by inserting indices into  $\Psi$  using the semi-standard tableau:

0	0	...	0	0	0		0
1	1		1				
⋮							
n-1	n-1		n-1				

### 5.3.2 Examples

We will consider Markov invariants of degree  $d = 1$  only. For the case of  $n=2, m=3$ , **Schur** shows that

$$(\times^3 \{1\})^{\otimes \{1\}} = \times^3 \{1\} \ni 2\{21\},$$

which implies that there are two Markov invariants corresponding to  $\{21\}$  with  $k=s=1$ .

There are two standard tableaux of shape  $\{2, 1\}$ :

1	2		
3			

1	3		
2			

The corresponding  $d=1$  Markov invariant follows from computing

$$\Psi_{i_1 i_2 i_3} := Y_{\{21\}} \psi_{i_1 i_2 i_3}$$

and then inserting indices according to the single semi-standard tableau:

$$\begin{array}{|c|c|} \hline 0 & 0 \\ \hline 1 & \\ \hline \end{array}.$$

For the first tableau we compute the symmetrized tensor

$$\Psi_{a_1 a_2 a_3}^{(1)} = \psi_{a_1 a_2 a_3} + \psi_{a_2 a_1 a_3} - \psi_{a_3 a_2 a_1} - \psi_{a_2 a_3 a_1}.$$

The single independent component gives the Markov invariant

$$\Psi_{001}^{(1)} = 2\psi_{001} - \psi_{100} - \psi_{010}.$$

The second tableau gives the symmetrized tensor

$$\Psi_{a_1 a_2 a_3}^{(2)} = \psi_{a_1 a_2 a_3} + \psi_{a_3 a_2 a_1} - \psi_{a_2 a_1 a_3} - \psi_{a_3 a_1 a_2}.$$

The single independent component gives the second Markov invariant

$$\Psi_{010}^{(2)} = 2\psi_{010} - \psi_{100} - \psi_{001}.$$

As a second example, consider the case  $n=2, m=4$ , with **Schur** giving

$$(\times^4\{1\})^{\otimes\{1\}} = \times^4\{1\} \ni 3\{31\},$$

so there are three Markov invariants with  $k=s=1$ . There are three standard tableaux and hence three candidate Young operators:

$$\begin{array}{|c|c|c|} \hline 1 & 2 & 3 \\ \hline 4 & & \\ \hline \end{array} \quad \begin{array}{|c|c|c|} \hline 1 & 2 & 4 \\ \hline 3 & & \\ \hline \end{array} \quad \begin{array}{|c|c|c|} \hline 1 & 3 & 4 \\ \hline 2 & & \\ \hline \end{array}.$$

The associated semi-standard tableau is

$$\begin{array}{|c|c|c|} \hline 0 & 0 & 0 \\ \hline 1 & & \\ \hline \end{array}.$$

For the first tableau we have the symmetrized tensor:

$$\begin{aligned} \Psi_{a_1 a_2 a_3 a_4} &= \psi_{a_1 a_2 a_3 a_4} + \psi_{a_2 a_1 a_3 a_4} + \psi_{a_3 a_2 a_1 a_4} + \psi_{a_1 a_3 a_2 a_4} \\ &+ \psi_{a_3 a_1 a_2 a_4} + \psi_{a_2 a_3 a_1 a_4} - \psi_{a_4 a_2 a_3 a_1} - \psi_{a_2 a_4 a_3 a_1} \\ &- \psi_{a_3 a_2 a_4 a_1} - \psi_{a_4 a_3 a_2 a_1} - \psi_{a_3 a_4 a_2 a_1} - \psi_{a_2 a_3 a_4 a_1}. \end{aligned}$$

By inserting the indices we get the Markov invariant

$$6\psi_{0001} - 2\psi_{1000} - 2\psi_{0100} - 2\psi_{0010}.$$

And by analogy for the remaining two Young operators (with the same semi-standard tableau) we have the Markov invariants

$$6\psi_{0010} - 2\psi_{1000} - 2\psi_{0100} - 2\psi_{0001},$$

and

$$6\psi_{0100} - 2\psi_{1000} - 2\psi_{0010} - 2\psi_{0001}.$$

Our final example is the case  $n=3, m=4$ , with **Schur** giving

$$(\times^4\{1\})^{\otimes 1} = \times^4\{1\} \ni 3\{21^2\},$$

so there are two Markov invariants with  $k = s = 1$ . Again, there are three standard tableaux

$$\begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & \\ \hline 4 & \\ \hline \end{array} \quad \begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 & \\ \hline 4 & \\ \hline \end{array} \quad \begin{array}{|c|c|} \hline 1 & 4 \\ \hline 2 & \\ \hline 3 & \\ \hline \end{array}$$

with associated semi-standard tableau

$$\begin{array}{|c|c|} \hline 0 & 0 \\ \hline 1 & \\ \hline 2 & \\ \hline \end{array}.$$

From the first standard tableau we compute the symmetrized tensor:

$$\begin{aligned} \Psi_{a_1 a_2 a_3 a_4}^{(1)} = & \psi_{a_1 a_2 a_3 a_4} + \psi_{a_2 a_1 a_3 a_4} - \psi_{a_3 a_2 a_1 a_4} - \psi_{a_2 a_3 a_1 a_4} - \psi_{a_4 a_2 a_3 a_1} \\ & - \psi_{a_2 a_4 a_3 a_1} - \psi_{a_1 a_2 a_4 a_3} - \psi_{a_2 a_1 a_4 a_3} + \psi_{a_4 a_2 a_1 a_3} \\ & + \psi_{a_2 a_4 a_1 a_3} + \psi_{a_3 a_2 a_4 a_1} + \psi_{a_2 a_3 a_4 a_1}. \end{aligned}$$

Again by filling the indices according to the semi-standard tableau we get the Markov invariant

$$\begin{aligned} \Psi_{0012}^{(1)} = & 2\psi_{0012} - \psi_{1002} - \psi_{0102} - \psi_{2010} - \psi_{0210} - 2\psi_{0021} \\ & + \psi_{2001} + \psi_{0201} + \psi_{1020} + \psi_{0120}. \end{aligned}$$

Similarly we find for the remaining two standard tableaux:

$$\begin{aligned} \Psi_{0102}^{(2)} = & -\psi_{0012} + \psi_{0021} + 2\psi_{0102} - \psi_{0120} - 2\psi_{0201} \\ & + \psi_{0210} - \psi_{1002} + \psi_{1200} + \psi_{2001} - \psi_{2100} \end{aligned}$$

and

$$\begin{aligned} \Psi_{0120}^{(3)} = & \psi_{0012} - \psi_{0021} - \psi_{0102} + 2\psi_{0120} + \psi_{0201} - 2\psi_{0210} \\ & - \psi_{1020} + \psi_{1200} + \psi_{2010} - \psi_{2100}. \end{aligned}$$

These three invariants are linearly independent, as required.

## 5.4 Markov invariants of $\times^m \mathcal{M}(n)$ on $V^{\otimes m}$

We now consider invariants of the group action  $\times^m \mathcal{M}(n)$  on  $V^{\otimes m}$  given by

$$\psi \rightarrow M_1 \otimes M_2 \otimes \dots \otimes M_m \psi; \quad M_a \in \mathcal{M}(n), \quad 1 \leq a \leq m.$$

According to Conjecture 5.1.2 there exists a Markov invariant,  $w$ , of degree  $d$  of this group action if

$$*^m \{k + s, k^{n-1}\} \ni \{d\},$$

for some  $nk + s = d$ . These Markov invariants will satisfy

$$w(M_1 \otimes M_2 \otimes \dots \otimes M_m \psi) = (\det(M_1) \det(M_2) \dots \det(M_m))^k w(\psi)$$

for all  $\psi \in V^{\otimes m}$ ,  $\forall M_a \in \mathcal{M}(n)$   $1 \leq a \leq m$ . The inner product multiplications computed for various cases by **Schur** are given in Table 5.1.

The Markov invariants can then be computed from

$$\Psi_{i_1 \dots i_{dm}} := Y_{\{k+s, k^{n-1}\}}^{(1)} Y_{\{k+s, k^{n-1}\}}^{(2)} \dots Y_{\{k+s, k^{n-1}\}}^{(m)} \psi_{i_1 \dots i_m} \psi_{i_{m+1} \dots i_{2m}} \dots \psi_{i_{(n-1)m} \dots i_{dm}},$$

where each Young operator  $Y_{\{k+s, k^{n-1}\}}^{(a)}$ ,  $1 \leq a \leq m$ , is generated from a standard tableau of shape  $\{k + s, k^{n-1}\}$  with integers chosen from the set  $\{a, m + a, \dots, (d - 1)m + a\}$ . The final step is to insert indices into  $\Psi$  using the semi-standard tableau:

0	0	...	0	0	0		0
1	1		1				
⋮							
n-1	n-1		n-1				

Again, the correct set of standard tableaux needed to generate a particular invariant is not certain, and we proceed by computing for different cases and checking for algebraic dependence until we get the correct number of algebraically independent invariants.

In what follows, we will adopt a notation where a Young operator corresponding to a certain tableau is written as  $Y_{a_1, a_2, \dots; b_1, b_2, \dots; c_1, \dots}$ , where the commas separate column entries in the tableau and semi-colons separate the rows.

n	2	2	3	3	4	4
m	{21}	{31}	{21 <sup>2</sup> }	{31 <sup>2</sup> }	{21 <sup>3</sup> }	{31 <sup>3</sup> }
2	1	1	1	1	1	1
3	1	1	1	1	0	1
4	3	4	4	13	4	16
5	5	10	10	61	6	137
6	11	31	31	397	40	1396

Table 5.1: Occurrences of  $\{d\}$  in  $*^m \{k + s, k^{n-1}\}$  with  $nk + s = d$



### 5.4.1 The stochastic invariant

For the group action of  $\times^m \mathcal{M}(n)$  there is always what is known as the degree  $d=1$  *stochastic invariant*,  $\Phi$ , for all  $m, n$  given by:

$$\Phi := \omega(\otimes^m \theta).$$

This corresponds to the trivial inner product multiplication

$$*^m \{1\} = \{1\},$$

with  $k = 0, s = 1$ . Evaluated on any tensor  $\psi \in V^{\otimes m}$  the stochastic invariant is simply the sum of the tensor components:

$$\Phi(\psi) = \sum_{i_1, i_2, \dots, i_m} \psi_{i_1 i_2 \dots i_m}.$$

In particular, evaluated on a phylogenetic tensor  $P$ :

$$\Phi(P) = \sum_{i_1, i_2, \dots, i_m} p_{i_1 i_2 \dots i_m} = 1,$$

which motivates the terminology.

### 5.4.2 The $n=2$ case

From Table 5.1 we see that for  $m = 2$  there is a single Markov invariant for each of  $d=3$  and  $d=4$ . These can be generated by simply taking pointwise products of the stochastic invariant with the general linear group invariant  $D_2$  (2.13):

$$\Phi \cdot D_2, \quad \Phi^2 \cdot D_2.$$

For  $m = 3$  there is a Markov invariant generated from  $\{21\}$ . We coin this invariant the *stangle* (stochastic tangle). By directed trial and error with various tableaux, this invariant was found by taking the composition of the three Young tableaux:

$$\begin{array}{|c|c|} \hline 1 & 7 \\ \hline 4 & \\ \hline \end{array} \quad \begin{array}{|c|c|} \hline 2 & 8 \\ \hline 5 & \\ \hline \end{array} \quad \begin{array}{|c|c|} \hline 3 & 9 \\ \hline 6 & \\ \hline \end{array} \quad .$$

This is written in our new notation as

$$\Psi_{i_1 i_2 i_3 i_4 i_5 i_6 i_7 i_8 i_9} := Y_{1,7;4} Y_{2,8;5} Y_{3,9;6} \psi_{i_1 i_2 i_3} \psi_{i_4 i_5 i_6} \psi_{i_7 i_8 i_9} \quad (5.5)$$

and we find that the stangle is

$$\begin{aligned} \mathcal{T}_2^s = \Psi_{000111000} = & -2\psi_{001}\psi_{010}\psi_{100} + \psi_{000}\psi_{011}\psi_{100} + \psi_{000}\psi_{010}\psi_{101} \\ & + \psi_{000}\psi_{001}\psi_{110} - \psi_{000}^2\psi_{111}. \end{aligned}$$

For  $m = 4$  there are three Markov invariants which we call the *squangles* (stochastic quangles). One of these Markov invariants can be generated simply by taking the pointwise product of the quangle multiplied by the stochastic invariant:

$$\Phi \cdot Q_2.$$

By directed trial and error the other two squangles have been found to be generated from

$$Y_{1,5;9}Y_{2,6;10}Y_{3,11;7}Y_{4,12;8}\psi_{i_1i_2i_3i_4}\psi_{i_5i_6i_7i_8}\psi_{i_9i_{10}i_{11}i_{12}} \quad (5.6)$$

and

$$Y_{1,5;9}Y_{2,10;6}Y_{3,11;7}Y_{4,12;8}\psi_{i_1i_2i_3i_4}\psi_{i_5i_6i_7i_8}\psi_{i_9i_{10}i_{11}i_{12}}.$$

Explicitly the first squangle is

$$\begin{aligned} Q_2^{s_1} = & \psi_{0011}\psi_{0100}\psi_{1000} + \psi_{0010}\psi_{0101}\psi_{1000} + \psi_{0001}\psi_{0110}\psi_{1000} - \psi_{0000}\psi_{0111}\psi_{1000} \\ & + \psi_{0010}\psi_{0100}\psi_{1001} + \psi_{0001}\psi_{0100}\psi_{1010} - \psi_{0000}\psi_{0100}\psi_{1011} - 2\psi_{0001}\psi_{0010}\psi_{1100} \\ & + 3\psi_{0000}\psi_{0011}\psi_{1100} - \psi_{0000}\psi_{0010}\psi_{1101} - \psi_{0000}\psi_{0001}\psi_{1110} + \psi_{0000}^2\psi_{1111}, \end{aligned}$$

and the second

$$\begin{aligned} Q_2^{s_2} = & \psi_{0011}\psi_{0100}\psi_{1000} - 2\psi_{0010}\psi_{0101}\psi_{1000} + \psi_{0001}\psi_{0110}\psi_{1000} - \psi_{0000}\psi_{0111}\psi_{1000} \\ & + \psi_{0010}\psi_{0100}\psi_{1001} - 2\psi_{0001}\psi_{0100}\psi_{1010} + 3\psi_{0000}\psi_{0101}\psi_{1010} - \psi_{0000}\psi_{0100}\psi_{1011} \\ & + \psi_{0001}\psi_{0010}\psi_{1100} - \psi_{0000}\psi_{0010}\psi_{1101} - \psi_{0000}\psi_{0001}\psi_{1110} + \psi_{0000}^2\psi_{1111}. \end{aligned}$$

The three degree  $d = 3$  Markov invariants  $\{\Phi \cdot Q_2, Q_2^{s_1}, Q_2^{s_2}\}$  have been shown by explicit computation to be linearly independent, as required.

### 5.4.3 The $n=3$ case

From Table 5.1, there are two Markov invariants for  $n = 3, m = 2$  of degree  $d = 4, 5$ . Again these invariants can be easily produced by taking products of the stochastic invariant with the determinant invariant (2.13):

$$\Phi \cdot D_3, \quad \Phi^2 \cdot D_3.$$

In the case  $m=3$  there is a single Markov invariant, which we also refer to as the *stangle*:

$$\begin{aligned} \Psi_{i_1i_2i_3i_4i_5i_6i_7i_8i_9i_{10}i_{11}i_{12}} \\ := Y_{1,4;7;10}Y_{2,8;5;11}Y_{3,12;6;9}\psi_{i_1i_2i_3}\psi_{i_4i_5i_6}\psi_{i_7i_8i_9}\psi_{i_{10}i_{11}i_{12}}, \end{aligned} \quad (5.7)$$

so that

$$\begin{aligned}
\mathcal{T}_3^{(s)} &= \Psi_{000011102220} \\
&= \psi_{012}\psi_{020}\psi_{101}\psi_{200} - \psi_{010}\psi_{022}\psi_{101}\psi_{200} - \psi_{011}\psi_{020}\psi_{102}\psi_{200} \\
&+ \psi_{010}\psi_{021}\psi_{102}\psi_{200} - \psi_{002}\psi_{021}\psi_{110}\psi_{200} + \psi_{001}\psi_{022}\psi_{110}\psi_{200} \\
&+ \psi_{002}\psi_{011}\psi_{120}\psi_{200} - \psi_{001}\psi_{012}\psi_{120}\psi_{200} - \psi_{012}\psi_{020}\psi_{100}\psi_{201} \\
&+ \psi_{010}\psi_{022}\psi_{100}\psi_{201} + \psi_{002}\psi_{020}\psi_{110}\psi_{201} - \psi_{000}\psi_{022}\psi_{110}\psi_{201} \\
&- \psi_{002}\psi_{010}\psi_{120}\psi_{201} + \psi_{000}\psi_{012}\psi_{120}\psi_{201} + \psi_{011}\psi_{020}\psi_{100}\psi_{202} \\
&- \psi_{010}\psi_{021}\psi_{100}\psi_{202} - \psi_{001}\psi_{020}\psi_{110}\psi_{202} + \psi_{000}\psi_{021}\psi_{110}\psi_{202} \\
&+ \psi_{001}\psi_{010}\psi_{120}\psi_{202} - \psi_{000}\psi_{011}\psi_{120}\psi_{202} + \psi_{002}\psi_{021}\psi_{100}\psi_{210} \\
&- \psi_{001}\psi_{022}\psi_{100}\psi_{210} - \psi_{002}\psi_{020}\psi_{101}\psi_{210} + \psi_{000}\psi_{022}\psi_{101}\psi_{210} \\
&+ \psi_{001}\psi_{020}\psi_{102}\psi_{210} - \psi_{000}\psi_{021}\psi_{102}\psi_{210} - \psi_{002}\psi_{011}\psi_{100}\psi_{220} \\
&+ \psi_{001}\psi_{012}\psi_{100}\psi_{220} + \psi_{002}\psi_{010}\psi_{101}\psi_{220} - \psi_{000}\psi_{012}\psi_{101}\psi_{220} \\
&- \psi_{001}\psi_{010}\psi_{102}\psi_{220} + \psi_{000}\psi_{011}\psi_{102}\psi_{220}.
\end{aligned}$$

In the case of  $m = 4$ , Table 5.1 predicts four Markov invariants, which we again refer to as squangles. One of the squangles can be inferred directly as the pointwise product:

$$\Phi \cdot Q_3,$$

and by directed trial and error we have shown that the other three can be generated from the Young operators:

$$\begin{aligned}
Q_3^{s_1} &\leftarrow Y_{1,5;9;13}Y_{2,6;10;14}Y_{3,7;11;15}Y_{4,8;12;16}, \\
Q_3^{s_2} &\leftarrow Y_{1,9;5;13}Y_{2,14;6;10}Y_{3,7;11;15}Y_{4,8;12;16}, \\
Q_3^{s_3} &\leftarrow Y_{1,9;5;13}Y_{2,10;6;14}Y_{3,7;11;15}Y_{4,8;12;16},
\end{aligned} \tag{5.8}$$

where  $\leftarrow$  indicates the implementation of our procedure with the indices of  $\Psi$  filled out to create the only semi-standard tableau of shape  $\{21^2\}$  using the integers  $\{0, 1, 2\}$ . The four invariants  $\{\Phi \cdot Q_3, Q_3^{s_1}, Q_3^{s_2}, Q_3^{s_3}\}$  have been shown by explicit computation to be linearly independent.

#### 5.4.4 The $n=4$ case

In the case of  $n = 4, m = 2$ , Table 5.1 predicts a Markov invariant of degree  $d = 5, 6$ . Again, these invariants can be generated easily as the pointwise products:

$$\Phi \cdot D_4, \quad \Phi^2 \cdot D_4.$$

In the case of  $m=3$  Table 5.1 predicts a degree  $d=6$  Markov invariant which we again refer to as the *stangle*. It is generated from the Young operator

$$\mathcal{T}_4^s \leftarrow Y_{1,4,13;7,10,16}Y_{2,8,17;5,11,14}Y_{3,12,18;6,9,15}. \tag{5.9}$$

Explicitly this polynomial has 1404 terms.

In the case of  $m=4$  there are four degree  $d=5$  Markov invariants which we again refer to as squangles. One of these is generated easily as

$$\Phi \cdot Q_4,$$

and by directed trial and error the other three have been found to be given by the Young operators:

$$\begin{aligned} Q_4^{s_1} &\leftarrow Y_{1,5;9;13;17} Y_{2,6;10;14;18} Y_{3,7;11;15;19} Y_{4,8;12;16;20}, \\ Q_4^{s_2} &\leftarrow Y_{1,9;5;13;17} Y_{2,14;6;10;18} Y_{3,7;11;15;19} Y_{4,8;12;16;20}, \\ Q_4^{s_3} &\leftarrow Y_{1,9;5;13;17} Y_{2,14;6;10;18} Y_{3,19;7;11;15} Y_{4,8;12;16;20}. \end{aligned} \quad (5.10)$$

The four degree  $d=5$  Markov invariants  $\{\Phi \cdot Q_4, Q_4^{s_1}, Q_4^{s_2}, Q_4^{s_3}\}$  have been shown by explicit computation to be linearly independent, as required.

## 5.5 What happens on a phylogenetic tree?

In this section we will examine the structure of the invariant functions we have discovered on phylogenetic trees. We will focus on the case of four characters  $n=4$  and three and four leaves  $m=3, 4$ .

We have discovered invariant functions which satisfy

$$w(g\psi) = \det(g)^k w(\psi),$$

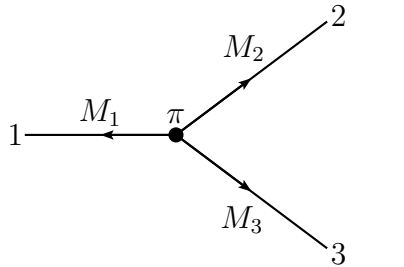
for all  $g \in \times^m \mathcal{M}(n)$  and  $\psi \in V^{\otimes m}$ . If we consider the case where these invariants are evaluated on the phylogenetic tensor  $P$ , the invariant takes the form

$$w(P) = \prod_{a=1}^m \det(M_a)^k w(\tilde{P}).$$

Our task is to examine the structure of the Markov invariants when evaluated on the phylogenetic tensor  $\tilde{P}$  corresponding to the various possible trees.

### 5.5.1 The stangle

As we saw in Chapter 4, we need only consider *unrooted* phylogenetic trees. For the case of three taxa the most general phylogenetic tree is:



The corresponding phylogenetic tensor can be expressed as

$$P = (M_1 \otimes M_2 \otimes M_3)1 \otimes \delta \cdot \delta \cdot \pi,$$

where

$$\delta^2 := 1 \otimes \delta \cdot \delta = \delta \otimes 1 \cdot \delta.$$

From the general properties of the Markov invariants we find that

$$\mathcal{T}^{(s)}(P) = \det(M_1) \det(M_2) \det(M_3) \mathcal{T}^{(s)}(\tilde{P}),$$

and by direct computation

$$\mathcal{T}^{(s)}(\tilde{P}) = 0.$$

It follows that evaluating the stangle on the general phylogenetic tensor of four leaves satisfies

$$\mathcal{T}^{(s)}(P) = 0.$$

This equation is independent of all the model parameters contained in the phylogenetic tree. This observation implies that this Markov invariant also satisfies the properties of a *phylogenetic invariant* for the general Markov model [1].

### 5.5.2 The squangles

For the case of four taxa there are three inequivalent unrooted phylogenetic trees as presented in Figure 5.1. The corresponding phylogenetic tensors are

- $P^{(1)} = M_1 \otimes M_2 \otimes M_3 \otimes M_4(1 \otimes 1 \otimes \delta M_5) \delta^2 \pi$
- $P^{(2)} = M_1 \otimes M_2 \otimes M_3 \otimes M_4(1 \otimes \delta M_5 \otimes 1) \delta^2 \pi$
- $P^{(3)} = M_1 \otimes M_2 \otimes M_3 \otimes M_4(\delta M_5 \otimes 1 \otimes 1) \delta^2 \pi.$

For any linear combination of the Markov invariants:

$$w = c\Phi \cdot Q_4 + c_1 Q_4^{s_1} + c_2 Q_4^{s_2} + c_3 Q_4^{s_3}$$

we have

$$w(P^{(1)}) = \det(M_1) \det(M_2) \det(M_3) \det(M_4) w(\tilde{P}^{(a)}), \quad a = 1, 2, 3.$$

Defining the linearly independent combinations

$$\begin{aligned} L_1 &= -\frac{3}{2} Q_4^{s_1} + Q_4^{s_2} + 2Q_4^{s_3}, \\ L_2 &= -\frac{3}{2} Q_4^{s_1} + 2Q_4^{s_2} + Q_4^{s_3}, \\ L_3 &= -Q_4^{s_2} + Q_4^{s_3}. \end{aligned}$$

it is possible to show by direct computation that the following relations hold:

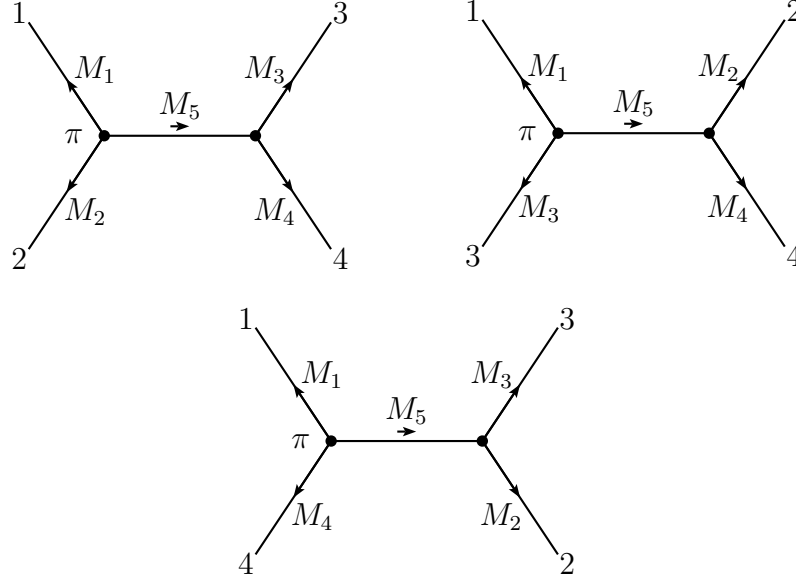


Figure 5.1: Three alternative quartet trees

- $L_1(\tilde{P}^{(1)}) = 0, \quad L_2(\tilde{P}^{(1)}) = -L_3(\tilde{P}^{(1)}) > 0;$
- $L_2(\tilde{P}^{(2)}) = 0, \quad L_1(\tilde{P}^{(1)}) = L_3(\tilde{P}^{(1)}) > 0;$
- $L_3(\tilde{P}^{(3)}) = 0, \quad L_1(\tilde{P}^{(1)}) = L_2(\tilde{P}^{(1)}) < 0.$

This implies that these linear combinations of the squangles are not only Markov invariants, but also phylogenetic invariants [1]. They are actually phylogenetically *informative* invariants because they can be used to distinguish between the three quartet topologies. Studying the statistical properties of this technique is a topic of ongoing work (see Appendix A).

## 5.6 Review of important invariants

We tabulate the invariant functions that have been of interest in this thesis in Table 5.2. It should be noted that in the case of the squangles the invariants of the general linear group are included with the invariants of the Markov semigroup.

## 5.7 Closing remarks

In this chapter we have defined and proved the existence of Markov invariants. We have shown how to derive their explicit polynomial form in interesting cases. We examined the structure of several invariants in the context of phylogenetic trees. Finally, we derived a novel technique of quartet tree

Name	Symbol	Schur multi.	Group	$(d, k)$	Ref.
<i>det</i>	$\det_2$	$*^2\{1^2\} = \{2\}$	$\times^2 GL(2)$	(2,1)	(2.13)
	$\det_3$	$*^2\{1^3\} = \{3\}$	$\times^2 GL(3)$	(3,1)	(2.13)
	$\det_4$	$*^2\{1^4\} = \{4\}$	$\times^2 GL(4)$	(4,1)	(2.13)
<i>tangle</i>	$\mathcal{T}_2$	$*^3\{2^2\} \ni \{4\}$	$\times^3 GL(2)$	(4,2)	(2.15)
	$\mathcal{T}_3$	$*^3\{2^3\} \ni \{6\}$	$\times^3 GL(3)$	(6,2)	(2.16)
	$\mathcal{T}_4$	$*^3\{2^4\} \ni \{8\}$	$\times^3 GL(4)$	(8,2)	(2.17)
<i>stangle</i>	$\mathcal{T}_2^s$	$*^3\{21\} \ni \{3\}$	$\times^3 \mathcal{M}(2)$	(3,1)	(5.5)
	$\mathcal{T}_3^s$	$*^3\{21^2\} \ni \{4\}$	$\times^3 \mathcal{M}(3)$	(4,1)	(5.7)
	$\mathcal{T}_4^s$	$*^3\{31^3\} \ni \{6\}$	$\times^3 \mathcal{M}(4)$	(6,1)	(5.9)
<i>quangle</i>	$Q_2$	$*^4\{1^2\} \ni \{2\}$	$\times^4 GL(2)$	(2,1)	(2.14)
	$Q_3$	$*^4\{1^3\} \ni \{3\}$	$\times^4 GL(3)$	(3,1)	(2.14)
	$Q_4$	$*^4\{1^4\} \ni \{4\}$	$\times^4 GL(4)$	(4,1)	(2.14)
<i>squangle</i>	$(Q_2, Q_2^{s_1}, Q_2^{s_2})$	$*^4\{21\} \ni 3\{3\}$	$\times^4 \mathcal{M}(2)$	(3,1)	(5.6)
	$(Q_3, Q_3^{s_1}, Q_3^{s_2}, Q_3^{s_3})$	$*^4\{21^2\} \ni 4\{4\}$	$\times^4 \mathcal{M}(3)$	(3,1)	(5.8)
	$(Q_4, Q_4^{s_1}, Q_4^{s_2}, Q_4^{s_3})$	$*^4\{21^3\} \ni 4\{5\}$	$\times^4 \mathcal{M}(4)$	(5,1)	(5.10)

Table 5.2: Invariant functions satisfying  $f \circ g = \det(g)^k f$ 

reconstruction which is valid under the assumptions of the general Markov model of sequence evolution.

## CHAPTER 6

# Conclusion

In this thesis we have examined the mathematical analogy between quantum physics and the Markov model of a phylogenetic tree.

In Chapter 2 we gave a review of group representation theory, established the Schur/Weyl duality and went on to show how one-dimensional representations and invariant functions of the general linear group can be put into coincidence. We also presented several examples of the explicit polynomial form of these invariants.

In Chapter 3 we concretely established the mathematical analogy between entanglement and that of phylogenetic relation. We showed that group invariant functions can be used to quantify a measure of phylogenetic relation.

In Chapter 4 we gave a review of pairwise phylogenetic distance measures and examined the use of the tangle in improving the calculation of pairwise distance measures from observed sequence data.

In Chapter 5 we defined and showed how to derive Markov invariant functions. We studied their properties in cases relevant to the problem of phylogenetic tree reconstruction. We derived a new technique for reconstruction of quartets which is valid under the assumptions of a general Markov model.

### Future investigations

There are several clear paths for continuing the work that has been presented in this thesis.

Rather than use the tangle to give improved pairwise distances it seems judicious to examine how the tangle could be used in more direct ways. The Neighbour-Joining (NJ) algorithm for tree reconstruction has at its core the concept of pairwise distances and in opposition to this the tangle polynomial actually gives a measure of the sum of the branch lengths for a triplet. Hence it seems that one possibility is to generalize the NJ algorithm in such a way that the tangle is incorporated explicitly into the procedure. Additionally, biologists are interested in the evolutionary distance between taxa and another possibility would be to use the tangle as a measure of the evolutionary distance



between triplets of taxa without decomposing this distance into pairs. Given a set of multiple taxa one could construct interesting questions comparing different triplets using the value of the tangle as a quantifier.

The stochastic tangle is a very interesting mathematical object as it simultaneously satisfies the properties of a Markov invariant and that of a phylogenetic invariant. In this thesis we have not investigated the potential of finding a practical role for the stochastic tangle in the problem of phylogenetic reconstruction. The possibilities of practical roles are similar to that of the tangle and we leave this as an open problem.

The squangles have been shown to give a new tree reconstruction algorithm for the case of quartets. The main path for future investigation is to study the statistical properties of such an algorithm. It is theoretically clear how to calculate unbiased forms of the squangles (see Appendix A) and this would be a desirable practical outcome as it will improve the performance of the quartet reconstruction in the case where the sequence data is of relatively short length. Unfortunately this calculation of an unbiased form is computationally difficult and has not been achieved. To further the complete statistical understanding it is necessary to calculate the variance of the squangles. Again this is theoretically clear but computationally difficult as one is required to square the polynomials.

In this thesis we have used the concept of a tree in a rather ad hoc way. Our procedure was to compute the explicit polynomial form of the invariant functions and then to impose a given tree structure onto the polynomial by choosing coordinates for the tensors selected to be consistent with the tree. Given that the existence of the invariant functions was proved using the Schur functions series, a natural corollary would be to ask if it is possible to identify the relationships between the invariant functions that occur on particular trees by simply studying the properties of the Schur functions in more detail. The branching operator  $\delta$  is technically an invertible linear operator on the expanded linear space known as a *Fock* space and it follows that the character theory of this action together with that of the Markov semigroup should introduce the possibility of “seeing” the tree structure within the Schur functions. Hence it seems feasible to identify the relationships between the invariant functions that occur on particular trees by simply studying the properties of the Schur functions in more detail.

The other clear course for theoretical investigation is to completely classify the ring of invariants for the Markov semigroup. This is not an easy problem as the Hilbert basis theorem states that the ring of invariants is guaranteed to be finitely generated if the group action is completely reducible [36]. However, the Markov group has an invariant subspace with no complementary invariant subspace and is hence not completely reducible. Further study is required to fully characterize the ring of Markov invariants. Additionally, the exact connection between the ring of Markov invariants and the ideal of phylogenetic invariants should be established concretely. In this thesis this connection was only made for the particular cases that were of interest. A well defined and complete description of the connection is required before one can speak with confidence on this matter.

# APPENDIX A

## Bias correction of invariant functions

### A.1 Multinomial distribution

Let  $X_a$ ,  $1 \leq a \leq n$ , be the random variable which counts the occurrences of character  $a$  in a finite subset of an infinite sequence consisting of the characters  $\{1, 2, \dots, n\}$ . If each character occurs with probability  $p_a$ , then for a subset of length  $N$  we have the standard multinomial distribution

$$\mathbb{P}(X_1 = k_1, X_2 = k_2, \dots, X_n = k_n) = \frac{N!}{k_1! k_2! \dots k_n!} p_1^{k_1} p_2^{k_2} \dots p_n^{k_n}. \quad (\text{A.1})$$

Defining the vector valued random variable  $X = (X_1, X_2, \dots, X_n) \in \mathbb{N}^n$ , we can express (A.1) as

$$\mathbb{P}(X = k) = \frac{N!}{\prod_{a=1}^n k_a!} \prod_{b=1}^n p_b^{k_b},$$

with  $k = (k_1, \dots, k_n) \in \mathbb{N}^n$  and  $k_1 + k_2 + \dots + k_n = N$ . Consider any function

$$\phi : \mathbb{C}^n \rightarrow \mathbb{C}^q, \quad q \in \mathbb{N}.$$

The expectation value of  $\phi(X)$  is then defined as

$$E[\phi(X)] = \sum_{k \in \mathbb{N}: k_1 + k_2 + \dots + k_n = N} \mathbb{P}(X = k) \phi(k).$$

### A.2 Generating function

For every  $s \in \mathbb{R}^n$  we define the generating function  $G : \mathbb{R}^n \rightarrow \mathbb{C}$  as

$$G(s) = E[e^{i(s, X)}],$$

where we have considered  $X \in \mathbb{N}^n \subset \mathbb{R}^n$  and  $(s, X) = s_1 X_1 + s_2 X_2 + \dots + s_n X_n$

and convergence is ensured by  $|e^{i(s,X)}| = 1$  and the triangle inequality. Observe that

$$\frac{\partial G(s)}{\partial s_j} = E[iX_j e^{i(s,X)}].$$

In particular we have

$$\frac{\partial G(s)}{\partial s_j} \Big|_{s=0} = iE[X_j].$$

We simplify notation by taking the Laplace transform

$$s \rightarrow is,$$

and find that in general

$$\frac{\partial^{b_1+b_2+\dots+b_m} G(s)}{\partial s_{a_1}^{b_1} \partial s_{a_2}^{b_2} \dots \partial s_{a_m}^{b_m}} \Big|_{s=0} = E[X_{a_1}^{b_1} X_{a_2}^{b_2} \dots X_{a_m}^{b_m}].$$

Computing a closed form of  $G(s)$  follows easily given the identity

$$(x_1 + x_2 + \dots + x_n)^N = \sum_{k \in \mathbb{N}^n: k_1+k_2+\dots+k_n=N} \frac{N!}{k_1!k_2!\dots k_n!} x_1^{k_1} x_2^{k_2} \dots x_n^{k_n},$$

so that

$$G(s) = (p_1 e^{s_1} + p_2 e^{s_2} + \dots + p_n e^{s_n})^N.$$

In particular  $G(0) = 1$ .

## A.3 Expectations of polynomials

We are particularly interested in the case when

$$\phi \in \mathbb{C}[V]_d, \quad V \cong \mathbb{C}^n.$$

In general we have

$$E[(\phi_1 + c\phi_2)(X)] = E[\phi_1(X)] + cE[\phi_2(X)],$$

but

$$E[\phi_1 \cdot \phi_2(X)] \neq E[\phi_1(X)]E[\phi_2(X)].$$

Thus in order to calculate the expected value of a polynomial we need only study expectation values of monomials:

$$E[X_{a_1}^{b_1} X_{a_2}^{b_2} \dots X_{a_m}^{b_m}], \quad m \leq n.$$

In particular we have

$$\begin{aligned}
E[X_a] &= \frac{\partial G(s)}{\partial s_a} \Big|_{s=0} \\
&= Np_a, \\
E[X_a X_b] &= \frac{\partial^2 G(s)}{\partial s_a \partial s_b} \Big|_{s=0} \\
&= N(N-1)p_a p_b + Np_a \delta_{ab}, \\
E[X_a X_b X_c] &= \frac{\partial^3 G(s)}{\partial s_a \partial s_b \partial s_c} \Big|_{s=0} \\
&= N(N-1)(N-2)p_a p_b p_c \\
&\quad + N(N-1)(p_a p_b \delta_{ac} + p_a p_c \delta_{ab} + p_b p_c \delta_{ab}) + Np_a \delta_{ab} \delta_{ac},
\end{aligned} \tag{A.2}$$

and for a set of distinct integers  $1 \leq a_1, a_2, \dots, a_d \leq n$  we have

$$E[X_{a_1} X_{a_2} \dots X_{a_d}] = \frac{N!}{(N-d)!} p_{a_1} p_{a_2} \dots p_{a_m}. \tag{A.3}$$

## A.4 Bias correction

For a given homogeneous polynomial  $\phi$  of degree  $d$ , we would like to find a polynomial  $\widetilde{\phi}$  such that

$$E[\widetilde{\phi}(X)] = \phi(p).$$

We refer to  $\widetilde{\phi}$  as the unbiased form of  $\phi$ .

By looking at the general form of the invariants  $\det_n$  it can be seen that every monomial term is of the form (A.3). It follows easily that

$$E[\det_n(X)] = \frac{N!}{(N-n)!} \det_n(p),$$

so that the unbiased version is given simply by

$$\widetilde{\det_n} := \frac{(N-d)!}{N!} \det_n.$$

It should be noted that this says nothing about what to do about finding an unbiased form of  $\log \det$ , because the  $\log$  function is *not* polynomial. For discussion on the bias correction of the  $\log \det$  function see [5].

We leave the computation of unbiased forms of the other invariants presented in this thesis as an open problem. However, the process is exemplified in the following.

Consider the expectation:

$$E[X_1 X_2 X_3] = N(N-1)(N-2)p_1 p_2 p_3.$$

Thus the unbiased form of this monomial is simply

$$\frac{(N-3)!}{N!} X_1 X_2 X_3.$$

Consider

$$E[X_1^2 X_2] = N(N-1)(N-2)p_1^2 p_2 + N(N-1)p_1 p_2.$$

The unbiased form of this monomial is then

$$\frac{(N-3)!}{N!} (X_1^2 X_2 - X_1 X_2),$$

since

$$E\left[\frac{(N-3)!}{N!} (X_1^2 X_2 - X_1 X_2)\right] = p_1^2 p_2.$$

By generalizing (A.2) for a set of distinct integers  $1 \leq a, b_1, b_2, \dots, b_m \leq n$  it follows that

$$E\left[\frac{N!}{(N-(m+1))!} (X_a^2 X_{b_1} X_{b_2} \dots X_{b_m} - X_a X_{b_1} X_{b_2} \dots X_{b_m})\right] = p_a^2 p_{b_1} p_{b_2} \dots p_{b_m}.$$

This is the first step to computing the unbiased form of general monomials. Clearly the process becomes more complicated as the degree of a given random variable within each monomial becomes larger.

# BIBLIOGRAPHY

- [1] E. S. Allman and J. A. Rhodes. Phylogenetic invariants of the general Markov model of sequence mutation. *Mathematical Biosciences*, 186:113–144, 2003.
- [2] E. Baake. What can and what cannot be inferred from pairwise sequence comparisons? *Mathematical Biosciences*, 154:1–21, 1998.
- [3] E. Baake and A. Haeseler. Distance measures in terms of substitution processes. *Theoretical Population Biology*, 55:166–175, 2001.
- [4] T. H. Baker. *Symmetric Functions and Infinite Dimensional Algebras*. PhD thesis, University of Tasmania, 1994.
- [5] D. Barry and J. A. Hartigan. Asynchronous distance between homologous DNA sequences. *Biometrics*, 43(2):261–276, 1987.
- [6] J. D. Bashford, P. D. Jarvis, J. G. Sumner, and M. A. Steel.  $U(1) \times U(1) \times U(1)$  symmetry of the Kimura 3ST model and phylogenetic branching process. *Journal of Physics A: Mathematical and General*, 37:L1–L9, 2004.
- [7] J. Bell. On the Einstein-Podolsky-Rosen paradox. *Physics*, 1:195–200, 1964.
- [8] B. A. Bernevig and H. D. Chen. Geometry of the three-qubit state, entanglement and division algebras. *Journal of Physics A: Mathematical and General*, 36(30):8325–8339, 2003.
- [9] D. Bryant and M. Steel. Constructing optimal trees from quartets. *Journal of Algorithms*, 38:237–259, 2001.
- [10] M. J. Carvalho and S. D’Agostino. Plethysms of Schur functions and the shell model. *Journal of Physics A: Mathematical and General*, 34:1375–1392, 2001.
- [11] J. A. Cavender and J. Felsenstein. Invariants of phylogenies in a simple case with discrete states. *Journal of Classification*, 4:57–71, 1987.
- [12] J. T. Chang. Full reconstruction of Markov models on evolutionary trees: identifiability and consistency. *Mathematical Biosciences*, 137(1):51–73, 1996.
- [13] P. A. M. Dirac. *The Principles of Quantum Mechanics*. Oxford Clarendon Press, 1958.

- [14] W. Dur, G. Vidal, and J. I. Cirac. Three qubits can be entangled in two inequivalent ways. *Physics Review A*, 62(6):062314, 2000.
- [15] S. N. Evans and T. P. Speed. Invariants of some probability models used in phylogenetic inference. *Annals of Statistics*, 21(1):355–377, 1993.
- [16] B. Fauser, P. D. Jarvis, and R. C. King. A Hopf algebraic approach to the theory of group branchings. In R.C. King, M. Bylicki, and J. Karwowski, editors, *Symmetry, Spectroscopy and SCHUR: Proceedings of the Professor Brian G. Wybourne Commemorative Meeting, Torun, Poland*. Nicolaus Copernicus University Press, 2006.
- [17] B. Fauser, P. D. Jarvis, R. C. King, and B. G. Wybourne. New branching rules induced by plethysm. *Journal of Physics A: Mathematical and General*, 39:2611–2655, 2005.
- [18] W. Feller. *An Introduction to Probability Theory and Its Applications*. John Wiley and Sons, Inc., 1968.
- [19] J. Felsenstein. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution*, 17:368–376, 1981.
- [20] J. Felsenstein. Counting phylogenetic invariants in some simple cases. *Journal of Theoretical Biology*, 152:357–376, 1991.
- [21] J. Felsenstein. *Inferring Phylogenies*. Sinauer Associates, 2004.
- [22] R. Feynman. *QED: The Strange Theory of Light and Matter*. Princeton University Press, 1988.
- [23] O. Gascuel. BIONJ: An improved version of the nj algorithm based on a simple model of sequence data. *Molecular Biology and Evolution*, 14(7):685–695, 1987.
- [24] G. S. Goodman. An intrinsic time for nonstationary finite Markov chains. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 16:165–180, 1973.
- [25] X. Gu and W. H. Li. Bias-corrected paralinear and logdet distances and tests of molecular clocks and phylogenies under non-stationary nucleotide frequencies. *Molecular Biology and Evolution*, 13(10):1375–1383, 1996.
- [26] O. Guhne and P. Hyllus. Investigating three qubit entanglement with local measurements. *International Journal of Theoretical Physics*, 42:1001–1013, 2003.
- [27] M. Hamermesh. *Group Theory and Its Application to Physical Problems*. Addison-Wesley, 1964.
- [28] K. Hoffman and R. Kunze. *Linear Algebra (2nd Edition)*. Prentice Hall, 1971.
- [29] M. Iosifescu. *Finite Markov Processes and Their Applications*. John Wiley and Sons, Chichester, 1980.

- [30] C. Itzykson and J-B. Zuber. *Quantum Field Theory*. McGraw-Hill, New York, 1980.
- [31] P. D. Jarvis and J. D. Bashford. Quantum field theory and phylogenetic branching. *Journal of Physics A: Mathematical and General*, 34:L703–L707, 2001.
- [32] P. D. Jarvis, J. D. Bashford, and J. G. Sumner. Path integral formulation and Feynman rules for phylogenetic branching models. *Journal of Physics A: Mathematical and General*, 38:9621–9647, 2005.
- [33] L. Jermiin, S. Y. Ho, F. Ababneh, J. Robinson, and A. W. Larkum. The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimated. *Systematic Biology*, 53(4):638–643, 2004.
- [34] J. E. Johnson. Markov-type Lie group in  $GL(n, \mathbb{R})$ . *Journal of Mathematical Physics*, 26(2):252–257, 1985.
- [35] R. Keown. *An Introduction To Group Representation Theory*. Academic Press, New York, 1975.
- [36] H. Kraft and C. Procesi. Classical Invariant Theory, A Primer. <http://www.math.unibas.ch/kraft/Papers/KP-Primer.pdf>, 2000.
- [37] J. A. Lake. Reconstructing evolutionary trees from DNA and protein sequences: Paralinear distances. *Proceedings of the National Academy of Sciences*, 91:1455–1459, 1994.
- [38] N. Linden and S. Popescu. On multi-particle entanglement. *Fortschritte der Physik*, 46:567–578, 1998.
- [39] D. E. Littlewood. *The Theory of Group Characters*. Oxford at the Clarendon Press, 1940.
- [40] P. J. Lockhart, M. A. Steel, M. D. Hendy, and D. Penny. Recovering evolutionary trees under a more realistic model of sequence evolution. *Molecular Biology and Evolution*, 11(4):605–612, 1994.
- [41] I. G. MacDonald. *Symmetric Functions and Hall Polynomials*. Clarendon Press, Oxford, 1979.
- [42] W. Miller. *Symmetry Groups and Their Applications*. Academic Press, New York, 1972.
- [43] A. Miyake. Classification of multiparticle entangled states by multidimensional determinants. *Physics Review A*, 67:012108, 2003.
- [44] M. Nei and S. Kumar. *Molecular Evolution and Phylogenetics*. Oxford University Press, Oxford, 2000.
- [45] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.



- [46] P. J. Olver. *Classical Invariant Theory*. Cambridge University Press, Cambridge, 2003.
- [47] A. Pais. *Inward Bound*. Oxford University Press, 1988.
- [48] J. Pearl and M. Tarsi. Structuring causal trees. *Journal of Complexity*, 2:66–77, 1986.
- [49] W. Pearson, G. Robins, and T. Zhang. Generalised neighbor-joining: More reliable phylogenetic tree reconstruction. *Molecular Biology and Evolution*, 16(6):806–816, 1999.
- [50] A. Rindos, S. Woollet, I. Viniotis, and K. Trivedi. Exact methods for the transient analysis of nonhomogeneous continuous time Markov chains. In William J. Stewart, editor, *2nd International Workshop on the Numerical Solution of Markov Chains*. Kluwer Academic Publishers, 1995.
- [51] F. Rodriguez, J. L. Oliver, A. Marin, and J. R. Medina. The general stochastic model of nucleotide substitution. *Journal of Theoretical Biology*, 142:485–501, 1990.
- [52] N. Saitou and M. Nei. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4(4):406–425, 1987.
- [53] C. Semple and M. Steel. *Phylogenetics*. Oxford Press, 2003.
- [54] M. Steel, M. D. Hendy, and D. Penny. Reconstructing phylogenies from nucleotide pattern probabilities: A survey and some new results. *Discrete Applied Mathematics*, 88:367–396, 1998.
- [55] M. Steel, L. Szekely, P. L. Erdos, and P. Waddell. A complete family of phylogenetic invariants for any number of taxa under Kimura’s 3ST model. *New Zealand Journal of Botany*, 31(31):289–296, 1993.
- [56] M. A. Steel. Recovering a tree from the leaf colourations it generates under a Markov model. *Applied Mathematics Letters*, 7(2):19–24, 1994.
- [57] M. A. Steel, L. Szekely, and M. Hendy. Reconstructing trees when sequence site evolve at variable rates. *Journal of Computational Biology*, 1(2):153–163, 1994.
- [58] K. Strimmer and A. Haeseler. Quartet puzzling: a quartet maximum-likelihood method for the reconstructing of tree topologies. *Molecular Biology and Evolution*, 13(7):964–969, 1996.
- [59] J. G. Sumner and P. D. Jarvis. Entanglement invariants and phylogenetic branching. *Journal of Mathematical Biology*, 51(1):18–36, 2005.
- [60] J. G. Sumner and P. D. Jarvis. Using the tangle: a consistent construction of phylogenetic distance matrices. *Mathematical Biosciences*, 204:49–67, 2006.

- [61] P. Szekeres. *A Course In Modern Mathematical Physics*. Cambridge University Press, 2004.
- [62] R. F. Werner and M. M. Wolf. Bell inequalities and entanglement. *Quantum Information and Computation*, 1(3):1–25, 2001.
- [63] Brian G Wybourne. **SCHUR**, Schur Group Theory Software. <http://smc.vnet.net/schur.html>, 2004.