Invariant polynomials and machine learning

Ward Issa Jereis Haddadin

Department of Applied Mathematics and Theoretical Physics
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

King's College June 2021

إلى

عيسى و عريب

Declaration

This dissertation is based on original research carried out while the author was a graduate student at the Department of Applied Mathematics and Theoretical Physics (DAMTP), University of Cambridge, from October 2018 to June 2021. The material in Chapters 3 and 4 is based on work done by the author with Ben Gripaios and Christopher G. Lester, most of which is published in [18]. More precisely, apart from Theorem 3.3.1 which was proved by B. Gripaios, all work in Chapters 3 and 4 relating to the orthogonal (and special orthogonal) group was done by the author under the supervision of B. Gripaios and all work relating to the euclidean (and special euclidean) group was done by the author alone. Chapters 5 and 6 are based on work done by the author, which is currently under review [20].

No part of this work has been submitted, or is being concurrently submitted, for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution.

Ward Issa Jereis Haddadin
June 2021

Abstract

Invariant polynomials and machine learning

Ward Issa Jereis Haddadin

In this thesis, we demonstrate the benefit of incorporating our knowledge of the symmetries of certain systems into the machine learning methods used to model them. By borrowing the necessary tools from commutative algebra and invariant theory, we construct systematic methods to obtain sets of invariant variables that describe two such systems: particle physics and physical chemistry. In both cases, our systems are described by a collection of vectors. In the former, these vectors represent the particle momenta in collision events where our system is Lorentz- and permutation-invariant. In the latter, the vectors represent the positions of atoms in molecules (or lattices) where our systems are Euclidean- and permutation-invariant.

We start by focusing on the algebras of invariant polynomials in these vectors and design systematic methods to obtain sets of generating variables. To do so, we build on two theorems of Weyl which tell us that the algebra of orthogonal group-invariant polynomials in n d-dimensional vectors is generated by the dot products and that the redundancies which arise when $n > d$ are generated by the $(d+1)$-minors of the $n \times n$ matrix of dot products. We prove the equivalent theorems for the algebra of polynomials invariant under the Euclidean group which provide us with a similar set of variables for describing molecular properties. We also extend these results to include the action of an arbitrary permutation group $P \subset S_n$ on the vectors. Doing so furnishes us with sets of variables for describing molecules with identical atoms or scattering processes involving identical particles, such as $pp \to jjj$, for which we provide an explicit minimal set of Lorentz- (and parity-) and permutation-invariant generators. Additionally, we use the Cohen-Macaulay structure of the Lorentz-invariant algebra to provide a more direct characterisation in terms of a Hironaka decomposition. Although such a characterisation would naïvely not be expected for the Euclidean-invariant

algebra (since the Euclidean group is not linearly reductive), we show that it is in fact also Cohen-Macaulay by establishing that it is isomorphic to another Cohen-Macaulay algebra, namely the Lorentz-invariant algebra in one vector fewer. Among the benefits of Hironaka decompositions is that they can be generalized straightforwardly to when parity is not a symmetry and to cases where a permutation group acts on the particles. In the first non-trivial case, $n = d + 1$, we give a homogeneous system of parameters that is valid for the action of an arbitrary permutation symmetry and make a conjecture for the full Hironaka decomposition in the case without permutation symmetry.

Armed with the knowledge of characterising invariant algebras, we then discuss why they are indeed sufficient for our purposes. More precisely, we discuss and prove some approximation theorems which allow us to use the generators of invariant algebras to approximate continuous invariant functions in machine learning algorithms in general and in neural networks specifically. Finally, we implement these variables in neural networks applied to regression tasks to test the efficiency improvements. We perform our testing for a variety of hyperparameter choices and find an overall reduction of the loss on training data and a significant reduction of the loss on validation data. For a different approach on quantifying the performance of these neural networks, we treat the problem from a Bayesian inference perspective and employ nested sampling techniques to perform model comparison. Beyond a certain network size, we find that networks utilising Hironaka decompositions perform the best.

Acknowledgements

First and foremost, I would like to thank Ben Gripaios. Much of this work would simply not be without his guidance and low tolerance for nonsense. His unique views on physics and his devotion to mathematical rigour have thoroughly shaped my way of thinking and will be with me for many years to come. He has also made me come to appreciate the pleasure of writing precisely but artfully and for that I am grateful.

I would also like to thank my examiners, Ben Allanach and Tobias Golling, for their valuable comments and suggestions. The viva voce was a thoroughly enjoyable and highly stimulating experience.

I am also grateful for Ben Allanach, Christopher G. Lester, and Scott Melville for their helpful advice and constant encouragement. Their kind words and friendly faces have never failed to mellow the occasional frustrations of work. I would also like to thank Will Handley whose advice and stimulating conversations have been with me ever since he supervised me for my master's thesis.

My friends at DAMTP and the Cavendish have made my time there incredibly enjoyable. I would like to especially thank Joseph Tooby-Smith for being a wonderful and intellectually stimulating PhD twin. I will cherish the memories of the long hours spent staring in confusion at the empty walls of his windowless Cavendish office. Additionally, I would like to thank Maeve Madigan, Daniel Zhang, Joe Davighi, Hasan Mahmood, Shayan Iranipour, Robin Croft, James Delaney, Josh Kirklin, João Melo, James Moore, Hannah Banks, Herschel Chawdhry, Andrei Popescu, Cameron Voisey and members of the Cambridge Pheno Working Group, especially Maria Ubiali, Alex Mitov, and Bryan Webber.

I owe very much to my close friends who have made my past seven years in Cambridge unforgettable. I am grateful to Jack for tolerating my constant music playing, Lisa for never letting me stray too far into capitalism (and for sending me excessively long voice notes), Mie for the many interesting book recommendations (which, admittedly, I've rarely read), Marci for introducing me to many new things but mostly for listening, Noah for the musical jam sessions and fiery pits, Katy for the never-ending supply of baked goods and elocution lessons, Am for the heated debates, exercise motivation, and film recommendations, Reef for being there even if for a call every fortnight, Chris, Fiona, Dino, Rebekka, and Daro.

Finally, I would like to thank my parents Issa and Orayb, to whom this thesis is dedicated, and my brother Karam. I would not be what I am today without their wisdom, constant encouragement, and polymorphic support.

Contents

List of Figures

List of Tables

Nomenclature

Other Symbols

$\hookrightarrow$	Inclusion
$\xrightarrow{\sim}$	Isomorphism
$\twoheadrightarrow$	Surjection

Acronyms / Abbreviations

FFT	First Fundamental Theorem
HD	Hironaka decomposition
MAG	Minimal algebra generators
MLE	Maximum likelihood estimation
SFT	Second Fundamental Theorem
SW	Stone-Weierstrass

1

Introduction

In the era of big data, solutions to most quantitative problems only seem to be a machine learning algorithm away. Much work has been done to develop and extend machine learning tools to cope with a diverse array of applications.[1] Applying these algorithms to data has allowed their users, with varying degrees of success, to model the underlying systems of study and extract interesting, often valuable, information previously unknown to us mere humans.

But prior to this data revolution, when one had to think analytically about the systems one was studying, generations of scientists uncovered a great deal of knowledge about physical systems and their properties. To simply ignore their work in our modern data-driven quests for knowledge seems somewhat ungrateful, and possibly wasteful. Take, for example, the setting of particle physics, where we not only know that collision processes must be Lorentz-invariant, but also that they can often be permutation-invariant since the particles involved can be identical and thus indistinguishable. For examples in slightly different settings, consider physical chemistry or material science, where the properties of molecules and lattices are rotation- and translation-invariant and also potentially permutation-invariant when they contain identical atoms. These seemingly simple facts turn out to have important consequences for our attempts to model the world through data.

In this thesis, we demonstrate the benefit of incorporating our knowledge of the symmetries of certain systems into the machine learning methods we use to describe them. We do this by borrowing tools from a branch of mathematics called invariant theory, which allow us to construct systematic methods to produce sets of variables that describe our systems whilst incorporating our knowledge of their symmetries. These

[1] The interested reader may search for "machine learning" on-line and will instantly be bombarded with more than they can ever read about the subject.

methods result in generators which come in two flavours; *minimal algebra generators* and *Hironaka decompositions*. Crucially, the description via these variables loses no information about the systems which ensures that we do not miss out on any possibly interesting information in our data. Furthermore, Hironaka decompositions have the added advantage that they are redundancy-free, a feature which is highly desirable in any machine learning endeavour.

1.1 Invariant systems

Let us begin by defining the problem more precisely. In data-driven approaches, one is often tasked with describing a system which one has collected some data about. This data set is usually a collection of labelled pairs $D = \{(x_i, y_i)\}$, where the $x_i \in X$ are a set of input variables which belong to some space X and the $y_i \in Y$ are some known target variables which belong to some other space Y.[2] It is then assumed that a property of the system is described by some "true" function $f : X \to Y : x_i \mapsto f(x_i) = y_i$ and the typical task of machine learning is to find a function $\hat{f} : X \to Y$ which approximates f in some sense.[3]

If we have no more information about the system at hand, then this completes our set-up and from here, one usually proceeds to pick an algorithm from the vast machine learning toolbox to apply it to the data. But, when working with systems that possess certain symmetries, we have additional *a priori* knowledge and there is more to be said about the function f. Consider a group G acting regularly on the space X, $G \times X \to X : (g, x) \mapsto g \circ x$ for $x \in X$ and $g \in G$. If a system is invariant under the action of G then the functions f describing its properties must also be G-invariant, $f(g \circ x) = f(x)$ for all $g \in G$. When attempting to use machine learning tools to approximate these properties, this G-invariance imposes a huge restriction on the possible candidates for f (and hence on its approximators) and it would be extremely wasteful to ignore it. If we deprive our algorithms of this knowledge of invariance, a lot of time, effort, and (often scarce) data will be wasted rediscovering what we already know rather than discovering new insights.

[2] In particle physics, X usually corresponds to the vector space of particle momenta and Y corresponds to a quantity of interest such as a cross-section.

[3] We restrict our discussion to the case of supervised learning for clarity, but these ideas are also applicable to the unsupervised case, where one lacks the knowledge of the property one is seeking. There, although we do not know what property the function f will describe, we still know that it must be invariant under the symmetries of the system.

One way to incorporate group invariance into our description relies on modifying the data input. This modification can be carried out in a couple of ways. The first, more brute-force, method is to create additional artificial data by acting with the symmetry group on the original data. The algorithm is then applied to the combined data set in the hope that it learns that the functions it is approximating are constant on group orbits. Apart from requiring additional, often significant, computational effort, this method has the added disadvantage of being impossible to carry out completely when the symmetry group is continuous. Another method relies on reformatting the input data. Instead of inputting the collected data, which in general is not invariant, directly into our models, we instead construct a new set of input variables which are invariant.[4] But, if this is done in any old arbitrary way (such as taking the `min`(), `max`(), or `sum`() of some combination of the inputs), then, even if the algorithm is non-linear and can deal with possibly missing information (which is not necessarily guaranteed), it will still waste resources constructing the missing information from the arbitrary invariant inputs. So crucially, one should reformat the input data without incurring any information loss. This requires finding a set of invariant variables which are able to construct all possible group-invariant functions that could describe the properties of our system.

Alas, the space of all invariant functions is too large to consider and we do not have a mathematical handle on it. That is, we do not know how to generate it. We thus restrict our attention to invariant polynomials which, when exploited via some approximation theorems, allow us to achieve the same goal. In particular, we highlight and prove some approximation theorems in Chapter 5 which show that invariant polynomials are indeed enough in the case of continuous functions invariant under compact groups or linearly reductive groups defined over the complex numbers. We also discuss the difficulties that arise when the reductive groups are defined over the reals.

Importantly however, the characterisation of invariant polynomials, unlike that of all invariant functions, is more tractable and is made possible by the tools of invariant theory.

[4] More precisely, we project X onto the orbit space X/G by $\pi : X \twoheadrightarrow X/G$ and define $\hat{f} : X/G \to Y$. The approximator function then becomes $\hat{f} \circ \pi : X \to Y$.

1.2 Invariant polynomials in physical sciences

Given the momentum vectors v_i of n particles in d spacetime dimensions, an old theorem of Weyl [42] tells us that the Lorentz- and parity-invariant polynomials are generated by the dot products $v_i \cdot v_j$. This theorem (or rather its obvious generalization from polynomials to the field of rational functions, the ring of formal power series, and thence to the whole gamut of functions typically considered in physics) has become so ubiquitous that it is, by and large, taken for granted nowadays.

Weyl's important result is actually composed of a pair of theorems, the precise statement of which goes as follows. By allowing the momenta to take values in the complex numbers rather than the reals, we can replace the action of the Lorentz group including parity transformations on the momenta with the action of the orthogonal group $O(d, \mathbb{C})$. Weyl's *first fundamental theorem* (FFT) [42] states that the algebra of polynomials in the v_i's invariant under $O(d, \mathbb{C})$ is generated by the $n(n+1)/2$ dot products $v_i \cdot v_j$.[5] Weyl's *second fundamental theorem* (SFT) characterises the relations between the generators: when $n \leq d$ there are no relations (so the dot products are algebraically independent and the algebra of invariants is a polynomial algebra), while when $n > d$, the relations are generated by the $(d+1)$-minors[6] of the $n \times n$ matrix whose entries are given by $v_i \cdot v_j$.

As useful as these results are in their current form, they are perhaps in need of a makeover given relatively recent developments in the area of commutative algebra and what we know about quantum field theory, namely that the particles that correspond to excitations of a single quantum field are indistinguishable. To do so, we make use of invariant theory methods and update these results via two approaches. First, we consider a system of n particles of which some subsets are identical (e.g. in a process in which two protons at the Large Hadron Collider collide to produce three jets). In this case, it is apposite to consider not just arbitrary Lorentz-invariant polynomials, but rather to restrict to those that are, in addition, invariant under the group of permutations of the identical particles (e.g. $S_2 \times S_3$ in our $pp \to jjj$ example). We attack the problem in a manner similar to Weyl's and provide a method for constructing minimal algebra generators of Lorentz- and permutation-invariant polynomials in the momenta, *i.e.* a generalisation of the FFT to include permutation groups. We do not however generalise the SFT (for technical reasons discussed later). Second, we make use of commutative algebra and invariant theory methods not available at the

[5] Without parity transformations, the group becomes $SO(d, \mathbb{C})$, and we have additional generators given by the possible contractions of the d-dimensional epsilon tensor with the momenta.

[6] We define a $(d+1)$-minor of a matrix to be the determinant of a $(d+1) \times (d+1)$ submatrix.

time of Weyl and provide an alternative, redundancy-free, description of Lorentz- and permutation-invariant polynomials via what is called a *Hironaka decomposition* where the invariant polynomials are uniquely expressed in terms of the generators of the decomposition. This has the added advantage of being easily generalised to cases when parity is not a symmetry.

To give a first explicit example of why this might be helpful in phenomenological analyses, it is useful to consider the situation in which the analysis is carried out, as is increasingly the case, by a supremely *un*intelligent being, namely via machine learning. There, experience has shown that, rather than let the machine learn about Lorentz invariance for itself, it is far more efficient to feed event data to the machine in a Lorentz-invariant form[7]. There is no reason to expect that permutation invariance should be any different. Symmetrising in this way has the related benefit of preventing the machine chasing wild geese, in the sense of looking for spurious Lorentz- or permutation-violating signals.[8] Similarly, a description via Hironaka decompositions, where invariant polynomials are uniquely expressed in terms of the decomposition, is also desirable to ensure that the algorithm wastes no time recognising redundancies in the input data. Symmetrising may even be an astute tactic in situations where the particles in question are known to be *not* identical, but where one wishes to deliberately blind oneself to the difference between them, because the associated physics is not under control. A good (though politically incorrect) example from the LHC might be a Swiss proton and a French proton (or rather beams thereof), where one can be fairly sure that there are observable differences between them, but one can be equally sure that such differences are not due to fundamental new physics, but have a rather more mundane, to wit intermural, origin.

It turns out that the results we obtain for the Lorentz- and permutation-invariant polynomials can, with some additional work, be readily extended to the Euclidean- and permutation-invariant polynomials. These polynomials are useful in cases when one not only knows that the system of study is rotation-invariant, but also translation-invariant. For example, in more recent attempts to model the world through data, chemists have become increasingly interested in teaching computers how to learn and predict the properties of molecules (solubility, polarity, etc.) for the purposes of drug discovery. There, the information one has about the molecules (apart from their charges and

[7]Indeed, as far as we are aware, no computer has yet discovered Lorentz invariance by itself. But, given an arbitrary symmetric metric, a neural network can be trained to converge on the Minkowski metric [8].

[8]Of course, this "benefit" will be considered a disbenefit by readers who are interested in the possibility that Lorentz invariance is violated, or that, say, 2 protons are not identical; we tactfully suggest that it would be better for all concerned if they were not to read any further.

masses) is the positions of their atoms, *i.e.* vectors, and any property of interest should be expressible in terms of the Euclidean-invariant polynomials in these vectors. When there are identical atoms in a molecule, one should similarly restrict further to the Euclidean- and permutation-invariant polynomials.

We also hope that these new descriptions using invariant polynomials will also be of use in analyses carried out by more intelligent beings. To give just one example, a common method for computing multi-loop amplitudes in quantum field theory is to first relate them using integration-by-parts identities [40, 28]. These are linear equations whose coefficients may be written as Lorentz- and permutation-invariant polynomials in the momenta of external particles. Thus, in setting up and carrying out such calculations, it would presumably be useful to know a (redundancy-free) set of generators of such polynomials.[9]

Similar ideas were explored in [23], in the context of classifying higher-dimensional operators in effective scalar field theories. A significant difference there is that one studies the action of permutations on quotient algebras with respect to an ideal which features the relation $\sum_i p_i = 0$ (corresponding to an integration-by-parts identity) in addition to the relations $p_i^2 = 0$ studied here (corresponding there to the leading order equations of motion). These additional relations make it difficult to compare our results directly with those in [23], though we hope that some of the results obtained here could nevertheless be usefully applied to the study of that problem. For a rather different approach, see [27], which studies permutation invariance directly at the level of quantum field theory amplitudes.

1.3 Layout and non-techincal statement of results

The layout of the thesis will be as follows. We begin with an overview of the mathematical background in Chapter 2. We review the necessary concepts of commutative algebra (rings, algebras, maps, etc.) and introduce the required results and theorems from invariant theory. In Chapter 3, we employ the tools of commutative algebra to formally understand the invariant algebras of interest. We first introduce Weyl's FFT, which is the statement that every Lorentz-invariant polynomial can be obtained by taking an arbitrary polynomial in variables y_{ij} (where $i, j \in \{1, \dots, n\}$ and $i \leq j$), and replacing $y_{ij} \mapsto v_i \cdot v_j$. From there, our first result is to extend Weyl's theorem to the polynomial invariants of the Euclidean group where we show that any Euclidean-

[9]Such permutation invariant polynomials may also be of use in analysing correlation functions in cosmology, but we will not consider this possibility further here. See [34] for more details.

invariant polynomial can be obtained by again taking an arbitrary polynomial in variables y_{ij} and replacing $y_{ij} \mapsto (v_i - v_{avg}) \cdot (v_j - v_{avg})$, where $v_{avg} = \frac{1}{n} \sum_i v_i$. We also show that the Euclidean-invariant algebra is isomorphic to the Lorentz-invariant algebra in one vector fewer, an essential result for obtaining Hironaka decompositions later on. Our second result, which follows almost immediately from Weyl's, is that every Lorentz- and permutation-invariant polynomial can be obtained by taking a permutation-invariant polynomial in y_{ij} (where the permutation group P acts on the indices i, j in the obvious way) and making the same replacement. Not only does this result seem completely trivial (though the proof will show it to be not quite so), but also it is completely useless as it stands, because of the difficulty of describing the permutation invariant polynomials in y_{ij}. Indeed, while permutations act in the natural way on the subset $\{y_{ii}\}$ and lead to a simple description of the invariant polynomials (going back, in the case $P = S_n$, to Gauss [16]), the action of permutations on $\{y_{ij} | i < j\}$ is non-standard and a description of the invariants (for the case $P = S_n$) is unknown for $n \geq 5$ [11]! Fortunately, such high multiplicities of identical particles are relatively rare in applications. To give ourselves a better chance of characterising these invariant polynomials, we therefore describe a more *ad hoc* method: the observables $v_i \cdot v_i$ for a particle are somewhat redundant, since they return the mass of the particle (for a jet, we assume that all jet masses are negligible, since to do otherwise would invalidate the assumption that jets are identical). As such, we are less interested in invariant polynomials involving $v_i \cdot v_i$. Unfortunately, one cannot simply throw them away, because when $n > d$ there are relations between $v_i \cdot v_j$ which mix pairs with $i = j$ and $i \neq j$ (with $n = 2$ and $d = 1$, for example, we have that $(v_1 \cdot v_1)(v_2 \cdot v_2) = (v_1 \cdot v_2)^2$). Our last result in this Chapter is therefore to construct a rigorous procedure (which is essentially forming a quotient with respect to the ideal generated by the polynomials $v_i \cdot v_i - m_i^2$, or rather the permutation invariant combinations thereof) to "get rid" of the $v_i \cdot v_i$.

Once we have the required knowledge of the invariant algebras and relations between them, we move onto finding sets of generators in Chapter 4. In the first approach, we use our direct generalisation of Weyl's result to describe and carry out a strategy for finding a set of generators of the permutation invariant polynomials in y_{ij} for specific cases of n and P, with at most 4 identical particles (such as for the $pp \to jjj$ example). The strategy uses well-known methods in invariant theory, relying crucially on the somewhat arcane Cohen-Macaulay property.[10] The list of generators obtained in this way is

[10]For readers who are not *au courant*, it is perhaps consoling to note that even Macaulay himself professed to being ignorant of this property.

somewhat lengthy in practice and so we turn to ways of shortening it. Again, there are standard ways in invariant theory of doing so, which we describe. As we will see, these results eventually lead to a manageable set of generators describing the Lorentz- or (Euclidean-) and permutation-invariant polynomials. In the example of $pp \to jjj$, we end up with a set of 26 generators, given explicitly in Table 4.2. In fact, this set of generators is minimal in number, so one can do no better. In the second approach, we use the fact that the Lorentz- (or Euclidean-) and permutation-invariant algebras are themselves Cohen-Macaulay and therefore admit Hironaka decompositions. That is, one may find a set of generators termed a *Homogeneous system of parameters* (HSOP), $\{\theta_i\}$, and *secondaries*, $\{\eta_j\}$, such that any invariant polynomial can be expressed *uniquely* as a linear sum of the $\{\eta_j\}$ with coefficients which are polynomials in the $\{\theta_i\}$, $\sum_j \eta_j h_j(\theta_i)$. We solve the hardest step, namely to find HSOPs for the Lorentz- and permutation-invariant algebras and show, using the results of Chapter 3, that they can be used as HSOPs for the Euclidean- and permutation-invariant algebras. From here, finding the secondaries reduces to an unwieldy algorithm, of which we give an explicit computation. Finally, we give examples of Hironaka decompositions of the case of $(n, d) = (5, 4)$ with no permutation symmetry and $(n, d) = (3, 2)$ with full permutation symmetry.

Armed with the knowledge of characterising invariant algebras, we show in Chapter 5 that our restriction from all invariant functions to the invariant polynomials is not all that restrictive. First, in the general setting of any machine learning algorithm, we discuss and prove some approximation theorems which show that the invariant polynomials can approximate continuous functions invariant under a special class of groups, linearly reductive groups over $\mathbb{C}$. We then focus our attention on neural networks. We extend a theorem that shows that neural networks using the minimal algebra generators as inputs can approximate any continuous invariant functions. But as mentioned previously, the minimal algebra generators contain redundancies which may reduce a network's efficiency. We therefore construct Hironaka decomposition based networks, which are redundancy-free, and prove that they too can approximate continuous invariant functions.

Finally, to breathe some life into all of this mathematics, we implement our generators of invariant algebras in neural networks in Chapter 6.[11] We perform a comprehensive comparison of neural networks that utilise the minimal algebra generators and Hironaka decompositions vs. ones that use the Weyl generators (dot

[11]Similar work has been carried out in [5] but there, the authors rely on the finite-dimensional representations of the Lorentz group to construct Lorentz-equivariant networks and they deal with permutation invariance by imposing sums at the appropriate parts of the networks.

products and epsilon tensors). We find an overall reduction of the loss on training data and a significant reduction of the loss on validation data for a wide range of hyperparameter choices. For a different approach on quantifying the performance of these neural networks, we treat the problem from a Bayesian inference perspective and employ nested sampling techniques to compute the model evidence which we then use to perform model comparisons. Beyond a certain network size, we find that networks utilising Hironaka decompositions outperform those using minimal algebra generators which in turn outperform those using the Weyl generators.

Finally, we summarise and discuss our results in Chapter 7. We also discuss some open problems and avenues of future work on both the mathematical and computational fronts.

2

Commutative algebra and invariant theory

We begin our journey by giving an overview of the required mathematical background. Although commutative algebra forms an integral part of any undergraduate mathematics curriculum, physicists are rarely formally introduced to the subject. In this Chapter, we provide a largely self-contained introduction to the most relevant concepts and objects considered in this thesis. We also provide an introduction to the even lesser known subject of invariant theory, on which we rely heavily in our later discussions. The material and results in this Chapter can be found in [26, 3, 35, 6, 11, 29, 10].

2.1 Commutative algebra

"Algebra is the offer made by the devil to the mathematician."
– Michael Atiyah, *on the dichotomy between algebra and geometry.*

2.1.1 Objects

All the objects we will encounter start their lives as lowly sets. By progressively endowing these sets with structure, we build up the main characters that will populate our discussions. To begin, we endow a set with one operation.

Definition 2.1.1 (Monoid)**.** *A monoid M is a set endowed with a single operation $\circ : M \times M \to M$ and a special element $e \in M$ such that we have*

$$\textit{(Associativity):} \quad m_1 \circ (m_2 \circ m_3) = (m_1 \circ m_2) \circ m_3, \quad \forall m_1, m_2, m_3 \in M,$$

(Identity): $m \circ e = e \circ m = m, \quad \forall m \in M.$

An example of a monoid is the set of positive integers $\mathbb{N} = \{1, 2, \dots\}$ under multiplication (with identity 1). By imposing an extra criterion on a monoid, we arrive at an object which needs no introduction for an audience of physicists.

Definition 2.1.2 (Group)**.** *A group G is a monoid in which each element $g \in G$ has a (unique) inverse $g^{-1} \in G$ such that $g \circ g^{-1} = g^{-1} \circ g = e$.*

An *Abelian* group is one where the operation $\circ$ is commutative, $g_1 \circ g_2 = g_2 \circ g_1, \forall g_1, g_2 \in G$. The set of integers $\mathbb{Z}$ endowed with addition is an example of an Abelian group (with identity 0). If we now allow ourselves a second operation, we can define a ring, one of the protagonists of this thesis.

Definition 2.1.3 (Ring)**.** *A ring R is an Abelian group, with respect to an operation $+$ (addition) and a corresponding identity* 0*, that is also a monoid, with respect to an operation $\cdot$ (multiplication), and a corresponding identity* 1*, such that $\cdot$ is distributive over $+$*

$$
\begin{aligned}
r_1 \cdot (r_2 + r_3) &= r_1 \cdot r_2 + r_1 \cdot r_3, \\
(r_2 + r_3) \cdot r_1 &= r_2 \cdot r_1 + r_3 \cdot r_1,
\end{aligned}
$$

for all $r_1, r_2, r_3 \in R$.

We often omit the $\cdot$ when the meaning is unambiguous. The "commutative" in the title of this Chapter indicates that we will take the multiplication operation to be commutative and henceforth restrict our attention to commutative rings.[1] We also impose that $1 \neq 0$.[2] An example of a commutative ring is the set of integers $\mathbb{Z}$ but now endowed not only with addition, but also multiplication.

In addition to the notion of a subring, which (unsurprisingly) is a subset of a ring that is itself a ring, a special subset that we will come across later on is called an ideal.

Definition 2.1.4 (Ideal)**.** *A subset $I \subset R$ is called an ideal if 1) $0 \in I$, 2) for any $a, b \in I$ we also have that $a + b \in I$, and 3) for any $a \in I$ and $r \in R$ we have that $ra \in I$.*

An example of an ideal of $\mathbb{Z}$ is the subset of all multiples of 2 (or of 3 or any other integer). We will also come across intersection ideals.

[1] The subject of non-commutative algebra is very rich and contains many intricacies that we will (thankfully) not encounter.

[2] If $1 = 0$, then the ring only has one element and is called the zero ring, $\{0\}$.

Definition 2.1.5 (Intersection ideal). *Consider a ring R and a subring $S \subset R$. If I is an ideal of R, then $J = I \cap S = \{a \mid a \in I, a \in S\}$ is an ideal of S.*

Given the ring of integers $\mathbb{Z}$, the subring of multiples of 3, and the ideal of multiples of 2 in $\mathbb{Z}$, the multiples of 2 and 3 form an example of an intersection ideal of the subring of multiples of 3.

The next piece of structure we allow ourselves is *scalar multiplication.*

Definition 2.1.6 (R-module). *Given a ring R, an R-module M is a an Abelian group, with respect to an operation $+$ (addition), with scalar multiplication $R \times M \to M : (r, m) \mapsto rm$ which is distributive over $+$ (in both R and M), associative, and is such that $(1, m) \mapsto 1m = m$.*

One example is an ideal $I \subset R$ (which is an R-module). Another example, which is close to any physicist's heart, is a real vector space V, which is an $\mathbb{R}$-module (also called an $\mathbb{R}$-vector space).

The final fundamental object we consider, usually the first object students of mathematics are introduced to as children, can be obtained by allowing ourselves the last obvious remaining property.

Definition 2.1.7 (Field). *A field K is a commutative ring in which every non-zero element has a multiplicative inverse and $1 \neq 0$.*

Examples of fields include the rational numbers $\mathbb{Q}$, the real numbers $\mathbb{R}$, and the complex numbers $\mathbb{C}$ (but not the integers $\mathbb{Z}$). The characteristic of a field is the smallest number n such that $\underbrace{1 + \cdots + 1}_{n \text{ summands}} = 0$. All the fields we will consider are fields of characteristic 0.[3]

2.1.2 Maps

As of now, all of the objects we have introduced are standalone. In the following Chapters, we will rely heavily on forming relations between these objects to reach our required results. To so, we introduce multiple notions of maps.

Definition 2.1.8 (Group map). *A group map $\phi : G \to H$ is a map that preserves the group operation, $\circ$, and the identity, e. We call ϕ an isomorphism if it is bijective and write $G \xrightarrow{\sim} H$.*

[3]The subject of algebra over finite fields (non-zero characteristic) is also a very rich one which we will (thankfully) not encounter.

The *kernel* $\ker\phi \subset G$ of a group map is the *preimage* of $e \in H$, $\phi^{-1}(e)$ (*i.e.* the set of all elements in G that map to e). The kernel of a group map is a normal subgroup of G.[4]

Definition 2.1.9 (Ring map)**.** *A ring map $\phi : R \to S$ (which we sometimes write less explicitly as $R \to S$) is a map that preserves the ring operations $(+, \cdot)$, and the multiplicative identity* 1*. We call ϕ an isomorphism if it is bijective and write $R \xrightarrow{\sim} S$.*

The *kernel* $\ker\phi \subset R$ of a ring map $\phi : R \to S$ is the *preimage* of 0, $\phi^{-1}(0)$ (*i.e.* the set of all elements in R that map to 0). As can easily be checked, the kernel of a ring map is an ideal of R. If we denote the *image* of a ring map by, $\operatorname{im}\phi = \phi(R) \subset S$, then by the first isomorphism theorem we have that $R/\ker\phi \xrightarrow{\sim} \operatorname{im}\phi$. If ϕ is surjective, then $\operatorname{im}\phi = S$ and we write $\phi : R \twoheadrightarrow S$.

Using the definition of a ring map, we define an algebra.

Definition 2.1.10 (R-algebra)**.** *Given a commutative ring R, an R-algebra (or algebra over R) is a ring S equipped with a ring map $\alpha : R \to S$ called the structure map.*

A subset $I \subset S$ of an R-algebra S is called an ideal if it is an ideal of S as a ring.

Before we define the final type of map that we will encounter, we must introduce the notion of a commutative diagram.

Definition 2.1.11 (Commutative diagram)**.** *A commutative diagram $\mathcal{C}$ is a collection of maps $A_i \xrightarrow{\phi_i} B_i$ in which all map compositions starting from the same set A and ending with the same set B agree.*

For example, if the following diagram

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[d, "g"] & B \arrow[d, "p"] \\ C \arrow[r, "h"] & D \end{tikzcd}$$

commutes, then $p \circ f = h \circ g$. We now define an algebra map.

Definition 2.1.12 (R-algebra map)**.** *Given two R-algebras S and T with structure maps α_S and α_T respectively, an R-algebra map is a ring map $\phi : S \to T$ such that $\phi \circ \alpha_S = \alpha_T$,* i.e. *the following diagram*

$$\begin{tikzcd} & S \arrow[dr, "\phi"] & \\ R \arrow[ur, "\alpha_S"] \arrow[rr, "\alpha_T"] & & T \end{tikzcd}$$

[4] A subgroup $N \subset G$ is called normal if and only if $gng^{-1} \in N$ for all $n \in N$ and $g \in G$.

commutes.

Similar to a ring map, we have that the $\ker\phi$ of an R-algebra map is an ideal and, by the first isomorphism, $S/\ker\phi \xrightarrow{\sim} \operatorname{im}\phi$.

2.1.3 Grading and generators

We say that a ring R is *graded* if we can write it as a direct sum $R = \bigoplus_{n\in\mathbb{N}_0} R_n$ of subgroups R_n such that $R_n R_m \subset R_{n+m}$. A homogeneous element (of degree n) is an element belonging to some factor (or specifically to R_n).

The ideal $\langle r_\lambda | \lambda \in \Lambda\rangle$, generated by some $r_\lambda \in R$ for some set Λ, is the smallest ideal in R that contains the r_λ. An element in $\langle r_\lambda | \lambda \in \Lambda\rangle$ can always be written as $\sum_{\lambda\in\Lambda} h_\lambda r_\lambda$ for some $h_\lambda \in R$. We call an ideal I of a graded ring $R = \bigoplus_n R_n$ homogeneous if it can be written as $I = \bigoplus_n (I \cap R_n)$. An ideal is homogeneous if and only if it is generated by homogeneous elements.

Given an R-algebra S, the subalgebra $R[s_\lambda | \lambda \in \Lambda]$ generated by $s_\lambda \in S$ is the smallest R-subalgebra that contains them. If there exists $s_1, \ldots, s_n \in S$ such that $S = R[s_1, \ldots, s_n]$, we say that S is *finitely-generated* (as an R-algebra). An R-algebra S is graded if it is graded as a ring. In what follows we will always give R the trivial grading (where we take $R = \mathbb{C}$) where $S_i S_j \subset S_{i+j}$ and $RS_i \subseteq S_i$. A graded R-algebra map $\phi : A \to S$ is an R-algebra map which preserves the original grading of $A = \bigoplus_i A_i$ such that $S = \bigoplus_i S_i$, $S_i = \phi(A_i)$, and $\phi(A_i)\phi(A_j) = S_i S_j \subset \phi(A_{i+j}) = S_{i+j}$.

We say that a subset $\{m_\lambda | \lambda \in \Lambda\} \subset M$ generates M (as an R-module) if M is the smallest submodule of M that contains $\{m_\lambda\}$. A module M is said to be *finitely-generated* if there exists a finite set of generators. Finally, the $\{m_\lambda\}$ are called *free* if $\sum_{\lambda\in\Lambda} r_\lambda m_\lambda = 0$, $r_\lambda \in R$, implies that $r_\lambda = 0$ for all λ and a *basis* if they also generate M. A *free module* is one that has a basis.

2.1.4 Quotient algebras

Quotient algebras will also form an integral part of our discussions in the following Chapters. We start by defining a quotient.

Definition 2.1.13 (Quotient)**.** *Consider the R-algebra A and an ideal $I \subset A$. The quotient A/I is defined with the quotient (projection) map*

$$q : A \twoheadrightarrow A/I, \ a \mapsto a + I, \quad a \in A.$$

In fact, this quotient is actually an algebra.

Lemma 2.1.1 (Quotient algebra)**.** *A/I is an R-algebra.*

Proof. We first check that the operations in the quotient are well-defined:

- Addition: I is a subgroup of A under addition and furthermore it is normal (A is Abelian under addition so any subgroup is a normal one). This means that A/I is a well-defined quotient group under addition by the quotient group theorem.
- Multiplication: For elements $x_1 + I = x_2 + I$ and $y_1 + I = y_2 + I$ in the quotient, we have that $x_1 - x_2, y_1 - y_2 \in I$. We can construct $y_1(x_1 - x_2) + x_2(y_1 - y_2) \in I$, which implies $x_1y_1 - x_2y_2 \in I$ and therefore $x_1y_1 + I = x_2y_2 + I$. Hence multiplication is also well-defined.

Next, we need to show that the quotient satisfies the ring axioms:

- It is a group under addition: There is an identity $(0 + I)$, there is an inverse for any element $x + I$ (namely $-x + I$), it is closed $(x_1 + I + x_2 + I = x_1 + x_2 + I)$, it is associative, and it is commutative.
- It is a monoid under multiplication: There is an identity $(1 + I)$, it is closed $((x_1 + I)(x_2 + I) = x_1x_2 + I)$, and it is associative.
- Multiplication is distributive over addition: $(x_1 + I)(x_2 + I + x_3 + I) = x_1x_2 + x_1x_3 + I$.

Finally to show it is an R-algebra, let α_A be the structure map of A, $\alpha_A : R \to A$. Then the structure map of A/I is $\alpha = q \circ \alpha_A : R \to A/I$. □

We can also show that this quotient algebra inherits the grading structure from its parent algebra if the ideal I is homogeneous.

Lemma 2.1.2 (Graded quotient algebra)**.** *For a graded algebra $A = \bigoplus_i A_i$ with a homogeneous ideal $I = \bigoplus_i (I \cap A_i)$, the grading structure is preserved in the quotient algebra $A/I = \bigoplus_i (A_i + I)/I$.*

Proof. It is easy to check that $A/I = \bigoplus_i A_i / \bigoplus_i (A_i \cap I) = \bigoplus_i A_i/(A_i \cap I)$ (*i.e.* quotients and direct sums "commute"). But using the second isomorphism theorem for rings, we have that $A_i/(I \cap A_i) \cong (A_i + I)/I$, and therefore $A/I \cong \bigoplus_i (A_i + I)/I$. □

Lastly, we prove a statement about induced algebra maps between quotient algebras.

Lemma 2.1.3 (Quotient algebra map)**.** *Consider two graded R-algebras A and S with ideals I_A and I_S respectively. If we have a graded surjective R-algebra map $\phi : A \twoheadrightarrow S$, with $\phi(I_A) \subset I_S$, then this induces another surjective map of R-algebras*

$$\phi^* : A/I_A \twoheadrightarrow S/I_S : a + I_A \mapsto s + I_S$$

Furthermore, it is graded if A/I_A and S/I_S are graded.

Proof. First, we note that for any $a + I_A \in A/I_A$, we have that $\phi(a + I_A) \in S/I_S$ since $\phi(I_A) \subset I_S$. We therefore have a well-defined restriction map $\phi^* : A/I_A \to S/I_S$. To show that it is surjective, take an element $s + I_S \in S/I_S$. Since ϕ is surjective, one can always find an $a \in A$ such that $\phi(a + I_A) = s + I_S$. Finally, if the quotient algebras are graded, $A/I_A = \bigoplus_i (A_i + I_A)/I_A$ and $S/I_S = \bigoplus_i (S_i + I_S)/I_S$ then ϕ^* preserves the grading as follows: $\phi^*((A_i + I_A)/I_A)\phi^*((A_j + I_A)/I_A) = ((S_i + I_S)/I_S)((S_j + I_S)/I_S) = \phi^*((A_{i+j} + I_A)/I_A) = (S_{i+j} + I_S)/I_S$. □

2.1.5 Polynomial algebras

The special case of algebras we will be primarily concerned with are called polynomial algebras. At the risk of stating the obvious, we first define monomials and then polynomials.

Definition 2.1.14 (Monomial). *A monomial in n indeterminates, $x_1, \ldots, x_n$, is a product of the form*

$$x_1^{\alpha_1} \cdot x_2^{\alpha_2} \cdot \cdots \cdot x_n^{\alpha_n}$$

(often simplified to just x^α) where all exponents α_i are non-negative integers. The total degree of a monomial, or deg(x^α), is the sum of its exponents, $\sum_i \alpha_i$.

Definition 2.1.15 (Polynomial). *A polynomial f in n indeterminates, $x_1, \ldots, x_n$, with coefficients in a field K is a finite linear sum of monomials*

$$f = \sum_\alpha a_\alpha x^\alpha, \ a_\alpha \in K.$$

The set of all polynomials forms a K-algebra which we denote by $K[x_1, \ldots, x_n]$ (the structure map of this algebra is $k \mapsto kx^0$, $k \in K$). We focus on the special case when these indeterminates are the coordinates of an n-dimensional K-vector space V in some basis, and we denote the K-algebra by $K[V]$ for short. In this thesis we always take $K = \mathbb{C}$. Trivially, this polynomial algebra is generated by the $x_1, \ldots, x_n$.

The last topic we introduce in our commutative algebra crash course, is the notoriously difficult ideal membership problem. That is, given a polynomial $f \in K[V]$ and an ideal $I = \langle f_1, \ldots, f_r \rangle \subset K[V]$ generated by $f_1, \ldots, f_r$, is $f \in I$? For uni-variate polynomials, this can easily be answered using the polynomial long division algorithm (which has been known about since medieval Arabic times). By checking whether f is divisible by any of the $f_1, \dots, f_r$, one can immediately establish whether $f \in I$. But, in the case of multi-variate polynomials, the situation is not as simple. Long division requires more care to be defined and even then, still poses problems. To answer the problem there, we begin by defining an ordering on the monomials.

Definition 2.1.16 (Monomial ordering)**.** *A monomial ordering is a total order "$>$" on the set of all monomials M such that for any monomials t, t_1, and t_2*

$$\begin{aligned} &t > 1 \textit{ for all } t \in M \setminus \{1\}, \\ &t_1 > t_2 \textit{ implies } t \cdot t_1 > t \cdot t_2. \end{aligned}$$

We denote the *leading monomial* of a polynomial f (with respect to an ordering "$>$") by $\mathrm{LM}(f)$. An example of a monomial order which we will use later on is *graded reverse lexicographic* ordering (or grevlex for short).

Definition 2.1.17 (Grevlex)**.** *For any monomials $t = x_1^{\alpha_1} \dots x_n^{\alpha_n}$ and $t' = x_1^{\alpha'_1} \dots x_n^{\alpha'_n}$, we say that $t >_{grevlex} t'$ if $deg(t) > deg(t')$, or if $deg(t) = deg(t')$ and $\alpha_i < \alpha'_i$ for the largest i with $\alpha_i \neq \alpha'_i$.*

Using this monomial ordering, one can define long division for multi-variate polynomials. We will not give the explicit algorithm here (it proceeds largely as one would expect, by imposing a monomial ordering and dividing the polynomial in that order). We instead give an example to illustrate how it can often fail to answer the question of ideal membership. Consider the polynomial algebra $\mathbb{R}[x, y]$ and the polynomials $f = xy^2 - x$, $f_1 = xy + 1$, and $f_2 = y^2 - 1$. It is obvious that $f = xf_2$ and hence is in the ideal $\langle f_1, f_2 \rangle$. But, if we impose grevlex ordering $x >_{\text{grevlex}} y$ (and so $f >_{\text{grevlex}} f_1 >_{\text{grevlex}} f_2$) and naïvely perform the division, we get $f = xy^2 - x = yf_1 + 0f_2 + (-x - y)$ which indicates that f is not actually in $\langle f_1, f_2 \rangle$. To remedy this problem, Buchberger introduced the notion of Gröbner bases.[5]

Definition 2.1.18 (Gröbner basis)**.** *Let $S \subseteq K[x_1, \dots, x_n]$ be a set of polynomials. We write $L(S) = \langle LM(g) | g \in S \rangle$ for the ideal generated by the leading monomials from S. Let $I \subseteq K[x_1, \dots, x_n]$ be an ideal. Then, a finite set $\mathcal{G} \subseteq I$ is called a Gröbner basis of I (with respect to some chosen monomial ordering) if $L(I) = L(\mathcal{G})$.*

Gröbner bases have many interesting properties (they generate their ideals, etc.), but one which is relevant for our discussion is the following. Given a Gröbner basis $\mathcal{G}$ of an ideal $I \subset K[x_1, \dots, x_n]$ and some $f \in K[x_1, \dots, x_n]$, then $f \in I$ if and only

[5]Named after his Ph.D. advisor Wolfgang Gröbner.

if the remainder of f by $\mathcal{G}$ is zero which gives us a infallible method to test for ideal membership. Obtaining Gröbner bases can be done using Buchberger's algorithm[6] [7] but it is in general inefficient and computationally expensive. In later Chapters, this inefficiency will impose computational constraints on our attempts to obtain generators of invariant polynomials.

2.2 Invariant theory

"Das ist nicht mathematik. Das ist theologie."
– Paul Gordan, *in response to Hilbert's proof of the finiteness theorem.*

We now move onto introducing invariant theory. The results and methods developed here will provide us with just the right tools to attack our problems in Chapter 4. For the rest of this Section, K denotes an algebraically closed field of characteristic zero. This restriction is necessary to make use of power theorems from invariant theory.

One of the fundamental problems of invariant theory is, given some group G acting on a polynomial algebra in some well-defined way, to find a set of invariant polynomials that generate the invariant subalgebra. Unlike the case of plain polynomial algebras, generating invariant algebras is highly non-trivial and a necessary question to ask is: when is an invariant algebra even finitely-generated? This is the question Hilbert himself posed in his 14-th problem.[7] As is usually the case in mathematics and science, the road to a discovery is winding and the answer to this question was no exception. But, having the privileged perspective gained by existing in the future light cone of Hilbert and co., we will present a more structured discussion of its resolution.

We start by discussing an important class of groups, namely the linear algebraic groups. Colloquially, these are the subgroups of the general linear group.[8] We will be further interested in their representations.

Definition 2.2.1 (Representation). *A representation of a group G is a finite n-dimensional K-vector space V together with a group map $G \to GL(V,K)$, where $GL(V,K)$ is the general linear group group of $n \times n$ invertible matrices with entries in K.*

[6] Named after himself this time.
[7] It is actually only part of the original problem.
[8] Formally, they are affine varieties which have the structure of a group such that the group multiplications and inversions are morphisms of affine varieties.

A representation is called *rational* if G acts regularly on V.[9] It turns out that an important subclass of the linear algebraic groups for our purposes are the reductive groups.[10]

Definition 2.2.2 (Linearly reductive group)**.** *A linear algebraic group G is called* linearly reductive *if for every rational representation V and every $v \in V^G \setminus \{0\}$, there exists a linear invariant function $f \in (V^*)^G$ such that $f(v) \neq 0$. Here, $V^G \subset V$ is the trivially invariant subspace ($g \circ v = v, \forall v \in V^G$) and V^* are the linear maps from $V \to K$.*

We need not worry too much about what this definition precisely means and how one can use it to prove that a linear algebraic group is linearly reductive.[11] It is known that all finite groups and most groups one encounters in physics (GL, SL, and, importantly for our purposes, the orthogonal group) are examples of such reductive groups.

Now consider a group G acting on a finite-dimensional K-vector space V via some representation, $\rho : G \to GL(V, K)$, as $G \times V \to V : (g, v) \mapsto \rho(g)v$. This action can be extended to the K-algebra of polynomials, $K[V]$, in the obvious way as $f(v) \mapsto g \circ f(v) = f(\rho(g)v)$ for $f \in K[V]$. We denote the subalgebra of all invariant polynomials by $K[V]^G = \{f \in K[V] | f = g \circ f, \forall g \in G\}$.[12] Although $K[V]$ is a polynomial algebra, the invariant algebra $K[V]^G$ often has a complicated structure and is general not polynomial.

One of the important properties of algebras invariant under reductive groups G is that there exists a projection from the algebra $K[V]$ to the invariant algebra $K[V]^G$ called a Reynolds operator.

Theorem 2.2.1 (Reynolds operator)**.** *A Reynolds operator $\mathcal{R}_G$ is a G-invariant projection,* i.e. *a linear map, $\mathcal{R}_G : K[V] \twoheadrightarrow K[V]^G$ such that $\mathcal{R}_G(f) = f$ and $\mathcal{R}_G(fh) = f\mathcal{R}_G(h)$ for all $h \in K[V]$ and $f \in K[V]^G$. If G is a linearly reductive group, then there exists a Reynolds operator $\mathcal{R}_G : K[V] \twoheadrightarrow K[V]^G$.*

[9] We say that a linear algebraic group G acts regularly on a K-vector space V if an action is given by a map $G \times V \to V$.

[10] There are multiple notions of reductivity in the literature; linearly reductive, geometrically reductive, and group theoretically reductive. In fields of characteristic zero, all notions coincide.

[11] For a simple example, consider the algebra in two variables, $\mathbb{C}[x, y]$, and the additive group $G_{\text{add}} = \mathbb{C}$ with a regular action on $(x, y) \in \mathbb{C}^2$ defined by $\lambda \circ (x, y) \mapsto (x + \lambda y, y), \lambda \in \mathbb{C}$. It is not too difficult to see that the invariant subalgebra is simply $\mathbb{C}[y]$ and that the trivially invariant subspace is $V^G = \mathbb{C} \times \{0\}$. Now, for every linear invariant function $f \in (V^*)^{G_{\text{add}}}$ (which is just $\propto y$) and any element $v \in V^{G_{\text{add}}} \setminus \{0\}$, $f(v) = 0$. Hence, G_{add} is not linearly reductive.

[12] The algebra $K[V]$ carries a grading with $(K[V])_0 = K$ which is inherited by the invariant subalgebra $K[V]^G$.

For finite groups H, Reynolds operators are as simple as averaging over the group

$$\mathcal{R}_H(f) = \frac{1}{|H|}\sum_{h\in H} h \circ f.$$

For continuous reductive groups, the situation is more complicated. In the case of reductive groups G defined over $\mathbb{C}$ with a maximal compact subgroup C, one may invoke the concept of a normalised Haar measure on C, μ_C, and the Reynolds operator becomes

$$\mathcal{R}_G(f) = \int_{c\in C} c \circ f \;\; d\mu_C.$$

For a general reductive group, the projection by a Reynolds operator can be found algorithmically [11]. An important result that follows from the proof of the existence of a Reynolds operator is the following

Corollary 2.2.1. *Let A and B be two K-algebras with a regular action of a linearly reductive group G on them, where $\mathcal{R}_A : A \twoheadrightarrow A^G$ and $\mathcal{R}_B : B \twoheadrightarrow B^G$ are Reynolds operators. Then any G-equivariant (*i.e.* commutes with G) K-algebra map $\phi : A \to B$ commutes with the Reynolds operators, $\mathcal{R}_B \circ \phi = \phi \circ \mathcal{R}_A$.*

We will use this fact repeatedly in our proofs of the following Chapters. Now, the initial question we (or rather Hilbert) posed finds its answer in perhaps the most important theorem in invariant theory, namely, Hilbert's finiteness theorem.

Theorem 2.2.2 (Hilbert's finiteness theorem)**.** *If G is a linearly reductive group and V is a rational representation of G, then $K[V]^G$ is finitely-generated over K.*

The original (non-constructive) proof was given for the classical groups ($GL, SL, \ldots$), but in fact, the arguments used are valid whenever a Reynolds operator exists (*i.e.* all linearly reductive groups).

Another important result of invariant theory, due to Noether, is that any finitely-generated graded algebra R with $R_0 = K$ admits a (not necessarily unique) homogeneous system of parameters (HSOP).

Definition 2.2.3 (Homogeneous system of parameters)**.** *A set of homogeneous elements, $\{\theta_1, \ldots, \theta_r\}$, of a graded algebra $R = \bigoplus_i R_i$, $R_0 = K$, is called a homogeneous system of parameters (also called primaries) if they are algebraically independent and if R is a finitely-generated module over $K[\theta_1, \ldots, \theta_r]$.*

The first condition of algebraic independence, although simple to state, will prove to be the more difficult to tackle in our proofs in Chapter 4. To unravel the second condition, we state the following result.

Lemma 2.2.1. *We call the zero set of all polynomials of strictly positive degree, $N_V := \{v \in V | f(v) = 0,\ \forall f \in K[V]^G_+\}$, the nullcone. If $\{\theta_1, \dots, \theta_r\} \in K[V]^G$ are homogeneous and the zero set $\{v \in V | \theta_i(v) = 0, \forall i\}$ coincides with the nullcone, then $K[V]^G$ is a finitely-generated module over $K[\theta_1, \dots, \theta_r]$.*

Checking the satisfaction of this *nullcone condition* is usually relatively straightforward. By combining these previous results we get that, given a HSOP of some invariant algebra $K[V]^G$, we can express the algebra as a finitely generated module over the subalgebra generated by the HSOP, $K[\theta_1 \dots, \theta_l]$. In particular, we may write $K[V]^G = \sum_k \eta_k K[\theta_1 \dots, \theta_l]$, where we call the η_j *secondary invariants.*

We again arrive at another significant result, namely that if $K[V]^G$ is Cohen-Macaulay, then it is a *free* (and as we have already seen, finitely-generated) module over any HSOP. Thus, we in fact have a *Hironaka decomposition* $K[V]^G = \bigoplus_k \eta_k K[\theta_1 \dots, \theta_l]$ and we are able to use the full power of linear algebra. In particular, each element in $K[V]^G$ can be written *uniquely* as $\sum_j \eta_j f^j$, where $f^j \in K[\theta_1 \dots, \theta_l]$, and the product of any two secondaries is uniquely given by $\eta_k \eta_m = \sum_j \eta_j f^j_{km}$, where $f^j_{km} \in K[\theta_1 \dots, \theta_l]$. This specifies the multiplication in $K[V]^G$ unambiguously.

So the important question now is when is an invariant algebra Cohen-Macaulay? This answer to this is provided by the relatively recent and very powerful *Hochster-Roberts* theorem (1974) [24].

Theorem 2.2.3 (Hochster-Roberts)**.** *An invariant algebra $K[V]^G$ is Cohen-Macaulay if G is a linearly reductive group.*

Some simple examples will perhaps be illuminating. When G is the trivial group acting on a basis vector $x \in \mathbb{C}$, we may set $\eta_1 = 1$ and $\theta_1 = x$, such that $K[V]^G = 1 \cdot \mathbb{C}[x]$. But we may also set $\eta_1 = 1, \eta_2 = x$, and $\theta_1 = x^2$, such that $K[V]^G = 1 \cdot \mathbb{C}[x^2] \oplus x \cdot \mathbb{C}[x^2]$. This already shows that a Hironaka decomposition is not unique. For a slightly less trivial example, let G be the group $\mathbb{Z}_2$ whose non-trivial element sends basis vectors $x, y \in \mathbb{C}^2$ to minus themselves. Then we may set $\eta_1 = 1, \eta_2 = xy$ and $\theta_1 = x^2, \theta_2 = y^2$, to get $\mathbb{C}[x, y]^{\mathbb{Z}_2} = 1 \cdot \mathbb{C}[x^2, y^2] \oplus xy \cdot \mathbb{C}[x^2, y^2]$.

As with all good things, finding these Hironaka decompositions is in general hard. One usually proceeds to find the HSOP of some invariant algebra. Proving that a proposed set of polynomials constitutes a HSOP is the most difficult step. Once a HSOP is found, one then finds the secondaries algorithmically. To present the algorithm, we begin by defining a very useful tool of invariant theory, namely Hilbert series. For a finitely-generated graded K-algebra $R = \bigoplus_{i=0}^{\infty} R_i$, the Hilbert series is defined as

$$H(R,t) = \sum_{i=0}^{\infty} \dim(R_i)t^i,$$

where $\dim(R_i)$ is the dimension of the (homogeneous) vector space R_i. Hilbert series of algebras with Hironaka decompositions can be re-expressed in a more useful form as

$$H\left(\bigoplus \eta K[\theta], t\right) = \frac{1 + \sum_k S_k t^k}{\prod_l (1-t^l)^{P_l}}.$$

where S_k is the number of secondaries at degree k and P_l is the number of primaries at degree l. Once the HSOP is found, the number and degrees of the secondaries are fixed which allows us to easily read-off their number and degrees. In Appendix A, we discuss how to compute the Hilbert series of algebras of interest and provide explicit examples. But now, given a Hilbert series of an invariant algebra and a Gröbner basis $\mathcal{G}$ of the ideal generated by the HSOP, $\langle \theta_1, \ldots, \theta_l \rangle \subset K[V]^G$, one can compute the secondary invariants using an algorithm as follows:[13]

- Read off the degrees of the secondaries $d_1, \ldots d_m$ from the Hilbert series.
- For $i = 1, \ldots, m$ perform the following two steps:

  - Calculate a basis of the homogeneous component $K[V]^G_{d_i}$ (invariant polynomials of degree d_i).

  - Select an element η_i from this basis such that the normal form $\mathrm{NF}_{\mathcal{G}}(\eta_i)$ (remainder on division by the Groebner basis) is non-zero and is not in the K-vector space generated by the polynomials $\mathrm{NF}_{\mathcal{G}}(\eta_1), \ldots, \mathrm{NF}_{\mathcal{G}}(\eta_{i-1})$.[14]
- The invariants $\eta_1, \ldots, \eta_k$ are the required secondary invariants.

As mentioned previously, the computation of Gröbner bases is usually expensive and the algorithms involved can very quickly overpower our current computational capabilities. This, in addition to the often large number of secondaries required, imposes a bottleneck on our ability to compute Hironaka decompositions in general. A way to remedy this is to make away with the nice module structure of Hironaka decompositions which

[13] A version of this algorithm is implemented in `Macaulay2` [17] (and other computer packages).

[14] This step is essentially a test of ideal "non-membership", hence our previous discussion of the ideal membership problem.

allows us to obtain a set of algebra generators of much fewer number called a set of *minimal algebra generators*. We discuss this in more detail in Chapter 4.

3

Algebras and maps

Before we embark on our quest to obtain generators of invariant polynomials, it is necessary to take a step back and better understand the structure of their invariant algebras. Our starting point will be the famous result of Weyl which states that the polynomial vector invariants under the (special) orthogonal group are generated by the dot products (and the possible contractions with the epsilon tensor). This result is actually twofold. The *first fundamental theorem* (FFT) is the statement that there exists a surjective map from a polynomial algebra to the invariant algebra and the *second fundamental theorem* (SFT) is the characterisation of the kernel of this map. In this Chapter, we build on these two theorems to obtain three different main results.

The first is to extend Weyl's theorems to the polynomial invariants of the Euclidean group (rotations and translations). Although the Euclidean group is not linearly reductive (due to the translations) and Hilbert's finiteness theorem does not apply, the algebra of Euclidean-invariant polynomials is indeed finitely-generated. Surprisingly, it is also Cohen-Macaulay and so admits Hironaka decompositions. We show this by establishing that it is isomorphic to another Cohen-Macaulay algebra, namely the algebra of invariant polynomials under the orthogonal group in one vector fewer.

Our second result, which follows almost immediately from Weyl's FFT, is to restrict Weyl's map to a surjective map onto the algebra of polynomials invariant under the orthogonal and permutation groups. Although not particularly useful in this form (as it does not provide us with a list of generators), it is an essential result for the work that we will carry out in Chapter 4 to obtain sets of generators. We also do this for the Euclidean-invariant algebra. There, the action of the permutation group on the algebra requires more care.

Finally, as we will see in the following Chapter, finding sets of generators when permutations are included is where the real hard work begins. Our third result is therefore to quotient out the dot products corresponding to the invariant masses (which in particle physics do not vary from one collision event to the other) in the orthogonal group-invariant algebra. By doing so, we give ourselves a better fighting chance when finding the generators as the algebra which surjects onto the resulting quotient algebra is simpler to characterise.

3.1 The FFT and SFT of the orthogonal group

Consider a d-dimensional $\mathbb{C}$-vector space V.[1] The set of all rotations (proper and improper) acting on the elements of this vector space forms the famous orthogonal group.

Definition 3.1.1 (Definite orthogonal group). *The orthogonal group* $O(d, K) \subset GL(d, K)$ *(where* $K = \mathbb{R}$ *or* $\mathbb{C}$*) is the subgroup of orthogonal matrices* $\{R \mid RR^T = I_d, \forall R \in GL(d, K)\}$.[2] *The special orthogonal group,* $SO(d, K) \subset O(d, K)$*, is the subgroup of proper rotations,* i.e. *orthogonal matrices with unit determinant* $\{R \mid det(R) = 1, \forall R \in O(d, K)\}$.

Another equivalent definition of the definite orthogonal group found in the literature is that it is the group of transformations on all vectors $v_i, v_j \in V$ that preserves the norm $\sum_{k=1}^{d} v_i^k v_j^k$. Over the complex numbers, the orthogonal group only has one possible flavour, the definite one. But over the reals, we also have the indefinite orthogonal groups.

Definition 3.1.2 (Real indefinite orthogonal group). *The real indefinite orthogonal group* $O(p, q, \mathbb{R}) \subset GL(p+q, \mathbb{R})$ *is the subgroup of matrices* $\{R \mid RR^T = I_p^q, \forall R \in GL(p+q, \mathbb{R})\}$.[3] *The indefinite special orthogonal group is the subgroup* $SO(p, q, \mathbb{R}) \subset O(p, q, \mathbb{R})$ *of matrices with unit determinant,* $\{R \mid det(R) = 1, \forall R \in O(p, q, \mathbb{R})\}$.

Another equivalent definition of the indefinite orthogonal group is that it is the group of transformations that preserves $\sum_{k=1}^{p} v_i^k v_j^k - \sum_{k=p+1}^{p+q} v_i^k v_j^k$. All the orthogonal groups (real, complex, definite, and indefinite) are linearly reductive groups.[4] As we will see shortly, the results of Weyl (and consequently the ones we obtain) are identical for all flavours of the orthogonal groups and so we will simply drop the distinction and denote the orthogonal group by $O(d)$ (or $SO(d)$).

[1] We work over $\mathbb{C}$ to make use of the powerful theorems from Chapter 2.

[2] The matrix I_d is the $d \times d$ identity matrix.

[3] The matrix I_p^q is defined as $I_p^q = \begin{pmatrix} I_p & \mathbf{0} \\ \mathbf{0} & -I_q \end{pmatrix}$.

[4] The Lorentz group (including parity transformations) in 4 dimensions is defined to be $O(1, 3, \mathbb{R})$.

Now consider n copies of the $\mathbb{C}$-vector space V, $V^n \cong \mathbb{C}^{nd}$. The polynomials in the coordinates of V^n in some basis, $v_1, \ldots, v_n$, form a graded $\mathbb{C}$-algebra $\mathbb{C}[V^n]$ which, for the sake of clarity, we will henceforth denote by $M_n = \mathbb{C}[V^n]$. The invariant polynomials under the orthogonal group form a subalgebra $M_n^{O(d)} \subset M_n$. A "set of generators" of $M_n^{O(d)}$ is equivalent to a surjective algebra map from some polynomial algebra to $M_n^{O(d)}$. Weyl's famous FFT states that

Theorem 3.1.1 (FFT of the orthogonal group)**.** *There exists a surjective map* $W : \mathbb{C}[y_{ij}] \twoheadrightarrow \mathbb{C}[V^n]^{O(d)}$*, where* $\mathbb{C}[y_{ij}]$ *is the polynomial algebra in variables* $y_{ij}, i, j \in \{1, \ldots, n\}, i \leq j$*, given explicitly on the generators by* $W : y_{ij} \mapsto v_i \cdot v_j$ *and extended to the polynomials in the obvious way.*

For the $SO(d)$-invariant algebra, $M_n^{SO(d)}$, when $n \geq d$, we additionally have the n choose d contractions of the n vectors with the d-dimensional Levi-Civita epsilon tensor, $\epsilon(v_{i_1}, \ldots, v_{i_d})$. The methods Weyl used to prove the above theorem are "signature independent".[5] That is, they are valid in the case of the definite orthogonal groups (over $\mathbb{R}$ and $\mathbb{C}$), where the dot product is defined as $v_i \cdot v_j = \sum_{k=1}^{d} v_i^k v_j^k$, and also in the case of the indefinite orthogonal groups (over $\mathbb{R}$), where $v_i \cdot v_j = \sum_{k=1}^{p} v_i^k v_j^k - \sum_{k=p+1}^{p+q} v_i^k v_j^k$. Now, Weyl's less famous second theorem characterises the relations between these generators.

Theorem 3.1.2 (SFT of the orthogonal group)**.** *The kernel* $\ker W$ *is non-trivial when* $n > d$ *and is the ideal* $I \subset \mathbb{C}[y_{ij}]$ *generated by the* $(d+1)$*-minors of the matrix whose* ij*-th entry is* y_{ij} *for* $i \leq j$ *and* y_{ji} *for* $i > j$*.*

With the $SO(d)$-invariants, one also has to take account of the relations between the epsilons tensors and the dot products. These two theorems will form the foundation of our proofs for the rest of this Chapter.

3.1.1 *Intermezzo*: Quotient-ing out the masses

Before we proceed to our main results, we take a brief detour to perform our quotient-ing out procedure mentioned earlier on the algebras invariant under the orthogonal group with no permutations. Although the results we prove will be unsurprisingly trivial, doing so rigorously will allow us to avoid any potential pitfalls when we attempt to include permutations as well.

Consider the quotient algebra formed by taking the quotient of $M_n^{O(d)}$ with respect to the ideal I generated by the $O(d)$-invariant elements $v_i \cdot v_i - m_i^2$, for all i, $\langle v_i \cdot v_i - m_i^2 | \forall i \rangle$

[5]The reference to this is slightly obscure. It can be found in [42] Section 2.12.

(where we allow the particle mass-squareds $m_i^2 \in \mathbb{C}$). We wish to show that there is a surjective algebra map $\mathbb{C}[y_{ij}, i < j] \twoheadrightarrow M_n^{O(d)}/I$, such that we can use the y_{ij} with $i < j$ as a set of generators.[6]

Theorem 3.1.3. *There exists a surjective* $\mathbb{C}$*-algebra map*

$$W^* : \mathbb{C}[y_{ij}, i < j] \twoheadrightarrow \mathbb{C}[V^n]^{O(d)}/\langle v_i \cdot v_i - m_i^2 | \forall i\rangle$$

which is graded if $m_i^2 = 0$ *for all* i*.*

Proof. The proof has two parts. One part is to show that the Weyl map W induces a surjective algebra map $\mathbb{C}[y_{ij}]/\langle y_{ii} - m_i^2 | \forall i\rangle \twoheadrightarrow M_n^{O(d)}/I$. The other part is to exhibit an algebra isomorphism $\mathbb{C}[y_{ij}]/\langle y_{ii} - m_i^2 | \forall i\rangle \xrightarrow{\sim} \mathbb{C}[y_{ij}, i < j]$.

For the first part, we have that the image of any element in $\langle y_{ii} - m_i^2 | \forall i\rangle$ is in I.[7] And so, by Lemma 2.1.3, we have a surjective algebra map $\mathbb{C}[y_{ij}]/\langle y_{ii} - m_i^2 | \forall i\rangle \twoheadrightarrow M_n^{O(d)}/I$.

To show the second part, consider the polynomial algebra $R[x]$ in one variable over an arbitrary ring R and let

$$\mathrm{ev} : R[x] \twoheadrightarrow R : x \mapsto r$$

be an R-algebra map which we call the *evaluation* map. Since $(x - r)$ is a monic polynomial, by the division algorithm we have that for any $f(x) \in R[x]$, $f(x) = g(x) \cdot (x - r) + s$, with $g(x) \in R[x]$ and $s \in R$. Thus $\mathrm{ev}(f(x)) = s$ and $\ker \mathrm{ev} = \langle x - r\rangle$. By the first isomorphism theorem, $R[x]/\langle x - r\rangle \xrightarrow{\sim} R$.

If we apply this result to $\mathbb{C}[y_{ij}] \cong \mathbb{C}[y_{ij} | (i, j) \neq (1, 1)][y_{11}]$, and then to $\mathbb{C}[y_{ij} | (i, j) \neq (1, 1)] \cong \mathbb{C}[y_{ij} | (i, j) \neq (1, 1), (2, 2)][y_{22}]$, and so on, we can obtain the desired result. Equivalently, an explicit isomorphism $\mathbb{C}[y_{ij}]/\langle y_{ii} - m_i^2 | \forall i\rangle \xrightarrow{\sim} \mathbb{C}[y_{ij}, i < j]$ can be obtained from the evaluation map (which is ungraded, except in the $m_i^2 = 0$ case) from $\mathbb{C}[y_{ij}]$ to $\mathbb{C}[y_{ij}, i < j]$ given by

$$\mathrm{ev} : \mathbb{C}[y_{ij}] \twoheadrightarrow \mathbb{C}[y_{ij}, i < j] : y_{ii}, y_{ij} \mapsto m_i^2, y_{ij} \tag{3.1}$$

whose kernel is indeed $\ker \mathrm{ev} = \langle y_{ii} - m_i^2 | \forall i\rangle$. □

[6] It is important to note that, unless $m_i^2 = 0$ for all i, such that I is homogeneous, $M_n^{O(d)}/I$ is not graded, and so nor is the map.

[7] Actually, $W(\langle y_{ii} - m_i^2 | \forall i\rangle) = \langle v_i \cdot v_i - m_i^2 | \forall i\rangle$, but equality is unnecessary for our purposes.

3.2 The FFT and SFT of the Euclidean group

We now move onto our first main result, namely obtaining the FFT and SFT of the Euclidean group. The Euclidean group in d-dimensions is a semi-direct product, $E(d) = O(d) \ltimes T(d)$, where $T(d) \cong V$ is the translation group (in our case $T(d) = \mathbb{C}^d$).[8] We wish to study the Euclidean-invariant algebra, $M_n^{E(d)}$. We do this by first considering the translation-invariant algebra and then introducing orthogonal group invariance to finally arrive at the Euclidean-invariant algebra.

Theorem 3.2.1. *There exists a surjective* $\mathbb{C}$*-algebra map* $\phi : \mathbb{C}[V^n] \twoheadrightarrow \mathbb{C}[V^n]^{T(d)}$ *given explicitly by* $v_i \mapsto v_i - v_{avg}$*, where* $v_{avg} = \frac{1}{n}\sum_{i=1}^n v_i$.

Proof. First, let us check that ϕ indeed maps elements of M_n to elements of $M_n^{T(d)}$. For any $f \in M_n$, where $f = f(v_1, \dots, v_n)$, the image of f under ϕ is $\phi(f) = f(v_1 - v_{avg}, \dots, v_n - v_{avg})$ which is T-invariant since for any element $t \in T(d)$, $v_i \mapsto v_i + t$ and $v_{avg} \mapsto v_{avg} + t$. This establishes a well-defined injective map $\phi : M_n \to M_n^{T(d)}$. So, we just need to show its surjectivity. For some T-invariant element $f \in M_n^{T(d)}$, $f(v_1, \dots v_n) = f(v_1 + t, \dots v_n + t)$ for any $t \in T(d)$. Since, $T(d) \cong V$, we can choose $t = -v_{avg}$ and so $f = f(v_1 - v_{avg}, \dots, v_n - v_{avg})$. Hence, ϕ is surjective. Finally, it is straightforward to check that ϕ is a $\mathbb{C}$-algebra map. $\square$

We can characterise the kernel of this map. To do so, we use the following theorem from [32].

Theorem 3.2.2. *Consider the algebra* $\mathbb{C}[x_1, \dots, x_n]$ *and the translation-invariant subalgebra* $\mathbb{C}[z_1, \dots, z_n]^{T(1)}$*. Let* ρ *be the* $\mathbb{C}$*-algebra map*

$$\rho : \mathbb{C}[x_1, \dots, x_n] \twoheadrightarrow \mathbb{C}[z_1, \dots, z_n]^{T(1)} : x_i \mapsto z_i - z_{avg}$$

where $z_{avg} = \frac{1}{n}\sum_i z_i$*. Then,* ρ *is surjective and* $\ker \rho = \langle x_{avg} \rangle$.

This is just the one-dimensional version of our theorem above. One can indeed generalise the proof of this theorem to our multi-dimensional case to obtain the kernel. But, doing so involves delving into notions of short exact sequences and isomorphisms in the category of chain complexes which we, and hopefully the reader, would like to avoid. We therefore directly employ the above theorem to obtain $\ker \phi$.

[8] $E(d)$ acts on $v \in V$ as $v \mapsto Rv + t$ where $v \in V$, $R \in O(d)$, and $t \in T(d)$. Note that $E(d)$ is a semi-direct product and the composition of two group elements is non-trivial, $(R_2, t_2) \circ (R_1, t_1) = (R_2R_1, R_2t_1 + t_2)$.

Theorem 3.2.3. *The kernel of the map* $\phi : \mathbb{C}[V^n] \twoheadrightarrow \mathbb{C}[V^n]^{T(d)} : v_i \mapsto v_i - v_{avg}$ *is* $\ker\phi = \langle v^1_{avg}, \ldots, v^d_{avg}\rangle$.

Proof. First, we start by considering $M_n \cong R_i[v^i_1, \ldots, v^i_n]$ as an algebra over R_i, where $R_i = \mathbb{C}[v^k_1, \ldots, v^k_n, \forall k \neq i]$. Now, let ϕ_i denote the following R_i-algebra map

$$\phi_i : R_i[v^i_1, \ldots, v^i_n] \to R_i[v^i_1, \ldots, v^i_n] : v^i_j \mapsto v^i_j - v^i_{avg},$$

whose image is $\operatorname{im}\phi = R_i[v^i_1, \ldots, v^i_n]^{T(1)}$ (where $T(1)$ acts only on v_i) and kernel is $\ker\phi_i = \langle v^i_{avg}\rangle$ by the theorem above. Now, if we compose the maps ϕ_i from $i = 1, \ldots, d$ we get

$$\phi_d \circ \cdots \circ \phi_1 : M_n \to M_n : v^k_j \mapsto v^k_j - v^k_{avg},\ \forall k$$

whose image, by Theorem 3.2.1, is precisely the translation-invariant algebra $M_n^{T(d)}$. This is exactly the map whose kernel we wish to characterise. The kernel can be found using the kernel theorem of the composition of algebra maps as $\ker\phi = \ker\phi_d \circ \cdots \circ \phi_1 = \phi_d^{-1} \circ \cdots \circ \phi_2^{-1}(\langle v^1_{avg}\rangle) = \phi_d^{-1} \circ \cdots \circ \phi_3^{-1}(\langle v^1_{avg}, v^2_{avg}\rangle) = \cdots = \langle v^1_{avg}, \ldots, v^d_{avg}\rangle$. □

By the first isomorphism theorem, we have that $M_n/\langle v^1_{avg}, \ldots, v^d_{avg}\rangle \xrightarrow{\sim} M_n^{T(d)}$. But since $\ker\phi$ is such a simple (and homogeneous) ideal to work with, using the arguments in the previous section (iteratively quotient-ing out the v^i_{avg}) we get that the quotient is just $M_{n-1} = \mathbb{C}[v_1, \ldots, v_{n-1}]$. We therefore have a direct and simple characterisation of the translation-invariant algebra as follows.

Theorem 3.2.4 (FFT and SFT of the translation group)**.** *There exists an isomorphism* $\phi^* : \mathbb{C}[V^{n-1}] \xrightarrow{\sim} \mathbb{C}[V^n]^{T(d)} : v_i \mapsto v_i - v_{avg},$ *for* $1 \leq i \leq n-1$*, where* $v_{avg} = \frac{1}{n}\sum_{i=1}^n v_i$*. That is, a set of algebraically independent generators of the translation-invariant algebra in* n *vectors is given by* $\{v_i - v_{avg},\ 1 \leq i \leq n-1\}$.

We now proceed to include the orthogonal group to obtain a characterisation of the required Euclidean-invariant algebra. Is it easily checked that ϕ^* (and indeed ϕ) is an $O(d)$-equivariant map. That is, given $R \in O(d)$, the diagram

$$\begin{tikzcd} M_{n-1} \arrow[r, two heads, "\phi^*"] \arrow[d, "R"] & M_n^{T(d)} \arrow[d, "R"] \\ M_{n-1} \arrow[r, two heads, "\phi^*"] & M_n^{T(d)} \end{tikzcd}$$

commutes. Since the orthogonal group is linearly reductive, there exists Reynolds operators $\mathcal{R}_{O(d)} : M_{n-1} \twoheadrightarrow M_{n-1}^{O(d)}$ and $\mathcal{R}^T_{O(d)} : M_n^{T(d)} \twoheadrightarrow M_n^{O(d)\ltimes T(d)}$ which furthermore commute with ϕ^*, $\phi^* \circ \mathcal{R}_{O(d)} = \mathcal{R}^T_{O(d)} \circ \phi^*$, since ϕ^* is $O(d)$-equivariant. Using this fact, we obtain the following isomorphism.

Theorem 3.2.5. *ϕ^* restricts*[9] *to an isomorphism $\phi^*| : \mathbb{C}[V^{n-1}]^{O(d)} \xrightarrow{\sim} \mathbb{C}[V^n]^{O(d)\ltimes T(d)}$, where $v_i \mapsto v_i - v_{avg}$.*

Proof. First note that ϕ^* sends an $O(d)$-invariant element in M_{n-1} to an $O(d)$-invariant element in $M_n^{T(d)}$ since ϕ^* is $O(d)$-equivariant. Therefore, we have a well-defined restriction map $\phi^*| : M_{n-1}^{O(d)} \to M_n^{O(d)\ltimes T(d)}$.

It remains to show that $\phi^*|$ is bijective. It is sufficient to show that it is surjective since ϕ^* is an isomorphism. To do so, let $q \in M_n^{O(d)\ltimes T(d)}$. Since ϕ^* is an isomorphism (and hence surjective), there exists an element $h \in M_{n-1}$ such that $\phi^*(h) = q$. But h might not be $O(d)$-invariant. So instead take $\mathcal{R}_{O(d)}(h)$. This is indeed $O(d)$-invariant and furthermore, $\phi^* \circ \mathcal{R}_{O(d)}(h) = \mathcal{R}^T_{O(d)} \circ \phi^*(h) = \mathcal{R}^T_{O(d)}(q) = q$, since ϕ^* commutes with the Reynolds operators and q is $O(d)$-invariant. Thus, $\phi^*|$ is surjective. □

That is, the Euclidean-invariant algebra is isomorphic to the orthogonal group-invariant algebra with one fewer vector. Intuitively, this is due to the fact that since we have translational symmetry, one can choose a "reference vector" in $\mathbb{C}^d$ to act as the origin thereby eliminating one degree of freedom. This isomorphism also shows up in a different, but mathematically identical, setting where one is working in a system where "momentum is conserved" such that $\sum_i v_i = 0$.

Theorem 3.2.6 (FFT of the Euclidean group)**.** *There exists a surjective map $\phi^*| \circ W : \mathbb{C}[y_{ij}, i \leq j \leq n-1] \twoheadrightarrow \mathbb{C}[V^n]^{E(d)}$, where $\mathbb{C}[y_{ij}, i \leq j \leq n-1]$ is the polynomial algebra in variables $y_{ij}, i, j \in \{1, \ldots, n-1\}, i \leq j$, given explicitly on the generators by $\phi^*| \circ W : y_{ij} \mapsto (v_i - v_{avg}) \cdot (v_j - v_{avg})$ and extended to the polynomials in the obvious way.*

Proof. Follows immediately from Theorem 3.2.5 and the FFT of the orthogonal group. □

Similarly to the special orthogonal group, if we consider the algebra invariant under the special Euclidean group $SE(d) \cong SO(d) \ltimes T(d)$, $\mathbb{C}[V^n]^{SE(d)}$, then we additionally have the $n-1$ choose d possible contractions of $v_i - v_{avg}$ with the d-dimensional Levi-Civita tensor as generators.

[9]Given an algebra map $a : R \to S$, and a subalgebra $T \subset R$, with an inclusion map $\iota : T \hookrightarrow R$, we denote the restricted map by $a| = a \circ \iota : T \to S$.

Theorem 3.2.7 (SFT of the Euclidean group). *The kernel* $\ker \phi^*| \circ W$ *is non-trivial when* $n > d+1$ *and is the ideal* $I \subset \mathbb{C}[y_{ij}, 1 \leq i \leq j \leq n-1]$ *generated by the* $(d+1)$*-minors of the matrix whose* ij*-th entry is* y_{ij} *for* $i \leq j \leq n-1$ *and* y_{ji} *for* $j < i \leq n-1$.

Proof. Follows immediately from Theorem 3.2.5 and the SFT of the orthogonal group. □

3.3 Including permutations

We now tackle our second main result, that is, including a permutation group acting on V^n. For the rest of this discussion we will restrict our attention to the orthogonal group. In Section 4.1.3, we discuss how to extend our work to the case of the special orthogonal group as well.

Consider a subgroup $P \subset S_n$ of the permutation group which acts in the permutation representation on the indices $i \in \{1, \dots n\}$. This action induces, in an obvious way, actions on $\{v_i\}$ and $\{y_{ij}\}$ (with the obvious rule that we replace y_{ij} by y_{ji} if $i > j$) and thence on $\mathbb{C}[y_{ij}]$, M_n, and (since the action of permutations commutes with that of the orthogonal group) on $M_n^{O(d)}$. Moreover, it is easily checked that the Weyl map $W : \mathbb{C}[y_{ij}] \twoheadrightarrow M_n^{O(d)}$ is equivariant with respect to P. That is, given $p \in P$, the diagram

$$\begin{array}{ccc} \mathbb{C}[y_{ij}] & \xrightarrow{W} & M_n^{O(d)} \\ \big\downarrow p & & \big\downarrow p \\ \mathbb{C}[y_{ij}] & \xrightarrow{W} & M_n^{O(d)} \end{array}$$

commutes. Additionally, since P is a finite group and hence linearly reductive, there exists Reynolds operators $\mathcal{R}_P : \mathbb{C}[y_{ij}] \twoheadrightarrow \mathbb{C}[y_{ij}]^P$ and $\mathcal{R}_P^O : M_n^{O(d)} \twoheadrightarrow M_n^{O(d)\times P}$ which furthermore commute with the map W since it is P-equivariant. From here, we wish to show that W restricts to a surjective map $\mathbb{C}[y_{ij}]^P \twoheadrightarrow M_n^{O(d)\times P}$, so that a set of generators of $\mathbb{C}[y_{ij}]^P$ furnishes us with a set of generators of $M_n^{O(d)\times P}$ via evaluating $y_{ij} \mapsto v_i \cdot v_j$.

Theorem 3.3.1. *There exists a surjective* $\mathbb{C}$*-algebra map* $W| : \mathbb{C}[y_{ij}]^P \twoheadrightarrow \mathbb{C}[V^n]^{O(d)\times P}$, *where the action of* P *on the indices* $i, j \in \{1, \dots, n\}$ *is via the permutation representation extended to the* y_{ij} *and* v_i *in the obvious way.*

Proof. We first note that W sends a P-invariant polynomial to a P-invariant polynomial; in other words $W(\mathbb{C}[y_{ij}]^P) \subset M_n^{O(d)\times P}$ and so there is a well-defined restriction map

$W| : \mathbb{C}[y_{ij}]^P \to M_n^{O(d)\times P}$. It remains to show that the $W|$ map surjects. To do so, let $q \in M_n^{O(d)\times P} \subset M_n^{O(d)}$. Since W is onto, there exists $r \in \mathbb{C}[y_{ij}]$ such that $W(r) = q$. But r is not necessarily P-invariant, so consider instead $\mathcal{R}_P(r)$. This is P-invariant and moreover, we have that $W| \circ \mathcal{R}_P(r) = \mathcal{R}_P^O \circ W|(r) = \mathcal{R}_P^O(q) = q$, where we used the facts that W commutes with the Reynolds operators and that q is P-invariant by assumption. Thus $W|$ is onto. □

The equivalent result for the algebra invariant under the Euclidean group can be obtained in two ways. The first is by exploiting the isomorphism we proved in the previous Section to exhibit an isomorphism $\phi^*| : M_{n-1}^{O(d)\times P} \xrightarrow{\sim} M_n^{E(d)\times P}$. When $P \subset S_{n-1}$ the result follows similarly to the proof of the above theorem since here, the action of P on the indices of the $v_i \in V^{n-1}$ (in the orthogonal invariant algebra) and the $v_i \in V^n$ (in the Euclidean invariant algebra) is via the permutation representation which commutes with the isomorphism $\phi^*|$. But if the full permutation group, $P = S_n$, acts on the $v_i \in V^n$ (in the Euclidean invariant algebra) via the permutation representation, then care needs to be taken since the representation of S_n on the $v_i \in V^{n-1}$ (in the orthogonal invariant algebra) is not simply the permutation representation. Instead here, S_n acts via its standard representation.[10] From here, it is easy to show that this action of S_n commutes with $\phi^*|$ and so the rest of the proof follows identically. We therefore obtain that

Theorem 3.3.2. *There exists an isomorphism $\phi^*|| : \mathbb{C}[V^{n-1}]^{O(d)\times P} \xrightarrow{\sim} \mathbb{C}[V^n]^{E(d)\times P}$, where, when $P \subset S_n$, P acts via the permutation representation on the indices of the $v_i \in V^{n-1}$ (in the orthogonal invariant algebra) and the $v_i \in V^n$ (in the Euclidean invariant algebra) and, when $P = S_n$, P acts via the permutation representation on $v_i \in V^n$ (in the Euclidean invariant algebra) and via the standard representation on $v_i \in V^{n-1}$ (in the orthogonal invariant algebra).*

Consequently, we also obtain the following

Corollary 3.3.1. *There exists a surjective $\mathbb{C}$-algebra map $\phi^*|| \circ W| : \mathbb{C}[y_{ij}, i \leq j \leq n-1]^P \twoheadrightarrow \mathbb{C}[V^n]^{E(d)\times P}$, where action of P on the indices is as given in Theorem 3.3.2 extended to the y_{ij} in the obvious way.*

The first result above is useful for mapping the Hironaka decompositions of the algebras $\mathbb{C}[V^{n-1}]^{O(d)\times P}$ to Hironaka decompositions of $\mathbb{C}[V^n]^{E(d)\times P}$. But when trying to

[10] The standard representation of the permutation group S_n on some basis $e_1, \dots e_n$ is via permutations but where $e_1 + \cdots + e_n = 0$. One can then choose a basis for this representation, usually $\{e_i - e_{i+1},\ 1 \leq i \leq n-1\}$. In our case, the basis is simply $\{v_i,\ 1 \leq i \leq n-1\}$.

find minimal algebra generators, the added nuisance of dealing with two representations of the permutation group in the second result is not worth the effort. In this case, it is more useful to work with the less direct characterisation of the translation-invariant algebra (without quotient-ing out the kernel) to obtain a surjective $\mathbb{C}$-algebra map $\phi| : \mathbb{C}[V^n]^{O(d)} \twoheadrightarrow \mathbb{C}[V^n]^{E(d)}$ with a non-trivial kernel. The advantage of this is that now, the action of $P \subseteq S_n$ is via the permutation representation on both algebras and $\phi|$ restricts to a surjective map $\phi|| : \mathbb{C}[V^n]^{O(d)\times P} \twoheadrightarrow \mathbb{C}[V^n]^{E(d)\times P}$. Therefore, we obtain the following surjective map

Theorem 3.3.3. *There exists a surjective $\mathbb{C}$-algebra map $\phi|| \circ W| : \mathbb{C}[y_{ij}]^P \twoheadrightarrow \mathbb{C}[V^n]^{E(d)\times P}$, where the action of P on the indices $i, j \in \{1, \dots, n\}$ is via the permutation representation extended to the y_{ij} and v_i in the obvious way.*

3.4 Quotient-ing out the masses: with permutations

Finally, we move onto our last main result where we carry out our quotient-ing procedure on the algebras $M_n^{O(d)\times P}$ invariant under the orthogonal and permutation groups. Our goal is to show that there exists a surjective algebra map[11] $\mathbb{C}[y_{ij}|i<j]^P \twoheadrightarrow M_n^{O(d)\times P}/J$ where $J = \langle v_i^2 - m_i^2 | \forall i\rangle \cap M_n^{O(d)\times P}$. Again, the proof has two parts. One is to show that the restricted Weyl map $W|$ induces a surjective algebra map $\mathbb{C}[y_{ij}]^P/J' \twoheadrightarrow M_n^{O(d)\times P}/J$, where $J' = \langle y_{ii} - m_i^2 | \forall i\rangle \cap \mathbb{C}[y_{ij}]^P$, and the other is to exhibit an algebra isomorphism $\mathbb{C}[y_{ij}]^P/J' \xrightarrow{\sim} \mathbb{C}[y_{ij}|i<j]^P$.

For the first part, we begin by showing that the image $W(J') \subset J$. For an element $j' \in J'$, $j' \in \langle y_{ii} - m_i^2 | \forall i\rangle$ and $j' \in \mathbb{C}[y_{ij}]^P$ by definition. But since the image $W(\langle y_{ii} - m_i^2 | \forall i\rangle) \subset I$, the image $W(j') \in I$. Furthermore, the element j' is P-invariant by assumption and as the map W is P-equivariant, the image $W(j')$ is also P-invariant. So, $W(j') \in J$ and hence $W(J') \subset J$.[12] And so, by Lemma 2.1.3, we again have a surjective algebra map.

For the second part, it turns out that the required result follows from a more general lemma.

Lemma 3.4.1. *Suppose that a linearly reductive group G acts reducibly on a $\mathbb{C}$-vector space $V = X \oplus Z$ and suppose that the representation carried by X is further reducible, containing the trivial representation. Let $\{x_i\}$ and $\{z_i\}$, respectively, be bases of the*

[11] Again, ungraded unless $m_i^2 = 0$.

[12] Actually, $W(J') = J$, but equality is unnecessary for our purposes.

dual spaces X^* *and* Z^*,[13] *respectively, and let* $a \in X$ *denote a* G*-invariant vector with components* $a_i = x_i(a) \in \mathbb{C}$. *Further, consider the algebras* $R = \mathbb{C}[x_1, \ldots, x_m, z_1, \ldots, z_n]$ *and* $S = \mathbb{C}[z_1, \ldots, z_n]$ *along with the evaluation map* $\mathrm{ev} : R \twoheadrightarrow S$, $x_i \mapsto a_i$, *with kernel* $\langle x_i - a_i | \forall i \rangle$. *Then, there exists an isomorphism of (ungraded if* $a_i \neq 0$*) algebras* $R^G/J \xrightarrow{\sim} S^G$, *where* R^G, S^G *are the* G*-invariant subalgebras of* R, S *respectively and* $J = \langle x_i - a_i | \forall i \rangle \cap R^G$ *is an ideal of* R^G.

Proof. To prove this, we start by explicitly defining the action of $g \in G$ on $h \in R^G$ and $f \in S$ via the reducible representation $\rho : G \to GL(V, \mathbb{C}) : g \mapsto \rho_X(g) \oplus \rho_Z(g)$ to be as follows

$$h \mapsto h^g = h(\rho_X(g)x_i, \rho_Z(g)z_i) = h(x_i, z_i) = h, \tag{3.2}$$

$$f \mapsto f^g = f(\rho_Z(g)z_i). \tag{3.3}$$

Next, we define the inclusion map $i : R^G \hookrightarrow R$ and compose it with the evaluation map to get the restricted algebra map $\mathrm{ev}| := \mathrm{ev} \circ i : R^G \to S$. It can then be checked that the evaluation map $\mathrm{ev}|$ is equivariant with respect to G. That is, given $g \in G$, the diagram

$$\begin{array}{ccc} R^G & \xrightarrow{\mathrm{ev}|} & S \\ \downarrow{\scriptstyle g} & & \downarrow{\scriptstyle g} \\ R^G & \xrightarrow{\mathrm{ev}|} & S \end{array}$$

commutes. Furthermore, since G is linearly reductive, there exists Reynolds operators $\mathcal{R}_G^R : R \twoheadrightarrow R^G$ and $\mathcal{R}_G^S : S \twoheadrightarrow S^G$, which commute with the map ev since it is G-equivariant. Now as the map $\mathrm{ev}|$ is G-equivariant, it sends a G-invariant polynomial to a G-invariant polynomial; in other words $\mathrm{ev}|(R^G) \subset S^G$ and so we have a well-defined restriction map $\mathrm{ev}| : R^G \to S^G$. It remains to show that $\mathrm{ev}|$ is surjective. To do so, let $s \in S^G \subset S$. Since ev is onto, there exists $r \in R$ such that $\mathrm{ev}(r) = s$. But r is not necessarily G-invariant, so consider instead $\mathcal{R}_G^R(r)$. This is G-invariant and we have, furthermore, that $\mathrm{ev}| \circ \mathcal{R}_G^R(r) = \mathcal{R}_G^S \circ \mathrm{ev}|(r) = \mathcal{R}_G^S(s) = s$, where we have used the fact that $\mathrm{ev}|$ commutes with the Reynolds operators and that s is G-invariant by assumption. Thus, $\mathrm{ev}|$ is onto. The last ingredient of the proof is to note that the kernel of the map $\ker \mathrm{ev}|$ is the restriction of the ideal $\langle x_i - a_i | \forall i \rangle$ to the G-invariant subalgebra $J = \langle x_i - a_i | \forall i \rangle \cap R^G$. Finally, by the first isomorphism theorem, $R^G/J \xrightarrow{\sim} S^G$. □

[13] The dual spaces are also denoted by $\mathrm{Hom}(X, \mathbb{C})$ and $\mathrm{Hom}(Z, \mathbb{C})$ since they are vector space homomorphisms to $\mathbb{C}$.

In our specific case, the variables y_{ii}, y_{ij} transform under reducible representations of the permutation group P with the representation of y_{ii}, $(1 \oplus (n-1,1))$, being further reducible containing the trivial representation. Furthermore, the masses m_i^2 clearly form an invariant vector when the particles (and hence the masses) are identical. Hence, the previous theorem applies and we have an isomorphism of (ungraded, except in the massless case) algebras $\mathbb{C}[y_{ij}]^P/J' \xrightarrow{\sim} \mathbb{C}[y_{ij}|i<j]^P$. Therefore, we have that

Theorem 3.4.1. *There exists a surjective* $\mathbb{C}$*-algebra map*

$$W|^* : \mathbb{C}[y_{ij}, i<j]^P \twoheadrightarrow \mathbb{C}[V^n]^{O(d)\times P}/(\mathbb{C}[V^n]^{O(d)\times P} \cap \langle v_i \cdot v_i - m_i^2 | \forall i \rangle)$$

which is graded if $m_i^2 = 0$ *for all* i*.*

3.5 Diagram

We summarise the results of this Chapter in the following commutative diagram

$$
\begin{tikzcd}
 & & & \mathbb{C}[V^n] \arrow[rr, two heads, "\phi^*"] \arrow[dl, two heads, "\mathcal{R}_P"] \arrow[dd, two heads, "\mathcal{R}_{O(d)}"] & & \mathbb{C}[V^{n+1}]^{T(d)} \arrow[dl, two heads, "\mathcal{R}_P"] \arrow[dd, two heads, "\mathcal{R}_{O(d)}"] \\
 & & \mathbb{C}[V^n]^P \arrow[rr, two heads] \arrow[dd, two heads, "\mathcal{R}_{O(d)}"] & & \mathbb{C}[V^{n+1}]^{T(d)\times P} \arrow[dd, two heads, "\mathcal{R}_{O(d)}"] & \\
 & \mathbb{C}[y_{ij}] \arrow[rr, two heads, "W"] \arrow[dl, two heads, "\mathcal{R}_P"] \arrow[dd, two heads, "\mathrm{ev}"] & & \mathbb{C}[V^n]^{O(d)} \arrow[rr, two heads, "\phi^*|"] \arrow[dl, two heads, "\mathcal{R}_P"] \arrow[dd, two heads] & & \mathbb{C}[V^{n+1}]^{E(d)} \arrow[dl, two heads, "\mathcal{R}_P"] \\
\mathbb{C}[y_{ij}]^P \arrow[rr, two heads, "W|"] \arrow[dd, two heads, "\mathrm{ev}|"] & & \mathbb{C}[V^n]^{O(d)\times P} \arrow[rr, two heads, "\phi^*||"] \arrow[dd, two heads] & & \mathbb{C}[V^{n+1}]^{E(d)\times P} & \\
 & \mathbb{C}[y_{ij}, i<j] \arrow[rr, two heads, "W^*"] \arrow[dl, two heads, "\mathcal{R}_P"] & & \mathbb{C}[V^n]^{O(d)}/I \arrow[dl, two heads, "\mathcal{R}_P"] & & \\
\mathbb{C}[y_{ij}, i<j]^P \arrow[rr, two heads, "W|^*"] & & \mathbb{C}[V^n]^{O(d)\times P}/J & & &
\end{tikzcd}
$$

Figure 3.1 Commutative diagram showing the algebras invariant under the orthogonal, Euclidean, and permutation groups along with the maps between them. In addition to the Reynolds operators, $\mathcal{R}$, we have the following (sometimes restricted) maps: the Weyl map $W : y_{ij} \mapsto v_i \cdot v_j$, the translation map $\phi^* : v_i \mapsto v_i - v_{avg}$, and the evaluation map $\mathrm{ev} : y_{ii} \mapsto m_i^2$.

4

Generators of invariant algebras

Having gained a solid understanding of the different algebras of interest, we now proceed to use the tools of invariant theory to find sets of generators for these algebras.

In the first part, we find generators of the algebras $\mathbb{C}[y_{ij}]^P$ since by the surjective maps established in the previous Chapter, a set of generators of $\mathbb{C}[y_{ij}]^P$ provides us with a set of generators of $\mathbb{C}[V^n]^{O(d)\times P}$ and $\mathbb{C}[V^n]^{E(d)\times P}$. Finding a set of generators of $\mathbb{C}[y_{ij}]^P$ is where the real hard work begins. Indeed, while the action of P on the subalgebra $\mathbb{C}[y_{ii}]$ is via the permutation representation, whose invariant algebra is well-understood (a result due to Gauss in the "worst-case scenario" $P = S_n$ tells us, for example, that $\mathbb{C}[y_{ii}]^{S_n}$ is isomorphic to the polynomial algebra in n variables with degrees $1, \ldots, n$), the invariants of the action of P on the subalgebra $\mathbb{C}[y_{ij}|i < j]$ are rather harder to describe, with a known description of $\mathbb{C}[y_{ij}|i < j]^{S_n}$ only known for $n < 5$, even though an algorithm is available [11].

Thus, we content ourselves with finding generators for n vectors (particles) in which at most 4 are identical, using the fact that the algebra of invariants is Cohen-Macaulay (as P is a finite group and thus linearly reductive) and therefore possesses a Hironaka decomposition. That is, we may write $\mathbb{C}[y_{ij}]^P = \bigoplus_k \eta_k \mathbb{C}[\theta_l]$, where η_k and θ_l, the secondaries and HSOP respectively, are polynomials in y_{ij}. Evidently, η_k and θ_l collectively generate $\mathbb{C}[y_{ij}]^P$.

There exist algorithms for computing η_k and θ_l, though even modern computers quickly run out of steam (hence the difficulties when $n \geq 5$). In this way, we are able to find a set of generators, whose number is typically rather large (for $pp \to jjj$, for example, we have 10 primaries and 360 secondaries for $\mathbb{C}[y_{ij}, i < j]^{S_2 \times S_3}$). To pare it down to a more manageable number, we employ two further strategies. Firstly, the form of the Hironaka decomposition implies that algebra multiplication, encoded in the relations $\eta_k \eta_m = \sum_j f^j_{km} \eta_j$, $f^n_{km} \in \mathbb{C}[\theta_l]$, can often be used to remove some generators which are redundant in the sense that they can be obtained as algebraic combinations of other generators. This reduction results in a minimal set of generators of the algebra. The price to pay is that the description of the algebra in terms of the remaining generators becomes more complicated and the resulting non-trivial relations between the generators introduces inefficiencies into our description (as we will see in Chapter 6). Secondly, in particle physics collisions, the dot products $v_i \cdot v_i$ do not vary from event to event, being fixed equal to the invariant masses m_i^2, and so we can use our construction of the quotient algebra $\mathbb{C}[y_{ij}|i < j]^P \twoheadrightarrow \mathbb{C}[V^n]^{O(d)\times P}/(\langle v_k^2 - m_k^2 | \forall k\rangle \cap \mathbb{C}[V^n]^{O(d)\times P})$ and work with the simpler algebra $\mathbb{C}[y_{ij}|i < j]^P$ to obtain a more manageable set of generators.

Even without permutation invariance, the SFTs imply that the maps to $\mathbb{C}[V^n]^{O(d)\times P}$ and $\mathbb{C}[V^n]^{E(d)\times P}$ do not inject for $n > d$ and $n > d+1$ respectively. This means that there are yet further relations between the generators of $\mathbb{C}[V^n]^{O(d)\times P}$ and $\mathbb{C}[V^n]^{EO(d)\times P}$ (beyond those in $\mathbb{C}[y_{ij}]^P$), which may be rather obscure and which may yet further frustrate phenomenological analyses. In Section 4.2, we exploit the fact that the algebras $\mathbb{C}[V^n]^{O(d)\times P}$ are themselves Cohen-Macaulay, meaning that they too admit Hironaka decompositions (which can also be mapped to Hironaka decompositions of $\mathbb{C}[V^{n+1}]^{E(d)\times P}$ via the isomorphism exhibited previously), to describe them directly and give some explicit examples.

We also describe the effects of removing parity-invariance (which amounts to replacing $O(d)$ and $E(d)$ by their subgroups $SO(d)$ and $SE(d)$). For the minimal algebra generators, this is conceptually straightforward, in that it can be achieved by adding further variables $z_{i_1\dots i_d}$ to the y_{ij}, which map under W to contractions of the epsilon tensor in d dimensions with n momenta (vectors). But in practice, elucidating the structure of the corresponding algebra of permutation invariants quickly becomes complicated. For Hironaka decompositions though, the extension requires no additional effort in finding the HSOPs, since any HSOP of $\mathbb{C}[V^n]^{O(d)\times P}$ can also be shown to be an HSOP of $\mathbb{C}[V^n]^{SO(d)\times P}$. The only extra work needed is in finding the secondaries.

4.1 Minimal algebra generators

We begin our expedition by tackling the first proposed problem, namely generalising Weyl's FFT. In layman's terms, the FFT is the statement that every Lorentz-invariant polynomial can be obtained by taking an arbitrary polynomial in variables y_{ij} (where $i, j \in \{1, \dots, n\}$ and $i \leq j$), and replacing $y_{ij} \mapsto v_i \cdot v_j$. The results of Section 3.3 show that every Lorentz- (or Euclidean-) and permutation-invariant polynomial can be obtained by taking a permutation-invariant polynomial in y_{ij} (where the permutation group P acts on the indices i, j as $y_{ij} \mapsto y_{\sigma(i)\sigma(j)}$, for $\sigma \in P$) and making the replacement $y_{ij} \mapsto v_i \cdot v_j$ (or $y_{ij} \mapsto (v_i - v_{avg}) \cdot (v_j - v_{avg})$). In a sense, these results *are* the generalizations of Weyl's FFT, but not only are they apparently completely trivial, but also they are completely useless as they stand, because of the difficulty of describing the permutation-invariant polynomials in y_{ij}. Our goal is therefore to find a set of generators of the algebra $\mathbb{C}[y_{ij}]^P$, which will in turn provide us with a set of generators for $\mathbb{C}[V^n]^{O(d)\times P}$. In the case considered by Weyl, where P is the trivial group, this is a triviality, since $\mathbb{C}[y_{ij}]$ is a polynomial algebra and so a set of generators (which is moreover a minimal set of generators) is given by $\{y_{ij}\}$. In cases where P is not the trivial group, finding a set of generators is rather harder than it may first appear. To see why this is the case, consider the "worst case scenario" $P = S_n$. The group S_n acts reducibly on the subspaces with bases $\{y_{ii}\}$ and $\{y_{ij}\}$, so there is a well-defined action (for any $P \subset S_n$, in fact) on the polynomial subalgebras $\mathbb{C}_= := \mathbb{C}[y_{ii}]$ and $\mathbb{C}_< := \mathbb{C}[y_{ij}|i < j]$; to begin with, it is helpful to consider these separately.

The action of S_n on $\{y_{ii}\}$ is via the natural permutation representation (in terms of irreducible representations in partition notation it is $1 \oplus (n-1,1)$) and a complete description of the invariant algebra $\mathbb{C}_=^{S_n}$ was given by Gauss: it is isomorphic (as a graded $\mathbb{C}$-algebra) to the polynomial algebra in n variables with degrees $1, \dots, n$. For an explicit isomorphism, one can take *e.g.* the *elementary symmetric polynomials* $e_1 = \sum_i y_{ii}, e_2 = \sum_{i<j} y_{ii}y_{jj}, \dots, e_n = \prod_i y_{ii}$ or the *power sum polynomials* $\sum_i y_{ii}^k$ with $k \in \{1, \dots n\}$.

The action of S_n on $\{y_{ij}\}$ is non-standard (in terms of irreducible representations it is $1 \oplus (n-1,1) \oplus (n-2,2)$ [15]). A description of the invariant algebra is trivial in $n = 2, 3$, being given by polynomial algebras in 1 and 3 variables, respectively, but was only determined relatively recently for $n = 4$ [2] and is unknown for $n \geq 5$. It is important to note that the invariant algebra is not a polynomial algebra for $n \geq 4$. Rather, like any algebra of invariants under the action of a finite group, it has the more general structure of a Cohen-Macaulay algebra. Such algebras admit a Hironaka decomposition as a free, finitely-generated module over a polynomial subalgebra.

Since an explicit description of $\mathbb{C}_<^{S_n}$ is, in general, unavailable, it is unrealistic to expect one to be available for the full invariant algebra $\mathbb{C}_\leq^P$ (where we use $\mathbb{C}_\leq$ as a shorthand to denote the full $\mathbb{C}[y_{ij}|i \leq j]$). But, since it too has the Cohen-Macaulay property, we can use the available algorithms to find a Hironaka decomposition in simple cases. As we will see, the number of primaries and secondaries that arise in such cases is rather large. This is where the benefit of the quotient-ing out procedure comes into play and pares down the number of generators by allowing us to work with the algebra $\mathbb{C}_{<}^{S_n}$.

4.1.1 Generators of $\mathbb{C}[y_{ij}]^P$

We now describe results from the theory of invariants which together may be used to find sets of generators for the algebras of permutation invariants, such as $\mathbb{C}_{<}^{P}$. For more details, see *e.g.* [11, 35].

As discussed previously, given a Hironaka decomposition of an invariant algebra $K[V]^G = \bigoplus_k \eta_k K[\theta_1 \ldots, \theta_l]$, the set containing the primary and secondary invariants, $\{\eta_i, \theta_j\}$, forms a set of generators of $K[V]^G$, which is what we seek. A Hironaka decomposition can be found by a two-step process. The first step is to find a HSOP. The necessary and sufficient conditions for a set of homogeneous elements in $K[V]^G$ to form such a system are that they be algebraically independent and that they satisfy the nullcone condition described in Section 2.2.

Finding a HSOP has been reduced to an (unwieldy) algorithm [7, 10], but we will not need it here. Indeed, the group P acts on $\mathbb{C}_{<}^{P}$, say (an analogous result holds for $\mathbb{C}_{\leq}^{P}$), by permuting the y_{ij} amongst themselves; but it is easily shown (*cf.* [11], Example 2.4.9) that for any permutation subgroup of $S_{n(n-1)/2}$, a HSOP is given by the $n(n-1)/2$ elementary symmetric polynomials in y_{ij}.

For our purposes, this HSOP is sometimes less than optimal, because it introduces primary invariants of unnecessarily high degrees, leading to more secondary invariants (as can easily be seen by considering the case where P is the trivial group, such that $\{y_{ij}\}$ is a HSOP, with primary invariants all of degree 1). A HSOP with primary invariants of lower degrees can be found by partitioning the y_{ij} into their orbits under P and forming the respective sets of elementary symmetric polynomials. Again, one may easily show that the union of these forms a HSOP.

Let us make this explicit in our $pp \to jjj$ example. Labelling the proton momenta by $4, 5$ and the jet momenta by $1, 2, 3$, we have the following orbits: $\{y_{45}\}, \{y_{12}, y_{13}, y_{23}\}$, $\{y_{14}, y_{15}, y_{24}, y_{25}, y_{34}, y_{35}\}$. Following our prescription, the HSOP will be

$$\begin{array}{llll} e_1(y_{45}), & e_1(y_{12}, y_{13}, y_{23}), & e_1(y_{14}, y_{24}, y_{34}, y_{15}, y_{25}, y_{35}), & e_4(y_{14}, y_{24}, y_{34}, y_{15}, y_{25}, y_{35}), \\ & e_2(y_{12}, y_{13}, y_{23}), & e_2(y_{14}, y_{24}, y_{34}, y_{15}, y_{25}, y_{35}), & e_5(y_{14}, y_{24}, y_{34}, y_{15}, y_{25}, y_{35}), \\ & e_3(y_{12}, y_{13}, y_{23}), & e_3(y_{14}, y_{24}, y_{34}, y_{15}, y_{25}, y_{35}), & e_6(y_{14}, y_{24}, y_{34}, y_{15}, y_{25}, y_{35}). \end{array} \tag{4.1}$$

When working with the full algebra $\mathbb{C}[y_{ij}, i \leq j]^P$, one must also include the elementary symmetric polynomials on the orbits of the $\{y_{ii}\}$.

Having found HSOPs of $\mathbb{C}[y_{ij}]^P$, we turn to the second step in finding a Hironaka decomposition, which is to find the corresponding secondary invariants. This is done via the secondaries algorithm in Section 2.2 which in turn requires the Hilbert series of these algebras. By way of illustration, Table A.1 of Appendix A lists the Hilbert series for a few of the algebras that we are interested in. Once the secondaries are found, the resulting Hironaka decomposition provides a complete characterisation of the invariant algebras.

4.1.2 Redundancies

In the previous Subsection, we described a systematic method for the construction of a Hironaka decomposition, and *ergo* a set of generators, for $\mathbb{C}^P_<$ (an analogous construction applies for $\mathbb{C}^P_\leq$). Unfortunately, the number of generators is rather large in all but the simplest cases. For the purpose of carrying out phenomenological analyses, one would like to have a set of generators that is as minimal as possible, in the sense of reducing both the number of generators and their degrees. In this Subsection, we will see that such a reduction is indeed possible, and leads to a set of generators whose cardinality is minimal (the degrees of the generators in such a set is moreover fixed). Unfortunately, the number of generators in such a set is still often rather large. But this is the best one can do.

The reduction may be achieved (at the cost of destroying the neat encoding of the algebraic structure in the Hironaka decomposition, which in itself is useful for phenomenological analyses as we will see in Chapter 6) via the following algorithm: For a set of generators S, choose an element $f \in S$ and set up a general element of the same degree as f in the algebra generated by $S \setminus f$ with unknown coefficients. Equate it to f and extract the corresponding system of linear equations by comparison of coefficients. The system is solvable if and only if f can be omitted from S. It turns out [11], though we will not show it here, that this procedure leads to a set of algebra generators whose cardinality is minimal; the degrees of the resulting generators are, moreover, uniquely determined.

It seems that we are home and dry, but there is one remaining issue: although the problem of finding the secondary generators is solved algorithmically, in most non-trivial cases, it is highly inefficient. Even modern computers using state-of-the-art algorithms start struggling with Hironaka decompositions containing more than a few hundred secondaries.[1] Our only hope is if we can somehow get away with finding some, but not all, of the secondaries before using the elimination procedure just described. Luckily, this hope comes in the form of Emmy Noether, who showed that the maximal degree of an algebra generator in a minimal set is $\leq |G|$ when G is a finite group. When G is non-cyclic (so $P \neq S_1, S_2$ in the case at hand), Noether's bound can be improved to $\frac{3}{4}|G|$ if $|G|$ is even and $\frac{5}{8}|G|$ if $|G|$ is odd [13].[2] Therefore, we only need to find the secondaries up to these bounds before discarding the redundant generators using the process outlined above. Of course, in many cases these bounds are practically useless; the order of S_n is $n!$. But for physically relevant examples such as $S_2 \times S_3$, they reduce the computation time significantly.

4.1.3 Parity

Finally, we briefly discuss the more general case where parity is not a symmetry. Weyl showed that a generating set of Lorentz invariants in d dimensions is given by the dot products, along with all the possible contractions of vectors with the anti-symmetric d dimensional Levi-Civita epsilon tensor[3]. To include these extra generators in our discussion, one could add some extra variables $z_{i_1,\cdots,i_d}$ which transform in a similar (anti-symmetric) manner to the epsilons under the action of the permutation group and are mapped to the epsilons in the appropriate way under the Weyl map. One then needs to study the algebra $\mathbb{C}[y_{ij}, z_{i_1,\cdots,i_d}]^P$ and find its Hironaka decomposition and consequently a set of minimal algebra generators.[4] The first challenge one runs into in trying to do so is the difficulty in finding a suitable HSOP. Since the elements in P act on $z_{i_1,\cdots,i_d}$ by permutation, a HSOP is given by the elementary symmetric polynomials in $z_{i_1,\cdots,i_d}$, but the degrees of the resulting generators are prohibitively large, with a consequent slew of secondaries. Given the inefficiencies of current algorithms, which

[1] This is partly due to the fact that it is computationally expensive to compute Gröbner bases, but also obtaining a basis of the homogeneous component $K[V]_{d_i}$ becomes increasingly expensive at higher degrees.

[2] In our $pp \to jjj$ example, we have $3|S_2 \times S_3|/4 = 9$, which comfortably exceeds our highest degree primary, of degree 6; we will see in the next Subsection that the highest degree generator in a minimal set of algebra generators is in fact 6.

[3] There are, of course, relations between the Levi-Civita tensors and the dot products, namely the product of two epsilon tensors contracted with some vectors v_i is equal to the corresponding minor of the $v_i \cdot v_j$ matrix.

[4] One could even quotient this algebra by the ideal generated by the relations between the y_{ij} and the $z_{i_1,\cdots,i_d}$ (corresponding to the relations between the epsilons and dot products) to make ones life simpler. It can then be shown that the HSOPs for this quotient algebra is the same as those for $\mathbb{C}[y_{ij}]^P$. The difficulty consequently gets condensed into finding the Hilbert series needed for the secondaries algorithm, since the problem of computing Hilbert series of quotient algebras is not completely solved yet.

already struggle with the simpler case of $\mathbb{C}^P_<$, it seems unlikely that one will be able to find a minimal set of generators in this way, in all but the simplest cases.[5]

4.1.4 Examples of minimal algebra generators

We will now apply the aforementioned techniques to find sets of generators for common examples of phenomenological interest. A list of minimal algebra generators of multiple cases of physical interest can be found on-line.[6]

$pp \to jj$

A common scattering problem is the two protons to two jets, $pp \to jj$, though of course jj could be any two objects that we do not want to or cannot distinguish, which corresponds to the $n = 4$ with $S_2 \times S_2$ case. Since this example is not too difficult to tackle, we will work with the full permutation invariant algebra $\mathbb{C}[y_{ij}, i \leq j \leq 4]^{S_2 \times S_2}$, that is, without quotient-ing out the masses (this corresponds to situations where the particles could be off-shell and hence $v_i \cdot v_i = m_i^2$ does not hold).

First, we find the primaries using our prescription. The invariant subspaces are $\{y_{12}\}$, $\{y_{34}\}$, $\{y_{13}, y_{14}, y_{23}, y_{24}\}$, $\{y_{11}, y_{22}\}$, $\{y_{33}, y_{44}\}$, and therefore we take the primaries to be

$$\begin{aligned}
&e_1(y_{12}), && e_1(y_{11}, y_{22}), \quad e_2(y_{11}, y_{22}),\\
&e_1(y_{34}), && e_1(y_{33}, y_{44}), \quad e_2(y_{33}, y_{44}),\\
&e_1(y_{13}, y_{14}, y_{23}, y_{24}), && e_3(y_{13}, y_{14}, y_{23}, y_{24}),\\
&e_2(y_{13}, y_{14}, y_{23}, y_{24}), && e_4(y_{13}, y_{14}, y_{23}, y_{24}).
\end{aligned}$$

We can already see directly from the improved Noether bound (which is $\frac{3}{4}(2!)(2!) = 3$ in this case) that these generators cannot be part of a minimal set. To read off the degrees of secondaries, we find the Hilbert series and cast it in the required form

$$H(\mathbb{C}[y_{ij}, i \leq j \leq 4]^{S_2 \times S_2}, t) = \frac{1 + 4t^2 + 3t^3 + 7t^4 + 4t^5 + 4t^6 + t^7}{(1-t)^5(1-t^2)^3(1-t^3)(1-t^4)}.$$

[5] The issue of parity was addressed in a different approach in another work [30] by attacking the problem in a more direct hands-on method.

[6] Available at: https://github.com/WardHaddadin/Invariant-polynomials-and-machine-learning.

Next, we use the algorithm to compute the secondaries. Using the bound, we only need to find the secondaries up to degree 3 (which saves us the effort of finding 16 secondries). Once found, we can start eliminating redundancies from the union of primaries and secondaries in the fashion described in Subsection 4.1.2. Once this is done, we are left with a set of 16 minimal algebra generators given in Table 4.1.

Degree = 1
$g_1 = y_{12}$,
$g_2 = y_{34}$,
$g_3 = y_{11} + y_{22}$,
$g_4 = y_{33} + y_{44}$,
$g_5 = y_{13} + y_{14} + y_{23} + y_{24}$,
Degree = 2
$g_6 = y_{11}y_{22}$,
$g_7 = y_{33}y_{44}$,
$g_8 = y_{13}y_{14} + y_{23}y_{24}$,
$g_9 = y_{13}y_{23} + y_{14}y_{24}$,
$g_{10} = y_{14}y_{23} + y_{13}y_{24}$,
$g_{11} = y_{11}y_{13} + y_{11}y_{14} + y_{22}y_{23} + y_{22}y_{24}$,
$g_{12} = y_{13}y_{33} + y_{23}y_{33} + y_{14}y_{44} + y_{24}y_{44}$,
Degree = 3
$g_{13} = y_{13}^3 + y_{14}^3 + y_{23}^3 + y_{24}^3$,
$g_{14} = y_{11}y_{13}^2 + y_{11}y_{14}^2 + y_{22}y_{23}^2 + y_{22}y_{24}^2$,
$g_{15} = y_{33}y_{13}^2 + y_{23}^2y_{33} + y_{14}^2y_{44} + y_{24}^2y_{44}$,
$g_{16} = y_{11}y_{13}y_{33} + y_{22}y_{23}y_{33} + y_{11}y_{14}y_{44} + y_{22}y_{24}y_{44}$.

Table 4.1 Table of minimal algebra generators for $\mathbb{C}[y_{ij}, i \le j \le 4]^{S_2 \times S_2}$.

$pp \to jjj$

We now ramp up the level of complexity, by considering $pp \to jjj$, which corresponds to the $n = 5$ with $P = S_2 \times S_3$ case. Here, we restrict to the case of $\mathbb{C}[y_{ij}, i < j \le 5]^{S_2 \times S_3}$.

The set of primaries was already given in Equation 4.1 of Subsection 4.1.1. Comparing to the Hilbert series in Table A.1, we see that they are non-optimal and we need to write the Hilbert series in the modified form

$$H(\mathbb{C}[y_{ij}, i < j \le 5]^{S_2 \times S_3}, t) =$$

$$\frac{1+3t^2+6t^3+12t^4+17t^5+32t^6+35t^7+47t^8+48t^9+49t^{10}+38t^{11}+34t^{12}+19t^{13}+12t^{14}+5t^{15}+2t^{16}}{(1-t)^3(1-t^2)^2(1-t^3)^2(1-t^4)(1-t^5)(1-t^6)}.$$

Using the algorithm to find the secondaries up to degree $\frac{3}{4}(2!)(3!) = 9$ (we can now see the full power of the degree bounds) and eliminating redundancies, we are left with a set of 26 minimal algebra generators. Table 4.2 contains the explicit list.

Degree = 1
$g_1 = y_{12}$, $g_2 = y_{34} + y_{35} + y_{45}$, $g_3 = y_{13} + y_{14} + y_{15} + y_{23} + y_{24} + y_{25}$,
Degree = 2
$g_4 = y_{13}y_{23} + y_{14}y_{24} + y_{15}y_{25}$, $g_5 = y_{34}y_{35} + y_{34}y_{45} + y_{35}y_{45}$, $g_6 = y_{13}y_{14} + y_{13}y_{15} + y_{14}y_{15} + y_{23}y_{24} + y_{23}y_{25} + y_{24}y_{25}$, $g_7 = y_{13}y_{34} + y_{14}y_{34} + y_{23}y_{34} + y_{24}y_{34} + y_{13}y_{35} + y_{15}y_{35} + y_{23}y_{35} + y_{25}y_{35} + y_{14}y_{45} + y_{15}y_{45} + y_{24}y_{45} + y_{25}y_{45}$, $g_8 = y_{13}y_{14} + y_{13}y_{15} + y_{14}y_{15} + y_{13}y_{23} + y_{14}y_{23} + y_{15}y_{23} + y_{13}y_{24} + y_{14}y_{24} + y_{15}y_{24} + y_{23}y_{24} + y_{13}y_{25} + y_{14}y_{25} + y_{15}y_{25} + y_{23}y_{25} + y_{24}y_{25}$,
Degree = 3
$g_9 = y_{34}y_{35}y_{45}$, $g_{10} = y_{13}y_{23}y_{34} + y_{14}y_{24}y_{34} + y_{13}y_{23}y_{35} + y_{15}y_{25}y_{35} + y_{14}y_{24}y_{45} + y_{15}y_{25}y_{45}$, $g_{11} = y_{13}y_{14}y_{34} + y_{23}y_{24}y_{34} + y_{13}y_{15}y_{35} + y_{23}y_{25}y_{35} + y_{14}y_{15}y_{45} + y_{24}y_{25}y_{45}$, $g_{12} = y_{13}y_{34}^2 + y_{14}y_{34}^2 + y_{23}y_{34}^2 + y_{24}y_{34}^2 + y_{13}y_{35}^2 + y_{15}y_{35}^2 + y_{23}y_{35}^2 + y_{25}y_{35}^2 + y_{14}y_{45}^2 + y_{15}y_{45}^2 + y_{24}y_{45}^2 + y_{25}y_{45}^2$, $g_{13} = y_{13}^2y_{34} + y_{14}^2y_{34} + y_{23}^2y_{34} + y_{24}^2y_{34} + y_{13}^2y_{35} + y_{15}^2y_{35} + y_{23}^2y_{35} + y_{25}^2y_{35} + y_{14}^2y_{45} + y_{15}^2y_{45} + y_{24}^2y_{45} + y_{25}^2y_{45}$, $g_{14} = y_{13}^2y_{23} + y_{13}y_{23}^2 + y_{14}^2y_{24} + y_{14}y_{24}^2 + y_{15}^2y_{25} + y_{15}y_{25}^2$, $g_{15} = y_{13}^2y_{14} + y_{13}y_{14}^2 + y_{13}^2y_{15} + y_{14}^2y_{15} + y_{13}y_{15}^2 + y_{14}y_{15}^2 + y_{23}^2y_{24} + y_{23}y_{24}^2 + y_{23}^2y_{25} + y_{24}^2y_{25} + y_{23}y_{25}^2 + y_{24}y_{25}^2$, $g_{16} = y_{13}y_{14}y_{15} + y_{13}y_{14}y_{23} + y_{13}y_{15}y_{23} + y_{14}y_{15}y_{23} + y_{13}y_{14}y_{24} + y_{13}y_{15}y_{24} + y_{14}y_{15}y_{24} + y_{13}y_{23}y_{24} + y_{14}y_{23}y_{24} + y_{15}y_{23}y_{24} + y_{13}y_{14}y_{25} + y_{13}y_{15}y_{25} + y_{14}y_{15}y_{25} + y_{13}y_{23}y_{25} + y_{14}y_{23}y_{25} + y_{15}y_{23}y_{25} + y_{13}y_{24}y_{25} + y_{14}y_{24}y_{25} + y_{15}y_{24}y_{25} + y_{23}y_{24}y_{25}$,
Degree = 4
$g_{17} = y_{13}^2y_{23}^2 + y_{14}^2y_{24}^2 + y_{15}^2y_{25}^2$,

$g_{18} = y_{13}y_{23}y_{34}^2 + y_{14}y_{24}y_{34}^2 + y_{13}y_{23}y_{35}^2 + y_{15}y_{25}y_{35}^2 + y_{14}y_{24}y_{45}^2 + y_{15}y_{25}y_{45}^2,$ $g_{19} = y_{13}y_{14}y_{34}^2 + y_{23}y_{24}y_{34}^2 + y_{13}y_{15}y_{35}^2 + y_{23}y_{25}y_{35}^2 + y_{14}y_{15}y_{45}^2 + y_{24}y_{25}y_{45}^2,$ $g_{20} = y_{13}^2y_{23}y_{34} + y_{13}y_{23}^2y_{34} + y_{14}^2y_{24}y_{34} + y_{14}y_{24}^2y_{34} + y_{13}^2y_{23}y_{35} + y_{13}y_{23}^2y_{35} + y_{15}^2y_{25}y_{35} + y_{15}y_{25}^2y_{35} + y_{14}^2y_{24}y_{45} + y_{14}y_{24}^2y_{45} + y_{15}^2y_{25}y_{45} + y_{15}y_{25}^2y_{45},$ $g_{21} = y_{13}^2y_{15}y_{34} + y_{14}^2y_{15}y_{34} + y_{23}^2y_{25}y_{34} + y_{24}^2y_{25}y_{34} + y_{13}^2y_{14}y_{35} + y_{14}y_{15}^2y_{35} + y_{23}^2y_{24}y_{35} + y_{24}y_{25}^2y_{35} + y_{13}y_{14}^2y_{45} + y_{13}y_{15}^2y_{45} + y_{23}y_{24}^2y_{45} + y_{23}y_{25}^2y_{45},$ $g_{22} = y_{13}^2y_{14}y_{23} + y_{13}^2y_{15}y_{23} + y_{13}y_{14}^2y_{24} + y_{14}^2y_{15}y_{24} + y_{13}y_{23}^2y_{24} + y_{14}y_{23}y_{24}^2 + y_{13}y_{15}^2y_{25} + y_{14}y_{15}^2y_{25} + y_{13}y_{23}^2y_{25} + y_{14}y_{24}^2y_{25} + y_{15}y_{23}y_{25}^2 + y_{15}y_{24}y_{25}^2,$ $g_{23} = y_{13}y_{14}y_{15}y_{23} + y_{13}y_{14}y_{15}y_{24} + y_{13}y_{14}y_{23}y_{24} + y_{13}y_{15}y_{23}y_{24} + y_{14}y_{15}y_{23}y_{24} + y_{13}y_{14}y_{15}y_{25} + y_{13}y_{14}y_{23}y_{25} + y_{13}y_{15}y_{23}y_{25} + y_{14}y_{15}y_{23}y_{25} + y_{13}y_{14}y_{24}y_{25} + y_{13}y_{15}y_{24}y_{25} + y_{14}y_{15}y_{24}y_{25} + y_{13}y_{23}y_{24}y_{25} + y_{14}y_{23}y_{24}y_{25} + y_{15}y_{23}y_{24}y_{25},$
Degree $= 5$
$g_{24} = y_{13}^2y_{23}^2y_{34} + y_{14}^2y_{24}^2y_{34} + y_{13}^2y_{23}^2y_{35} + y_{15}^2y_{25}^2y_{35} + y_{14}^2y_{24}^2y_{45} + y_{15}^2y_{25}^2y_{45}$ $g_{25} = y_{13}y_{14}y_{15}y_{23}y_{24} + y_{13}y_{14}y_{15}y_{23}y_{25} + y_{13}y_{14}y_{15}y_{24}y_{25} + y_{13}y_{14}y_{23}y_{24}y_{25} + y_{13}y_{15}y_{23}y_{24}y_{25} + y_{14}y_{15}y_{23}y_{24}y_{25},$
Degree $= 6$
$g_{26} = y_{13}y_{14}y_{15}y_{23}y_{24}y_{25}.$

Table 4.2 Table of minimal algebra generators for $\mathbb{C}[y_{ij}, i < j \leq 5]^{S_2 \times S_3}$.

4.2 Hironaka decompositions

In the previous Section, we generalized the FFT of the orthogonal and Euclidean groups to include the action of an arbitrary group of permutations of the n particles and provided a systematic method of constructing a set of minimal algebra generators of Lorentz- (or Euclidean-) and permutation-invariant polynomials. A major difference is that, even when $n \leq d$, the algebra of invariants is not a polynomial algebra once we include permutations. This simple observation already suggests that attempts to generalise the SFT to the case where permutations are included will lead to unpleasantness.

In this Section, we replace the FFT and SFT by a more direct description of the algebra of Lorentz- (or Euclidean-) and permutation-invariants, using tools of commutative algebra and invariant theory which were not available to Weyl. In particular, we use the fact that (via a theorem of Hochster and Roberts [24]) the algebras of invariants are Cohen-Macaulay, and so admit Hironaka decompositions as free, finitely-generated modules over polynomial subalgebras. Thus, a direct description of the invariant algebras can be given in terms of a set of generators of such polynomial subalgebras, termed either a *homogeneous system of parameters* (HSOP) or *primaries*, and a set of basis elements for the modules, called *secondaries*. In particular, every element in the algebras can be expressed *uniquely* in terms of primaries and secondaries, and multiplication in the algebras is completely encoded in the finite set of products of secondaries.

Again, the difficulty is in finding these Hironaka decompositions explicitly. In the what follows, we employ the necessary results of invariant theory to find Hironaka decompositions in the trivial cases ($n \leq d$ for $\mathbb{C}[V^n]^{O(d)\times P}$ and $n \leq d{+}1$ for $\mathbb{C}[V^n]^{E(d)\times P}$) and in the first non-trivial case ($n = d{+}1$ for $\mathbb{C}[V^n]^{O(d)\times P}$ and $n = d{+}2$ for $\mathbb{C}[V^n]^{E(d)\times P}$). We solve the hardest step in the procedure, namely to find HSOPs. Unfortunately, even though the remaining step of finding the secondaries reduces to a conceptually straightforward exercise in linear algebra, the available algorithm proceeds by brute-force Grörbner basis methods [11] and runs out of steam in cases with more than a few particles (vectors). But we hope that our results, modest though they are, will inspire others to make more targeted attacks on the problem. In Subsection 4.2.5, we present Hironaka decompositions of $\mathbb{C}[V^n]^{O(d)\times P}$ in the cases with $(n, d) = (5, 4)$ with no permutations included, $(n, d) = (3, 2)$ with all permutations included, and a conjecture for the Hironaka decomposition of the first non-trivial case with no permutations. Appendix A gives the details of the relevant Hilbert series computations.

4.2.1 Technical statement of results

In Section 4.1, we generalised Weyl's FFT to include permutation invariance. That is, we constructed a general method for finding a set of generators of the algebra $\mathbb{C}[y_{ij}]^P$ which surjects onto the Lorentz- (or Euclidean-) and permutation-invariant subalgebra, $W| : \mathbb{C}[y_{ij}]^P \twoheadrightarrow \mathbb{C}[V^n]^{O(d)\times P}$ (or $\phi|| \circ W| : \mathbb{C}[y_{ij}]^P \twoheadrightarrow \mathbb{C}[V^n]^{E(d)\times P}$). What our work did not include is the generalisation of the SFT which amounts to characterising the kernel of the restriction map $\ker W|$ (or $\ker \phi|| \circ W|$). Formally, the kernel of $W|$ is the intersection of the invariant algebra with the ideal $I \subset \mathbb{C}[y_{ij}]$ generated by the $(d + 1)$-minors of the matrix whose ij-th entry is y_{ij} for $i \leq j$ and y_{ji} for $i > j$, *i.e.* $\ker W| = I \cap \mathbb{C}[y_{ij}]^P$ (the situation is more complicated for the Euclidean- and permutation-invariants). In practice however, it is difficult to explicitly describe $\ker W|$, for a couple of reasons. For one thing, as stated previously, whereas $\mathbb{C}[y_{ij}]$ is a polynomial algebra, the invariant algebra $\mathbb{C}[y_{ij}]^P$ has a more complicated structure in general: it is Cohen-Macaulay and therefore can be expressed as a free, finitely-generated algebra over a polynomial subalgebra. For another, it turns out that the generators of the ideal I transform in an unpleasant representation of the permutation group, making finding the corresponding permutation-invariant generators difficult.

We therefore follow an alternative approach, seeking a more direct description of the Lorentz- (or Euclidean-) and permutation-invariant algebra $\mathbb{C}[V^n]^{O(d)\times P}$ ($\mathbb{C}[V^n]^{E(d)\times P}$). The *Hochster-Roberts* theorem [24] states that an invariant algebra $K[V]^G$ is Cohen-Macaulay if G is a linearly reductive group. Since $O(d) \times P$ is linearly reductive, the theorem applies and the algebra $\mathbb{C}[V^n]^{O(d)\times P}$ can be expressed as a free, finitely-generated module over a polynomial subalgebra. That is, the algebra can be expressed in terms of a Hironaka decomposition as $\mathbb{C}[V^n]^{O(d)\times P} = \bigoplus_k \eta_k \mathbb{C}[\theta_l]$ where the $\{\eta_k\}$ are the secondaries, the $\{\theta_l\}$ form a HSOP, and multiplication in the algebra is uniquely defined via $\eta_k\eta_m = \sum_j f^j_{km}\eta_j$, with $f^j_{km} \in \mathbb{C}[\theta_l]$. Every element in the algebra is then uniquely expressed as a linear sum of secondaries with coefficients which are polynomials in the HSOP. Furthermore, we can map this Hironaka decomposition under the isomorphism $\phi^*|| : \mathbb{C}[V^{n-1}]^{O(d)\times P} \xrightarrow{\sim} \mathbb{C}[V^n]^{E(d)\times P}$ when $P \subset S_n$ (but *not* $P = S_n$) to obtain Hironaka decompositions of the Euclidean- and permutation-invariant algebra.

The difficult part of finding Hironaka decompositions begins when one tries to find valid HSOPs as, apart from using inefficient algorithms [7], there is no obvious way to obtain them. Furthermore, the properties that a valid HSOP needs to satisfy are non-trivial and difficult to check. Previously, we were able to sidestep this by repurposing Gauss's results on permutation-invariants. Here, we are not so lucky. In Subsection 4.2.2, we propose HSOPs for the algebras $\mathbb{C}[V^n]^{O(d)\times P}$, with $n = d+1$, in the two cases where $P = 1$ (with 1 denoting the trivial group) and $P = S_n$ and explicitly verify that they satisfy the necessary conditions.

4.2.2 HSOPs for $\mathbb{C}[V^n]^{O(d)\times P}$ and $\mathbb{C}[V^n]^{E(d)\times P}$

In this Subsection, we find HSOPs for the algebras $\mathbb{C}[V^n]^{O(d)\times P}$ in the $n \leq d$ case for any $P \subseteq S_n$ and in the $n = d+1$ case with no permutation symmetry, $P = 1$, and with full permutation symmetry, $P = S_n$. In fact, the HSOP for $P = S_n$ is also a HSOP for any $P \subset S_n$, and so we obtain a complete solution of this part of the problem. Since we also have an isomorphism between $\mathbb{C}[V^{n-1}]^{O(d)\times P}$ and $\mathbb{C}[V^n]^{E(d)\times P}$, we additionally obtain Hironaka decompositions for the Euclidean invariant algebras for $n \leq d+2$. Crucially though, this only works for $P \subset S_n$ (and not $P = S_n$, since the action of the permutation group on the algebras is different here).

The necessary conditions for a set of polynomials to constitute a HSOP are twofold: firstly, the polynomials must be algebraically independent; secondly, they must satisfy the nullcone condition [11].

A set of polynomials $f_1, \ldots, f_m \in K[x_1, \ldots, x_k]$ is said to be *algebraically independent* if the only polynomial $h \in K[z_1, \ldots, z_m]$ satisfying $h(f_1, \ldots, f_m) = 0$ is the zero polynomial. Although trivial to define, the algebraic independence of polynomials is less trivial to check. One method proceeds via calculation of a Gröbner basis, while another uses the Jacobi criterion. The former quickly becomes inefficient when used with many polynomials of high degree, but more importantly, it is difficult to apply in an abstract way. We therefore resort to using the Jacobi criterion[7] which states that a set of polynomials, $f_1, \ldots, f_m \in K[x_1, \ldots, x_k]$, is algebraically independent if and only if the wedge product of the exterior derivatives[8] of the polynomials is non-zero, *i.e.*

$$df_1 \wedge \cdots \wedge df_m \neq 0.$$

As regards the nullcone, a set of polynomials, $\{f_1, \ldots, f_m\}$, is said to satisfy the nullcone condition if the vanishing locus of all of its constituent polynomials coincides with $N_V = \{v \in V |\ f(v) = 0, \forall f \in K[V]^G_+\}$. We remark that, in the case of the Lorentz- and permutation-invariant algebra $\mathbb{C}[V^n]^{O(d)\times P}$, the nullcone is the set $\{v_i \cdot v_j = 0, \forall i \leq j\}$. The fact that this does not depend on the choice of P will prove to be important when we come to construct a HSOP for arbitrary P.

HSOPs in the trivial case ($n \leq d$)

The case when $n \leq d$ is relatively straightforward. Here, Weyl's SFT tells us that there are no relations between the dot products (the kernel $\ker W$ is trivial) and so the map $W|$ is actually an isomorphism, $W| : \mathbb{C}[y_{ij}]^P \xrightarrow{\sim} \mathbb{C}[V^n]^{O(d)\times P}$. Therefore, HSOPs for $\mathbb{C}[V^n]^{O(d)\times P}$ can be obtained by mapping the HSOPs found for $\mathbb{C}[y_{ij}]^P$ using the methods discussed in Subsection 4.1.1 under $W|$. We consequently have the following

Theorem 4.2.1. *A HSOP of $\mathbb{C}[V^n]^{O(d)\times P}$ when $n \leq d$ can be obtained by partitioning the $v_i \cdot v_j$ into their orbits under P and forming the respective sets of elementary symmetric polynomials on them.*

Via the isomorphism $\phi^*||$, we further have the following

[7]For a proof of the Jacobi criterion, see for example [14] or [4].

[8]The definition of a derivative requires some care for fields where limits are not defined [4], but here we will only need to consider the case $K = \mathbb{C}$.

Corollary 4.2.1. *A HSOP of* $\mathbb{C}[V^n]^{E(d)\times P}$ *for* $P \subset S_n$ *(but not* $P = S_n$*) when* $n \leq d+1$ *can be obtained by mapping the HSOPs of* $\mathbb{C}[V^{n-1}]^{O(d)\times P}$ *under* $\phi^*|| : v_i \cdot v_j \mapsto (v_i - v_{avg}) \cdot (v_j - v_{avg})$*, where* $v_{avg} = \frac{1}{n}\sum_{i=1}^{n} v_i$.

A HSOP in the first non-trivial case ($n = d+1$) with $P = 1$

For the first non-trivial case, let us warm up by considering the case without permutations. With $n = d+1$, the SFT tells us that the relations between the dot products $v_i \cdot v_j$ are generated by the image of a single element under the Weyl map W, namely the determinant of the matrix whose ij-th entry is y_{ij} for $i \leq j$ and y_{ji} for $i > j$. Thus, $W(\det(y_{ij})) = \det(v_i \cdot v_j) = 0$ where $\det(v_i \cdot v_j) \in \mathbb{C}[V^n]^{O(d)}$. This will be important for proving that our proposed HSOP satisfies the nullcone condition. We now get the following

Theorem 4.2.2. *A HSOP for the algebra* $\mathbb{C}[V^n]^{O(d)}$*, with* $n = d+1$*, is given by the* $d(d+3)/2$ *polynomials*

$$\begin{aligned} \theta_i &= v_1 \cdot v_1 + v_i \cdot v_i, \quad 2 \leq i \leq d+1, \\ \alpha_{ij} &= v_i \cdot v_j, \qquad\qquad 1 \leq i < j \leq d+1. \end{aligned} \tag{4.2}$$

Proof. We first check that these polynomials satisfy the nullcone condition. Evidently, if all dot products vanish, then both θ_i and α_{ij} vanish. Proceeding in the other direction, suppose that θ_i and α_{ij} vanish. The vanishing of α_{ij} implies not only the vanishing of the dot products with $i < j$, but also implies, together with the vanishing determinant relation, that $\prod_{i=1}^{d+1}(v_i \cdot v_i) = 0$. So either $(v_1 \cdot v_1) = 0$ or $(v_k \cdot v_k) = 0$ for some $2 \leq k \leq d+1$. If the former, then the fact that $\theta_i = 0$ implies $v_i \cdot v_i = 0$. If the latter, then $\theta_k = 0$ implies $v_1 \cdot v_1 = 0$, while the vanishing of the other θ_i implies the vanishing of all other $v_i \cdot v_i$ with $i \neq 1, k$. Either way, all dot products vanish and the nullcone condition is satisfied.

To prove algebraic independence, it is sufficient to show that the wedge product of the exterior derivatives of θ_i and α_{ij} is non-zero on at least a single point. We choose to evaluate the wedge product at the point

$$\begin{aligned} p_1 &= (0, 0, 0, \ldots, 0), \\ p_2 &= (1, 0, 0, \ldots, 0), \\ p_3 &= (0, 1, 0, \ldots, 0), \\ &\vdots \end{aligned}$$

$$p_{d+1} = (0, 0, \dots, 0, 1),$$

where the unit entry moves progressively along, as indicated. We claim that the component of the wedge product proportional to

$$\omega = dv_1^1 \wedge \dots \wedge dv_1^d \wedge dv_2^1 \wedge \dots \wedge dv_2^d \wedge dv_3^2 \wedge \dots \wedge dv_3^d \wedge dv_4^3 \wedge \dots \wedge dv_4^d \wedge \dots \wedge dv_d^{d-1} \wedge dv_d^d \wedge dv_{d+1}^d,$$

has coefficient at this point given by $2^d \neq 0$ (up to an irrelevant minus sign) and so the Jacobi criterion is satisfied. To establish the claim in detail, one starts by showing that the only non-zero contribution to the wedge product is

$$\begin{aligned} &d(v_2 \cdot v_2) \wedge \dots \wedge d(v_{d+1} \cdot v_{d+1}) \wedge d(v_1 \cdot v_2) \wedge \dots \wedge d(v_1 \cdot v_{d+1}) \\ &\qquad \wedge d(v_2 \cdot v_3) \wedge \dots \wedge d(v_2 \cdot v_{d+1}) \wedge d(v_3 \cdot v_4) \wedge \dots \wedge d(v_d \cdot v_{d+1}), \end{aligned}$$

as contributions with more than one $d(v_1 \cdot v_1)$ vanish trivially and contributions with a single $d(v_1 \cdot v_1)$ vanish on the specified point as $d(v_1 \cdot v_1)|_{v_i = p_i} = 2 \sum_i p_1^i dv_1^i = 0$ there. Now, the coefficient of the component proportional to ω can be thought of as the determinant of an associated matrix.[9] In that form, after some row and column swaps (hence the irrelevant minus sign), one can show that the coefficient is the determinant of a diagonal matrix whose entries are all 1's except for d instances of 2's which come from the $d(v_i \cdot v_i)|_{v_i = p_i} = 2 \sum_j p_i^j dv_i^j = 2dv_i^{i-1}$, for $2 \leq i \leq d+1$. Hence, the coefficient of ω is 2^d as claimed.

Therefore, the proposed set of polynomials θ_i and α_{ij} satisfies the algebraic independence and nullcone conditions and so constitutes a valid HSOP. $\square$

Combining the above theorem with the isomorphism $\phi^*| : \mathbb{C}[V^n]^{O(d)} \xrightarrow{\sim} \mathbb{C}[V^{n+1}]^{E(d)}$, we therefore have the following corollary

Corollary 4.2.2. *A HSOP for the algebra $\mathbb{C}[V^n]^{E(d)}$, with $n = d+2$, is given by the $d(d+3)/2$ polynomials*

$$\begin{aligned} \theta_i &= (v_1 - v_{avg}) \cdot (v_1 - v_{avg}) + (v_i - v_{avg}) \cdot (v_i - v_{avg}), && 2 \leq i \leq d+1, \\ \alpha_{ij} &= (v_i - v_{avg}) \cdot (v_j - v_{avg}), && 1 \leq i < j \leq d+1. \end{aligned}$$

[9] Explicitly, it is the matrix with ij-th entry being $\frac{\partial f_i}{\partial x_j}$ where $f_i \in \{v_r \cdot v_s \mid r \leq s,\ s \neq 1\}$ and $x_j \in \{v_1^1, \dots, v_1^d, v_2^1, \dots, v_2^d, v_3^2, \dots, v_3^d, v_4^3, \dots, v_4^d, \dots, v_d^{d-1}, v_d^d, v_{d+1}^d\}$.

A HSOP in the first non-trivial case ($n = d+1$) with $P = S_n$

We now move on to the full-permutation case. Previously, in the $n = d+1$ case with no permutations, the SFT indicated that the relations between the dot products are generated by a single element, $\det(v_i \cdot v_j) \in \mathbb{C}[V^n]^{O(d)}$, where $\det(v_i \cdot v_j) = 0$. Here, we consider the full permutations case and work in the permutation-invariant subalgebra $\mathbb{C}[V^n]^{O(d)\times S_n}$. But, since the determinant relation, $\det(v_i \cdot v_j)$, is permutation-invariant, it is also an element of the permutation-invariant subalgebra, $\det(v_i \cdot v_j) \in \mathbb{C}[V^n]^{O(d)\times S_n}$, and therefore can be safely used in our proof of the validity of the HSOP for $P = S_n$.[10]

As to the HSOP itself, we take inspiration from Gauss who tells us that the m symmetric polynomials in m independent variables satisfy the necessary HSOP conditions. Therefore, the obvious candidates in our case are given by symmetric polynomials in the $d+1$ variables $v_i \cdot v_i$ and $d(d+1)/2$ variables $v_i \cdot v_j$ (with $i < j$), giving a total of $d(d+3)/2+1$. But, these cannot satisfy the algebraic independence condition since the dot products are not independent variables. It is therefore reasonable to suppose that in order to fix this, we need to judiciously discard one symmetric polynomial from this set to obtain a valid HSOP. As we will see, taking the power sum polynomials and discarding the highest degree polynomial in $v_i \cdot v_i$ does the job. In fact, taking any set of symmetric polynomials (elementary[11] or complete homogeneous) and discarding the highest degree polynomial in $v_i \cdot v_i$ also does the job. This can be seen by using Newton's identities for the elementary symmetric polynomials or the equivalent relations for the complete homogeneous symmetric polynomials.[12] Indeed, we have the following

[10]More precisely, the kernel of the restriction map $\ker W| = \langle \det(v_i \cdot v_j)\rangle \cap \mathbb{C}[V^n]^{O(d)\times P} = \langle \det(v_i \cdot v_j)\rangle$ as an ideal of $\mathbb{C}[V^n]^{O(d)\times P}$.

[11]The k-th elementary symmetric polynomial, e_k, on the the variables $x_1, \ldots, x_n$ is defined as

$$e_k(x_1, \ldots, x_n) = \sum_{1 \le j_1 < j_2 < \cdots < j_k \le n} x_{j_1} \ldots x_{j_k}.$$

[12]Newton's identities relate the m-th power sum symmetric polynomial, Pow_m, to the first m elementary symmetric polynomials, e_i, via:

$$\mathrm{Pow}_m = \sum_{\substack{r_1+2r_2+\cdots+mr_m=m,\\ r_1\ge 0,\ldots,r_m\ge 0}} (-1)^m \frac{m(r_1+\cdots+r_m-1)!}{r_1!r_2!\ldots r_m!} \prod_{i=1}^{m} (-e_i)^{r_i}.$$

Equivalent relations relating the power sum symmetric polynomials to the complete homogeneous symmetric polynomials also exist.

Theorem 4.2.3. *A HSOP for the algebra $\mathbb{C}[V^n]^{O(d)\times S_n}$, with $n = d+1$, is given by the $d(d+3)/2$ permutation-invariant polynomials*

$$
\begin{aligned}
\theta_k &= \mathrm{Pow}_k(v_i \cdot v_i) := \sum_{i=1}^{d+1} (v_i \cdot v_i)^k, \quad 1 \le k \le d, \\
\alpha_k &= \mathrm{Pow}_k(v_i \cdot v_j) := \sum_{i<j}^{d+1} (v_i \cdot v_j)^k, \quad 1 \le k \le d(d+1)/2,
\end{aligned}
\tag{4.3}
$$

where Pow_k *is the k-th power symmetric polynomial.*

Proof. We first check that these polynomials satisfy the nullcone condition. Evidently, if all the dot products vanish, then both θ_k and α_k vanish. Proceeding in the other direction, suppose that θ_k and α_k vanish. Using Newton's identities, one can show that the vanishing of the first r power symmetric polynomials implies the vanishing of the first r elementary symmetric polynomials. Therefore, $\{\alpha_k = 0,\ \forall k\}$ implies the vanishing of all the elementary symmetric polynomials on the $v_i \cdot v_j, i < j$. Now, the vanishing of the highest degree elementary symmetric polynomial, $\prod_{i<j}^{n} v_i \cdot v_j = 0$, implies the vanishing of at least one $v_i \cdot v_j,\ i < j$. This then implies the vanishing of the $d(d+1)/2 - 1$ elementary symmetric polynomials on the remaining $d(d+1)/2 - 1$ dot products $v_i \cdot v_j, i < j$. Repeating this process recursively, one sees that the vanishing of the α_k implies the vanishing of all $v_i \cdot v_j, i < j$. This result, combined together with $\det(v_i \cdot v_j) = 0$, implies that $\prod_{i=1}^{d+1}(v_i \cdot v_i) = 0$. But $\prod_{i=1}^{d+1}(v_i \cdot v_i)$ is the elementary symmetric polynomial of highest degree, so $\prod_{i=1}^{d+1}(v_i \cdot v_i) = 0$, together with the vanishing of the θ_k, implies the vanishing of all $d+1$ elementary symmetric polynomials in $v_i \cdot v_i$. From here, one can again recursively show that all $v_i \cdot v_i$ must vanish, so the nullcone condition is satisfied.

To prove algebraic independence, we evaluate (a component of) the wedge product of the exterior derivatives of θ_k and α_k at the point

$$
\begin{aligned}
p_1 &= (2, 0, \ldots, 0), \\
p_2 &= (3, 0, \ldots, 0), \\
p_3 &= (l_m, 1, 0, \ldots, 0), \\
p_4 &= (l_{m+1}, 0, 1, 0, \ldots, 0), \\
&\vdots \\
p_i &= (l_{m+i-3}, 0, \ldots, 0, 1, 0, \ldots, 0), \\
&\vdots
\end{aligned}
$$

$$p_{d+1} = (l_{m+d-2}, 0, \ldots, 0, 1),$$

where l_i denotes the i-th prime number (with $l_1 = 2$) and $m \geq 3$ and where the unit entry moves progressively along, as indicated. The prime numbers will prove useful soon when we require that the dot products $v_i \cdot v_j$ are all distinct. We claim that the component of the wedge product proportional to

$$\omega = dv_1^1 \wedge \cdots \wedge dv_1^d \wedge dv_2^1 \wedge \cdots \wedge dv_2^d \wedge dv_3^2 \wedge \cdots \wedge dv_3^d \wedge dv_4^3 \wedge \cdots \wedge dv_4^d \wedge \cdots \wedge dv_d^{d-1} \wedge dv_d^d \wedge dv_{d+1}^d,$$

has a non-zero coefficient. To establish this claim in detail, we first note that the wedge product can be re-expressed as

$$d!\left(\frac{d(d+1)}{2}\right)!\det(M)\sum_{k=1}^{d+1}\det(L_k)\Omega_k,$$

where M is the Vandermonde matrix[13] on the $p_i \cdot p_j, i < j$, L_k is the Vandermonde matrix on the $p_i \cdot p_i, i \neq k$, and

$$\Omega_k = d(v_1 \cdot v_1) \wedge \cdots \wedge \widehat{d(v_k \cdot v_k)} \wedge \cdots \wedge d(v_{d+1} \cdot v_{d+1}) \wedge d(v_1 \cdot v_2) \wedge d(v_1 \cdot v_3) \wedge \cdots \wedge d(v_d \cdot v_{d+1}),$$

where $\widehat{\ }$ over a term indicates that that term should be omitted. By considering the coefficient of the component proportional to ω of Ω_k as the determinant of an associated matrix[14], one can show that the only contributions to the sum on the specified point come from the instances with $k = 1, 2$. Therefore, the coefficient of the component proportional to ω of the wedge product is

$$2^d d!\left(\frac{d(d+1)}{2}\right)!\det(M)\left(9\det(L_1) - 4\det(L_2)\right),$$

up to an irrelevant overall minus sign (from row and column swaps). The $\det(M)$ term is non-zero as every dot product $p_i \cdot p_j$, $i < j$, is distinct (our use of prime numbers guarantees that $l_i l_j = l_m l_n$ if and only if either $l_i = l_n$ and $l_j = l_m$ or $l_i = l_m$ and

[13] The Vandermonde matrix V on a set of variables x_i, $i \in \{1, \ldots, n\}$, is the $n \times n$ matrix with entries $V_{ij} = x_i^{j-1}$. The determinant of this matrix can be nicely expressed as $\det(V) = \prod_{1 \leq i < j \leq n}(x_i - x_j)$ and is non-zero only if all the x_i's are distinct.

[14] Explicitly, it is the matrix with ij-th entry being $\frac{\partial f_i}{\partial x_j}$ where $f_i \in \{v_r \cdot v_s \mid r \leq s,\ s \neq k\}$ and $x_j \in \{v_1^1, \ldots, v_1^d, v_2^1, \ldots, v_2^d, v_3^2, \ldots, v_3^d, v_4^3, \ldots, v_4^d, \ldots, v_d^{d-1}, v_d^d, v_{d+1}^d\}$.

$l_j = l_n$). To show the last term is non zero, we expand it as

$$\begin{aligned}(9\det(L_1) - 4\det(L_2)) &= 9\prod_{i<j\neq 1}^{d+1}(p_i\cdot p_i - p_j\cdot p_j) - 4\prod_{i<j\neq 2}^{d+1}(p_i\cdot p_i - p_j\cdot p_j)\\ &= \prod_{i<j\neq 1,2}^{d+1}(p_i\cdot p_i - p_j\cdot p_j)\left(9\prod_{i=3}^{d+1}(p_2\cdot p_2 - p_i\cdot p_i) - 4\prod_{i=3}^{d+1}(p_1\cdot p_1 - p_i\cdot p_i)\right).\end{aligned}$$

Since we have the freedom to choose m to be as large as we want (there are infinitely many primes), we can see that this term is non-zero as follows: for large m, where $p_i \cdot p_i \gg p_1 \cdot p_1$, $p_2 \cdot p_2$, it tends to $\sim 5\prod_{i=3}^{d+1}(p_i\cdot p_i)\prod_{i<j\neq 1,2}^{d+1}(p_i\cdot p_i - p_j\cdot p_j)$ which is non-zero as the $p_i \cdot p_i$ are non-zero and distinct (again by our use of the prime numbers). Therefore, the Jacobi criterion is satisfied.

Hence, these polynomials are algebraically independent and satisfy the nullcone condition and so constitute a valid HSOP. □

As we have already remarked, the nullcone of $\mathbb{C}[V^n]^{O(d)\times P}$, being given by the vanishing locus of the dot products $v_i \cdot v_j$, is independent of the choice of P. Moreover, an algebraically independent set of $O(d,\mathbb{C}) \times S_n$-invariant polynomials is also an algebraically independent set of $O(d,\mathbb{C}) \times P$-invariant polynomials, for any $P \subset S_n$. We thus have the important

Corollary 4.2.3. *A HSOP for the algebra $\mathbb{C}[V^n]^{O(d)\times P}$, with $n = d+1$, is given by the permutation-invariant polynomials in Equation 4.3, for any $P \subset S_n$.*

As we shall see, this gives us a starting point for finding a Hironaka decomposition for any P in the case $n = d+1$. We can also use Corollary 4.2.3 to obtain HSOPs for the Euclidean- and permutation-invariant algebra via the isomorphism $\phi^*|| : \mathbb{C}[V^{n-1}]^{O(d)\times P} \xrightarrow{\sim} \mathbb{C}[V^n]^{E(d)\times P}$. Unfortunately though, since the action of P on the algebra $\mathbb{C}[V^{n-1}]^{O(d)\times P}$ here is via the permutation representation for $P \subset S_n$ (and not $P = S_n$) only, we can only map this Hironaka decomposition to the Euclidean- and permutation-invariant algebra in those cases. And so, we have the following

Corollary 4.2.4. *A HSOP for the algebra $\mathbb{C}[V^n]^{E(d)\times P}$, with $n = d+2$ and $P \subset S_n$ (and not $P = S_n$), is given by the $d(d+3)/2$ permutation-invariant polynomials*

$$\begin{aligned}\theta_k &= \mathrm{Pow}_k((v_i - v_{avg})\cdot(v_i - v_{avg})) := \sum_{i=1}^{d+1}\left((v_i - v_{avg})\cdot(v_i - v_{avg})\right)^k, \quad 1 \le k \le d,\\ \alpha_k &= \mathrm{Pow}_k((v_i - v_{avg})\cdot(v_j - v_{avg})) := \sum_{i<j}^{d+1}\left((v_i - v_{avg})\cdot(v_j - v_{avg})\right)^k, \quad 1 \le k \le d(d+1)/2,\end{aligned} \tag{4.4}$$

where Pow_k *is the* k*-th power symmetric polynomial.*

A remark on HSOPs for $n \geq d+2$

It would obviously be desirable to generalise our methods to cases with $n \geq d+2$. The first obstacle in doing so is that the relations between the dot products $v_i \cdot v_j$ given by the higher minors of the matrix whose entries are $v_i \cdot v_j$, are not S_n-invariant. Thus, they do not belong to $\mathbb{C}[V^n]^{O(d)\times S_n}$ and cannot be used directly in the proofs. To overcome this, one presumably needs to first find a set of invariant polynomials which generate the relations and then work with these. But it is not clear to us what form a HSOP might take.

4.2.3 Secondaries

Now that we can write down HSOPs of our invariant algebras at will in cases with $n \leq d+1$, the corresponding secondaries may be computed via an algorithm (which can be found in [11]). Here, we illustrate the algorithm by applying it to a simple example, namely $(n,d) = (3,2)$ with no permutation symmetry, *i.e.* the algebra $\mathbb{C}[V^3]^{O(2)}$, with $V \cong \mathbb{C}^6$.

The algorithm is based on the following two observations. Firstly, the number of secondaries required can be read off (along with their degrees) from the Hilbert series, which itself can be computed using standard methods from invariant theory (as we review in Appendix A). Indeed, given a Hironaka decomposition of an invariant algebra $K[V]^G = \bigoplus_i \eta_i K[\theta_j]$, its Hilbert series $H(K[V]^G, t)$, takes the form $\frac{1+\sum_{k=1} S_k t^k}{\prod_{l=1}(1-t^l)^{P_l}}$ where S_k is the number of secondary invariants η_i of degree k and P_l is the number of primary invariants (HSOP) θ_j of degree l. Therefore, given a HSOP, which fixes the P_l, one can read off the number and degrees of the secondaries from the numerator of the Hilbert series. Secondly, given a set of polynomial invariants $\{\eta_1, \dots, \eta_m\}$ of the right cardinality, the set forms the secondaries of the invariant algebra if and only if its constituent polynomials are linearly independent modulo the ideal $I := \langle \theta_1, \dots, \theta_r \rangle \in K[V]$ generated by the HSOP $\{\theta_i\}$. To show linear independence of a set of polynomials modulo an ideal, one can compute the remainders of the polynomials upon division by a Gröbner basis of that ideal and check that the remainders are themselves linearly independent [11].

Turning to our example, the methods described in Appendix A show that the Hilbert series is given by

$$H(\mathbb{C}[V^3]^{O(2)}, t) = \frac{1 + t^2 + t^4}{(1 - t^2)^5}.$$

Here we have written the series in a form such that the denominator reproduces the five primaries of degree 2 corresponding to the HSOP given in Equation 4.2, namely

$$\{(p \cdot p) + (q \cdot q), (p \cdot p) + (r \cdot r), (p \cdot q), (p \cdot r), (q \cdot r)\},$$

where we have labelled the momenta (vectors) by p, q, and r (we denote the corresponding components of V by $\{p_1, p_2, q_1, q_2, r_1, r_2\}$). We thus read off from the numerator that there is 1 secondary of degree 2 and 1 secondary of degree 4 (and of course the trivial secondary, 1, of degree 0).

The next step in the algorithm is to compute a Gröbner basis of the ideal generated by the HSOP, which will later be used to verify the linear independence of the secondaries. To do so, one must first choose a monomial ordering.[15] A common (and often very efficient) choice is graded reverse lexicographic order as defined in Section 2.2. In this ordering, a Gröbner basis of the ideal generated by our HSOP is given by the set of 20 polynomials

$$\{q_1r_1 + q_2r_2, p_1r_1 + p_2r_2, q_1^2 + q_2^2 - r_1^2 - r_2^2, p_1q_1 + p_2q_2, p_1^2 + p_2^2 + r_1^2 + r_2^2, p_2q_1r_2 - p_1q_2r_2, q_2^2r_1 - q_1q_2r_2 - r_1^3 - r_2^2r_1, p_2q_2r_1 - p_1q_2r_2, p_2^2r_1 - p_1p_2r_2 + r_1^3 + r_2^2r_1, -p_1q_2^2 + p_2q_1q_2 + p_1r_2^2 - p_2r_1r_2, p_2^2q_1 - p_1p_2q_2 + q_1r_2^2 - q_2r_1r_2, q_2r_1r_2^2 - q_1r_2^3, p_2r_1r_2^2 - p_1r_2^3, r_2r_1^3 + r_2^3r_1, q_2r_2^3 + q_2r_1^2r_2, p_2r_2^3 + p_2r_1^2r_2, r_1^4 - r_2^4, q_2r_1^3 + q_1r_2^3, p_2r_1^3 + p_1r_2^3, r_2^5 + r_1^2r_2^3\}.$$

We then proceed to generate a basis of homogeneous invariant polynomials in the algebra of degree d_i, corresponding to the degrees of secondaries read off of the Hilbert series, using linear algebra methods. If G were a finite group, this would be a simple matter of averaging all possible monomials of degree d_i over G to obtain a basis of invariant polynomials at that degree.[16] But for us G is infinite, so things are not so straightforward. We use the additional information that $\mathbb{C}[V^n]^{O(d)\times P} \subset \mathbb{C}[V^n]^{O(d)}$ and

[15] Readers seeking a gentle introduction to Gröbner basis methods may wish to consult [10].

[16] The average of a polynomial f over a finite group G is found by applying the Reynolds operator $\mathcal{R}_G(f) = \frac{1}{|G|}\sum_{g\in G} g \circ f$.

that by the FFT, $\mathbb{C}[V^n]^{O(d)}$ is generated by the set of dot products in the momenta.[17] This allows one to obtain a basis of homogeneous polynomials in $\mathbb{C}[V^n]^{O(d)\times P}$ of degree d_i by averaging all possible products of $d_i/2$ dot products over the (finite) permutation group P.

From this basis, we consecutively choose elements and compute their remainders upon division by the Gröbner basis (also called the normal forms) and keep them only if their remainders are non-zero and lie outside the $\mathbb{C}$-vector space generated by the remainders of previously found secondaries (*i.e.* the remainders are linearly independent). Once the required number of secondaries is obtained, one proceeds to the next degree and so on until all the secondaries have been found.

In our case (skipping over the trivial case of the secondary 1), we start at degree 2. Here, the basis of polynomials is just the set of dot products. Choosing $p \cdot p$, we compute the remainder upon division to be $-(r_1^2 + r_2^2)$, which is non-zero and so we have the required secondary of degree 2. We then move on to degree 4. Here, the basis of polynomials is all possible products of two dot products. We choose $(p \cdot p)^2$ and compute the remainder upon division to be $2r_2^2(r_1^2 + r_2^2)$, which is non-zero and is obviously linearly independent from the remainder of the previous secondary, since it does not have the same degree. We therefore have our required secondary of degree 4.[18] Finally, we obtain the Hironaka decomposition of the algebra as follows

$$\mathbb{C}[V^3]^{O(2)} = \Big(1 \oplus (p \cdot p) \oplus (p \cdot p)^2\Big) \cdot \mathbb{C}[p \cdot p + q \cdot q, p \cdot p + r \cdot r, p \cdot q, p \cdot r, q \cdot r]. \quad (4.5)$$

Simple though it is, our example already hints at the two bottlenecks that arise when computing the secondaries of $\mathbb{C}[V^n]^{O(d)\times P}$ in high dimensions with large permutation symmetry. One is the computation of the Gröbner basis of the ideal and the other is the computation of a basis of invariant polynomials of a certain degree, which becomes progressively more costly at higher degrees. There are multiple tricks which can be used to mitigate the latter bottleneck [11] (*e.g.*, using products of lower degree secondaries as candidates), but there is still no really effective way of tackling the inefficiency of the Gröbner basis computations.

In Subsection 4.2.5, we employ the algorithm to provide Hironaka decompositions for computationally tractable cases.

[17] Equivalently, $\mathbb{C}[V^n]^{E(d)\times P} \subset \mathbb{C}[V^n]^{E(d)}$ and by the FFT, $\mathbb{C}[V^n]^{E(d)}$ is generated by the set of dot products in the $v_i - v_{avg}$.

[18] We could have equally chosen either $(q \cdot q)$ or $(r \cdot r)$ for the degree 2 secondary and either $(q \cdot q)^2$ or $(r \cdot r)^2$ for the degree 4 one. It is also interesting to note that the remainders upon division by the Gröbner basis of $(p \cdot p)^3, (q \cdot q)^3$, and $(r \cdot r)^3$ are zero and so they lie *in* the ideal generated by the HSOP.

4.2.4 Parity

A significant advantage of characterising the Lorentz- (or Euclidean-) and permutation-invariant algebra directly via Hironaka decompositions is that the description can be readily extended to the case where parity is not a symmetry. The reason for this is that the algebras with and without parity as a symmetry, *i.e.* $\mathbb{C}[V^n]^{O(d)\times P}$ (or $\mathbb{C}[V^n]^{E(d)\times P}$) and $\mathbb{C}[V^n]^{SO(d)\times P}$ (or $\mathbb{C}[V^n]^{SE(d)\times P}$) respectively, have the same nullcone. This important fact can be traced back to existence of relations between the d-dimensional epsilon tensor contracted with the n momenta and the dot products (the square of a d-dimensional epsilon tensor is equal to the corresponding $d \times d$ subdeterminant of the matrix of dot products). Using these relations, one can readily show that the vanishing set of all the dot products then immediately implies the vanishing set of the epsilon tensors and consequently that the nullcones of the two algebras coincide.

Therefore, the HSOPs found in Subsection 4.2.2, which have been shown to satisfy the nullcone condition for $\mathbb{C}[V^n]^{O(d)\times P}$, also satisfy the nullcone condition for $\mathbb{C}[V^n]^{SO(d)\times P}$. Hence, we arrive at the following

Corollary 4.2.5. *A HSOP for the parity-non-invariant algebra* $\mathbb{C}[V^n]^{SO(d)\times P}$ *is given by the HSOPs for the parity-invariant algebra* $\mathbb{C}[V^n]^{O(d)\times P}$.

We also have an identical statement for the Euclidean- and permutation-invariant algebras $\mathbb{C}[V^n]^{SE(d)\times P}$.

Corollary 4.2.6. *A HSOP for the parity-non-invariant algebra* $\mathbb{C}[V^n]^{SE(d)\times P}$ *is given by the HSOPs for the parity-invariant algebra* $\mathbb{C}[V^n]^{E(d)\times P}$.

The only extra work needed to find the complete Hironaka decomposition for the algebras $\mathbb{C}[V^n]^{SO(d)\times P}$ (or $\mathbb{C}[V^n]^{SE(d)\times P}$) is in computing the secondaries. One now additionally needs to consider the epsilon tensors contracted with the momenta when constructing the basis of monomials and care must be taken to keep track of minus signs that appear as a result of the antisymmetric structure of the epsilon tensors when symmetrising under the permutation group.

4.2.5 Examples of Hironaka decompositions in $n = d + 1$

We now present two examples with explicit Hironaka decompositions using the above prescriptions. A list of Hironaka decompositions of multiple cases of physical interest can be found on-line.[19]

[19] Available at: https://github.com/WardHaddadin/Invariant-polynomials-and-machine-learning.

The case of $(n, d) = (5, 4)$ with $P = 1$

For the no permutation case with $(n, d) = (5, 4)$, we start by finding the HSOP for the algebra $\mathbb{C}[V^5]^{O(4)}$ in the way described in Subsection 4.2.2. This results in the following set of polynomials

$$\begin{aligned} \theta_i &= v_1 \cdot v_1 + v_i \cdot v_i, \quad 2 \leq i \leq 5, \\ \alpha_{ij} &= v_i \cdot v_j, \quad 1 \leq i < j \leq 5. \end{aligned}$$

Using the algorithm described in Subsection 4.2.3, we proceed to find the secondaries in a similar manner. The Hilbert series of the algebra, computed using methods described in Appendix A, is

$$H(\mathbb{C}[V^5]^{O(4)}, t) = \frac{1 + t^2 + t^4 + t^6 + t^8}{(1 - t^2)^{14}}.$$

We are therefore looking for 1 secondary at each of the degrees $0, 2, 4, 6$, and 8. We find that the following set of polynomials

$$1, (v_1 \cdot v_1), (v_1 \cdot v_1)^2, (v_1 \cdot v_1)^3, (v_1 \cdot v_1)^4,$$

have remainders upon division by the Gröbner basis of the ideal generated by the HSOP which are non-zero and linearly independent. Therefore, we obtain a Hironaka decomposition of the algebra as follows

$$\mathbb{C}[V^5]^{O(4)} = \Big(1 \oplus (v_1 \cdot v_1) \oplus (v_1 \cdot v_1)^2 \oplus (v_1 \cdot v_1)^3 \oplus (v_1 \cdot v_1)^4\Big) \cdot \mathbb{C}[\{\theta_i, \alpha_{ij}\}].$$

The forms of the Hironaka decompositions for $d = 2$ and 4 given here and in 4.5 invite an obvious conjecture for their form in arbitrary dimension d. Namely, the secondaries are given by the dot product of any one momenta with itself, raised to the 0-th all the way to the d-th powers. An explicit computation shows this to be the case also in $d = 1$ and 3. Therefore, we are led to the following

Conjecture 4.2.1. *The Hironaka decomposition of Lorentz-invariant algebras, $\mathbb{C}[V^n]^{O(d)}$, in the case of $n = d + 1$, is given by*

$$\mathbb{C}[V^n]^{O(d)} = \bigoplus_{m=0}^{d} (v_1 \cdot v_1)^m \; \mathbb{C}[\{\theta_i, \alpha_{ij}\}],$$

where the HSOP $\{\theta_i, \alpha_{ij}\}$ are as given by Equation 4.2.

The case of $(n,d) = (3,2)$ with $P = S_3$

For the full permutation case with $(n,d) = (3,2)$, we find the HSOP for the algebra $\mathbb{C}[V^3]^{O(2)\times S_3}$ in the way described in Subsection 4.2.2. This results in the following set

$$\theta_k = \mathrm{Pow}_k(v_i \cdot v_i) = \sum_{i=1}^{3} (v_i \cdot v_i)^k, \quad 1 \leq k \leq 2,$$
$$\alpha_k = \mathrm{Pow}_k(v_i \cdot v_j) = \sum_{i<j}^{3} (v_i \cdot v_j)^k, \quad 1 \leq k \leq 3.$$

Using the algorithm described in Subsection 4.2.3, we proceed to find the secondaries in a similar manner. The Hilbert series of the algebra, computed using methods described in Appendix A, is

$$H(\mathbb{C}[V^3]^{O(2)\times S_3}, t) = \frac{1 + t^4 + 2t^6 + t^8 + t^{12}}{(1-t^2)^2 (1-t^4)^2 (1-t^6)}.$$

We are therefore looking for 1 secondary at degree 0, 1 at degree 4, 2 at degree 6, 1 at degree 8, and 1 at degree 12. We find that the following set of polynomials

$$\begin{aligned}
\eta_1 &= 1,\\
\eta_2 &= (v_1 \cdot v_1)(v_2 \cdot v_3) + (v_2 \cdot v_2)(v_1 \cdot v_3) + (v_3 \cdot v_3)(v_1 \cdot v_2),\\
\eta_3 &= (v_1 \cdot v_1)^2(v_2 \cdot v_3) + (v_2 \cdot v_2)^2(v_1 \cdot v_3) + (v_3 \cdot v_3)^2(v_1 \cdot v_2),\\
\eta_4 &= (v_1 \cdot v_1)(v_2 \cdot v_3)^2 + (v_2 \cdot v_2)(v_1 \cdot v_3)^2 + (v_3 \cdot v_3)(v_1 \cdot v_2)^2,\\
\eta_5 &= (v_1 \cdot v_1)^2(v_2 \cdot v_3)^2 + (v_2 \cdot v_2)^2(v_1 \cdot v_3)^2 + (v_3 \cdot v_3)^2(v_1 \cdot v_2)^2,\\
\eta_6 &= (v_1 \cdot v_1)^5(v_2 \cdot v_3) + (v_2 \cdot v_2)^5(v_1 \cdot v_3) + (v_3 \cdot v_3)^5(v_1 \cdot v_2),
\end{aligned}$$

have remainders upon division by the Gröbner basis of the ideal generated by the HSOP which are non-zero and linearly independent. Therefore, we obtain a Hironaka decomposition of the algebra as follows

$$\mathbb{C}[V]^{O(2)\times S_3} = \bigoplus_{i=1}^{6} \eta_i \mathbb{C}[\{\theta_k, \alpha_k\}].$$

5

Approximation theorems

Armed with the generators of invariant algebras, we now move onto discussing how they can be exploited in machine learning via approximation theorems. The initial aim of this thesis was to find sets of variables able to approximate any invariant function corresponding to some property of an invariant system. But unfortunately, generating all invariant functions is a feat which, as of yet (and possibly for ever after), is mathematically insurmountable. We therefore restricted our attention to the polynomials, where developments in the field of invariant theory over the past century allowed us to obtain sets of generators of the invariant polynomials. Our goal in this Chapter is to show that this restriction is in fact not all that restrictive. That is, we wish to demonstrate that the invariant polynomials can indeed approximate all invariant functions.

It turns out that this is not true in general.[1] In what follows, we discuss and prove some approximation theorems which show that the invariant polynomials can approximate all continuous functions invariant under a compact group or a linearly reductive group defined over $\mathbb{C}$. To do this, we build on some recent work [43] and the Stone-Weierstrass theorem.

We then restrict our attention to the infamous machine learning tool, neural networks. There, we generalise a result from [43] that shows that neural networks utilising the minimal algebra generators (MAG) can approximate any continuous invariant function and we prove a similar, but new, result, namely that neural networks

[1]For a counterexample, consider the algebra in two variables $\mathbb{R}[x,y]^{\mathbb{R}^\times}$ invariant under the multiplicative group $\mathbb{R}^\times \cong \mathbb{R} \setminus \{0\}$ (which is linearly reductive over $\mathbb{R}$) that sends $(x,y) \mapsto (\lambda x, \lambda y)$, for $\lambda \in \mathbb{R}^\times$. It is not too difficult to see that the only invariant polynomials are the constants $\mathbb{R}[x,y]^{\mathbb{R}^\times} \cong \mathbb{R}$. Now consider the function $f(x,y) = \frac{x}{y}$ defined on any compact subspace $X \subset \mathbb{R} \times (\mathbb{R} \setminus \{0\})$ (*i.e.* $\mathbb{R}^2$ excluding $y = 0$). It is invariant (and continuous) but obviously cannot be approximated by the constant generators of $\mathbb{R}[x,y]^{\mathbb{R}^\times}$.

"endowed" with a Hironaka decomposition (HD) structure can also approximate all continuous invariant functions. Furthermore, these Hironaka decomposition networks are redundancy-free, since any invariant polynomial is uniquely expressed in terms of a Hironaka decomposition. We will see this advantage materialise in Chapter 6 when we put all these networks to the test.

5.1 Invariant functions

We will continue working in the field of complex numbers $\mathbb{C}$, but some of the results we obtain are also valid when restricting to the reals $\mathbb{R}$. The famous Stone-Weierstrass (SW) theorem [39, 38], which forms the cornerstone of approximation theory, states that any continuous function defined on a compact space can be approximated by a polynomial. The most general form of the SW theorem is very powerful.[2] Here, we concern ourselves with the special version regarding complex-valued functions defined on compact subspaces of $\mathbb{C}^n$.

Let $X \subset \mathbb{C}^n$ be a compact subspace and denote by $C(X)$ the set of all continuous complex-valued functions $f : X \to \mathbb{C}$.[3] As can easily be checked, the set $C(X)$ can be endowed with the structure of an algebra (with operations $+$, addition, and $\cdot$, multiplication). In fact, it is a special type of algebra called a $C^\star$-algebra possessing a norm $| - |$ defined for our purposes to be the uniform norm $|f| = \sup_{x\in X}|f(x)|$ and complex conjugation.[4] Now, the Stone-Weierestrass theorem states that

Theorem 5.1.1 (Stone-Weierstrass)**.** *The polynomial subalgebra* $\mathbb{C}[X]$ *is dense in* $C(X)$*. That is, for any* $f \in C(X)$ *and any real* $\epsilon > 0$*, one can find a polynomial* $r \in \mathbb{C}[X]$ *such that* $|r(x) - f(x)| < \epsilon$ *for all* $x \in X$*.*

This allows us to approximate all continuous functions (and thus all continuous invariant ones) using the polynomials in the coordinates x_i of $\mathbb{C}^n$ in some basis. We, however, would like to employ the MAG and HDs in our machine learning quests to approximate invariant functions. This crucially hinges on our ability to approximate all invariant functions using invariant polynomials which can then be expressed via MAG and HDs. Therefore, we must generalise (or rather restrict) the above theorem to the case of invariant functions and invariant polynomials. It turns out that this is non-trivial in general and we have not been able to demonstrate it completely. We

[2]The full theorem can be found in our discussions of Appendix B.
[3]Topologically, a continuous map is one which maps open sets to open sets.
[4]Details can be found in Appendix B.

sketch the roadmap to a full proof using two different approaches in Appendix B and discuss the bottlenecks that arise in each case.

In the case of continuous real-valued functions invariant under compact groups, the problem is relatively straightforward and has been solved in [43]. Here, we will prove the generalisation for the case of continuous complex-valued functions invariant under linearly reductive groups defined over $\mathbb{C}$. To do this, we first state a result which can be found in [11] or [29].

Lemma 5.1.1. *Let G be a linearly reductive group defined over $\mathbb{C}$ and let H be a maximal compact subgroup $H \subset G$. Consider the algebras of invariant polynomials $\mathbb{C}[V]^H$ and $\mathbb{C}[V]^G$. Then, $\mathbb{C}[V]^H = \mathbb{C}[V]^G$.*

The above Lemma does not hold for linearly reductive groups defined over $\mathbb{R}$.[5] The following arguments therefore only restrict to linearly reductive groups G defined over $\mathbb{R}$ if G itself is compact.

We also require the concept of a (normalised) Haar measure of a compact group H, μ_H. This is a group invariant measure which allows us to define an integral over the group, $\int_{h\in H} d\mu_H$, such that $\int_{h\in H} 1 d\mu_H = 1$. By borrowing the arguments from [43], we can now prove our generalisation as follows.

Theorem 5.1.2. *Let G be a linearly reductive group over $\mathbb{C}$ acting regularly on a finite-dimensional vector space $V \cong \mathbb{C}^n$ and denote by $\mathbb{C}[V]^G$ the algebra over $\mathbb{C}$ of G-invariant polynomials. Then, any continuous complex-valued G-invariant function $f : V \to \mathbb{C}$ can be approximated by an invariant polynomial $r_{sym} \in \mathbb{C}[V]^G$ on a compact subspace $X \subset V$.*

Proof. Let X be a compact subspace of V and consider the maximal compact subgroup $H \subset G$. Further consider the symmetrised set $X_{sym} = \cup_{h\in H}(h \circ X)$ which is compact as it is the image of a compact set $H \times X$ under a continuous map. By the Stone-Weierstrass theorem, for any $\epsilon > 0$ there exists a polynomial $r \in \mathbb{C}[V]$ such that $|r(x) - f(x)| < \epsilon$ for $x \in X_{sym}$. But r might not be G-invariant. So instead, consider the symmetrised function $r_{sym}(x) = \int_{h\in H} r(h \circ x) d\mu_H$ which is H-invariant. We note that r_{sym} is an H-invariant polynomial, since $r(h \circ x)$ is a fixed degree polynomial in x for any $h \in H$, and therefore $r_{sym} \in \mathbb{C}[V]^H$ (effectively, r_{sym} is the resulting projection of the Reynolds operator acting on r). By Lemma 5.1.1, we consequently have that

[5] This is because the result relies crucially on the fact that when G is a linearly reductive group defined over $\mathbb{C}$, the maximal compact subgroup is Zariski-dense in G. This is not guaranteed when G is defined over $\mathbb{R}$.

$r_{sym} \in \mathbb{C}[V]^G$. Finally,

$$\begin{aligned}|r_{sym}(x) - f(x)| &= \left|\int_{h\in H} r(h\circ x) - f(h\circ x)d\mu_H\right| \\ &\leq \int_{h\in H} |r(h\circ x) - f(h\circ x)|d\mu_H < \int_{h\in H} \epsilon d\mu_H = \epsilon\end{aligned}$$

for all $x \in X$, where we used the fact that f is G-invariant (and thus H-invariant), that $|r(x)-f(x)| < \epsilon$, and importantly, that $h\circ x \in X_{sym}, \forall h \in H$ and so the approximation is valid throughout the integral. □

This important theorem now allows us to safely use the invariant polynomials (and hence the generators we found in Chapter 4, MAG and HDs) to approximate continuous functions invariant under any compact groups or linearly reductive groups defined over $\mathbb{C}$ in machine learning algorithms. Crucially though, this theorem does not apply to the Lorentz group, $O(1, d-1, \mathbb{R})$, which is non-compact and linearly reductive over $\mathbb{R}$. That is, the assumption that the dot products and epsilon tensors can approximate any continuous Lorentz-invariant function is still merely an assumption. This could have important consequences for the way particles physicists process and analyse data, since without a rigorous proof, we cannot rule out the possibility of some exotic looking Lorentz-invariant functions which cannot be expressed or approximated by the dot products and epsilon tensors. Nonetheless, we believe that this is unlikely and that with an expert attempt at the bottlenecks in the proposed approaches of Appendix B, the problem can hopefully be resolved. For now, we brush this issue under the rug.

5.2 Neural Networks

For the rest of this Chapter, we restrict our attention to the infamous machine learning tool, neural networks. It is now a well-known fact that a neural network with a (large enough) single layer and a non-polynomial activation function can approximate any continuous real-valued function defined on some compact subspace of $\mathbb{R}^n$ [36].

Theorem 5.2.1 (Pinkus)**.** *Let $\sigma : \mathbb{R} \to \mathbb{R}$ be a non-polynomial function and let $V \cong \mathbb{R}^n$ be a real finite-dimensional vector space. Then any continuous function $f : V \to \mathbb{R}$ can be approximated, in the sense of uniform convergence on compact sets*[6]*, by functions*

[6] That is, for any compact subspace $X \subset V$ and any real $\epsilon > 0$, one can find a function $\hat{f}$ such that $|f(x) - \hat{f}(x)| < \epsilon$ for all $x \in X$.

$\hat{f} : V \to \mathbb{R}$ *of the form*

$$\hat{f}(x) = \sum_{i=1}^{N} c_i \sigma \left(\sum_{s=1}^{n} w_{is} x_s + h_i \right)$$

for $x \in V$*, some parameter* N*, and real coefficients* c_i, w_{is}, h_i*.*

Recently, a generalisation of this theorem to complex-valued functions was proved in [41].

Theorem 5.2.2 (Voigtlaender)**.** *Let* $\sigma : \mathbb{C} \to \mathbb{C}$ *be a function that is* not *almost polyharmonic.*[7] *Let* $V \cong \mathbb{C}^n$ *be a complex finite-dimensional vector space. Then any continuous function* $f : V \to \mathbb{C}$ *can be approximated, in the sense of uniform convergence on compact sets, by functions* $\hat{f} : V \to \mathbb{C}$ *of the form*

$$\hat{f}(x) = \sum_{i=1}^{N} c_i \sigma \left(\sum_{s=1}^{n} w_{is} x_s + h_i \right)$$

for $x \in V$*, some parameter* N*, and complex coefficients* c_i, w_{is}, h_i*.*

The function $\hat{f}$ in the above theorem represents a complex-valued neural network with a single layer of width N. Interestingly, unlike real-valued neural networks, the criterion on the activation function of a single layer (shallow) and a multiple layer (deep) complex-valued networks is different.[8] This theorem can also be employed to approximate any continuous complex-valued *multi-dimensional* function $f : V \to \mathbb{C}^m$ by "stitching" together m networks each approximating one of the components of f. One thus obtains a network of width of mN which can approximate any function $f : V \to \mathbb{C}^m$.

In [43], real-valued neural networks were shown to be able to approximate any continuous real-valued function invariant under a compact group using the MAG of the invariant algebra of polynomials as inputs. Here, we will prove the generalisation of that theorem to complex-valued neural networks and continuous complex-valued functions invariant under linearly reductive groups defined over $\mathbb{C}$.

Theorem 5.2.3 (MAG networks)**.** *Let* G *be a linearly reductive group over* $\mathbb{C}$ *acting regularly on a finite-dimensional vector space* $V \cong \mathbb{C}^n$ *and let* $g_1, \ldots, g_m : V \to \mathbb{C}$

[7] A function σ is almost polyharmonic if there exist $m \in \mathbb{N}$ and an infinitely differentiable $g : \mathbb{C} \to \mathbb{C}$ with $\Delta^m g = 0$ such that $\sigma = g$ almost everywhere. Here, $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ is the usual Laplace operator on $\mathbb{C} \cong \mathbb{R}^2$.

[8] For multiple layer neural networks, σ must be neither a polynomial, a holomorphic function, nor an antiholomorphic function.

be a set of minimal algebra generators of the invariant algebra $\mathbb{C}[V]^G$. Then, any continuous G-invariant function $f : V \to \mathbb{C}$ can be approximated, in the sense of uniform convergence on compact sets, by invariant functions $\hat{f} : V \to \mathbb{C}$ of the form

$$\hat{f}(x) = \sum_{i=1}^{N} c_i \sigma \left(\sum_{s=1}^{m} w_{is} g_s(x) + h_i \right)$$

for $x \in V$, some parameter N, and complex coefficients c_i, w_{is}, h_i.

Proof. It is obvious that the functions $\hat{f}$ are G-invariant, so we only need to prove completeness.

From theorem 5.1.2, we can approximate f by an invariant polynomial $r_{sym} \in \mathbb{C}[V]^G$ on a compact subspace $X \subset V$. Next, we use the MAG to express $r_{sym} = h(g_1, \ldots, g_m)$ for some polynomial h. It remains to approximate the polynomial $h(g_1, \ldots, g_m)$ by an expression of the form $H(x) = \sum_{i=1}^{N} c_i \sigma(\sum_{s=1}^{m} w_{is} g_s(x) + h_i)$ on the compact set $\{g_1(x), \ldots, g_m(x) | x \in X\}$. Using Theorem 5.2.2, we can do this with accuracy ϵ', $|H(x) - h(x)| < \epsilon'$ for all $x \in X$. Finally, setting $\hat{f}(x) = H(x)$, we obtain an $\hat{f}$ of the required form such that $|\hat{f}(x) - f(x)| < \epsilon$ for $x \in X$. $\square$

As we have mentioned in Section 4.1, the MAG are not algebraically independent in general. This leads to redundancies in the description of invariant polynomials via the MAG which, consequently, reduces the efficiency of neural networks using the MAG as inputs. HDs however are redundancy-free and express any invariant polynomial uniquely. To make use of this, we generalise Theorem 5.2.3 to the case of HDs by constructing a neural network which makes use of the nice HD structure.

Theorem 5.2.4 (HD networks)**.** *Let G be a linearly reductive group over $\mathbb{C}$ acting regularly on a finite-dimensional vector space $V \cong \mathbb{C}^n$ and let $\theta_1, \ldots, \theta_m : V \to \mathbb{C}$ and $\eta_1, \ldots, \eta_p : V \to \mathbb{C}$ be a Hironaka decomposition (primaries and secondaries respectively) of the invariant algebra $\mathbb{C}[V]^G = \bigoplus_i \eta_i \mathbb{C}[\theta_j]$. Then, any continuous G-invariant function $f : V \to \mathbb{C}$ can be approximated, in the sense of uniform convergence on compact sets, by invariant functions $\hat{f} : V \to \mathbb{C}$ of the form*

$$\hat{f}(x) = \sum_{k=1}^{p} \eta_k(x) \sum_{i=1}^{N} c_{ki} \sigma \left(\sum_{s=1}^{m} w_{is} \theta_s(x) + h_i \right)$$

for $x \in V$, some parameter N, and complex coefficients c_{ki}, w_{is}, h_i.

Proof. It is obvious that the functions $\hat{f}$ are G-invariant, so we only need to prove completeness.

From theorem 5.1.2, we can approximate f by an invariant polynomial $r_{sym} \in \mathbb{C}[V]^G$ on a compact subspace $X \subset V$. Next, we use a HD to express $r_{sym} = \sum_{k=1}^{p} \eta_k h_k(\theta_1, \ldots, \theta_m)$ uniquely. It remains to approximate the polynomials $h_k(\theta_1, \ldots, \theta_m)$ by expressions of the form $H_k(x) = \sum_{i=1}^{N} c_{ki}\sigma(\sum_{s=1}^{m} w_{is}\theta_s(x) + h_i)$ on the compact set $\{\theta_1(x), \ldots, \theta_m(x) | x \in X\}$. Using the multi-dimensional version of Theorem 5.2.2, we can do this with accuracy ϵ', $|H_k(x) - h_k(x)| < \epsilon'$ for all k and $x \in X$. Finally, setting $\hat{f}(x) = \sum_{k=1}^{p} \eta_k(x) H_k(x)$, we obtain an $\hat{f}$ of the required form such that $|\hat{f}(x) - f(x)| < \epsilon$ for $x \in X$. □

As we will see in Section 6.2, the width of the network, N, is very important in HD networks. It must be large enough and comparable to the number of secondaries so that the network can approximate the (in general distinct) polynomials $h_k(\theta_1, \ldots, \theta_m)$. When this is satisfied, the full power of HD networks becomes apparent over MAG networks.

A basic `C++` implementation of these networks can be found on-line.[9]

[9] Available at: https://github.com/WardHaddadin/Invariant-polynomials-and-machine-learning.

6

Application to neural networks

We finally arrive at the application portion of the thesis. In this Chapter, we perform comprehensive testing of the neural networks proposed in Chapter 5. We first do this by direct brute-force testing on simulated data under a variety of hyperparameter choices. We find that minimal algebra generators and Hironaka decomposition networks outperform ordinary ones, with Hironka decomposition networks performing the best.

For a different perspective on quantifying the performance of these networks, we turn to Bayesian inference methods and treat the neural networks as models. Using nested sampling, we compute the evidence of these models and show that, beyond a certain network size, Hironaka decomposition networks again come out on top.

6.1 Brute-force testing

We will perform the first part of our testing by applying the above networks to regression problems.[1] That is, given some data set $D = \{(x_i, y_i)\}$ of size $|D|$, described by a true function f, we would like to train our networks, $\hat{f}$, by tuning their coefficients to approximate f. Goodness-of-fit will be defined here using the loss metric of *mean squared error*, $\mathrm{MSE}(D) = \frac{1}{|D|}\sum_i(\hat{f}(x_i) - y_i)^2$. Instead of performing regression on arbitrary invariant functions, it is a better idea to do regression on polynomials.[2] This is because by regressing on polynomials of a certain maximum degree, we are able to control the "difficulty" of the regression problem which in turn allows us to quantify

[1]We restrict to real-valued functions throughout our testing.

[2]There is some theoretical work on networks fitting polynomials [9, 1] which discusses and proves some results regarding the approximating power of networks in relation to polynomial degree and sparsity.

the performance of the networks when applied to problems of varying complexity. Additionally, we can easily produce randomly generated polynomials for testing.

We consider the Lorentz- and permutation-invariant polynomials in particle momenta and perform testing for the invariant algebras listed in Table A.1 of Appendix A. It is obvious that using the dot products and epsilon tensor contractions as inputs is a better idea than just using the plain momenta. We therefore restrict our testing to compare the performance of networks utilising HDs and MAG, which we call *HD* and *MAG networks* respectively, against networks utilising the Weyl generators (dot products and epsilon tensor contractions), or *Weyl networks*.

6.1.1 Summary of results

For each hyperparameter, testing was carried out on 100 repetitions of randomly generated invariant polynomials and randomly generated data.[3] The data was split into an $80:20$ train-to-validation ratio and L_2-regularisation methods were used to prevent overfitting (this is discussed in the context of Bayesian neural networks in Section 6.2).

The general theme that can be seen from the results in Figures 6.2, 6.3, and 6.4 is that HD networks perform better than MAG networks which in turn perform better that Weyl networks. As expected, the improvement in performance becomes more apparent in the presence of permutation groups of higher order since the permutation-invariant generators save the networks the effort required to learn permutation-invariance. We can also see that the improvement is more significant with lower degree polynomials, and while still present at higher degrees, plateaus. Notably, HD networks tend to prefer tanh activation functions while MAG and Weyl networks perform better with ReLU and Leaky ReLU.

Below, we provide a more detailed description of the behaviour of the networks under a variety of hyperparameters.

Number of network parameters

We tested how the benefit or dis-benefit of each network scales with the total number of parameters. That is, at what size does the investment in the MAG and HD networks become worth it? We will get a more complete picture of this in the next Section,

[3] The input and output data was normalised to be in the range $[-1,1]$.

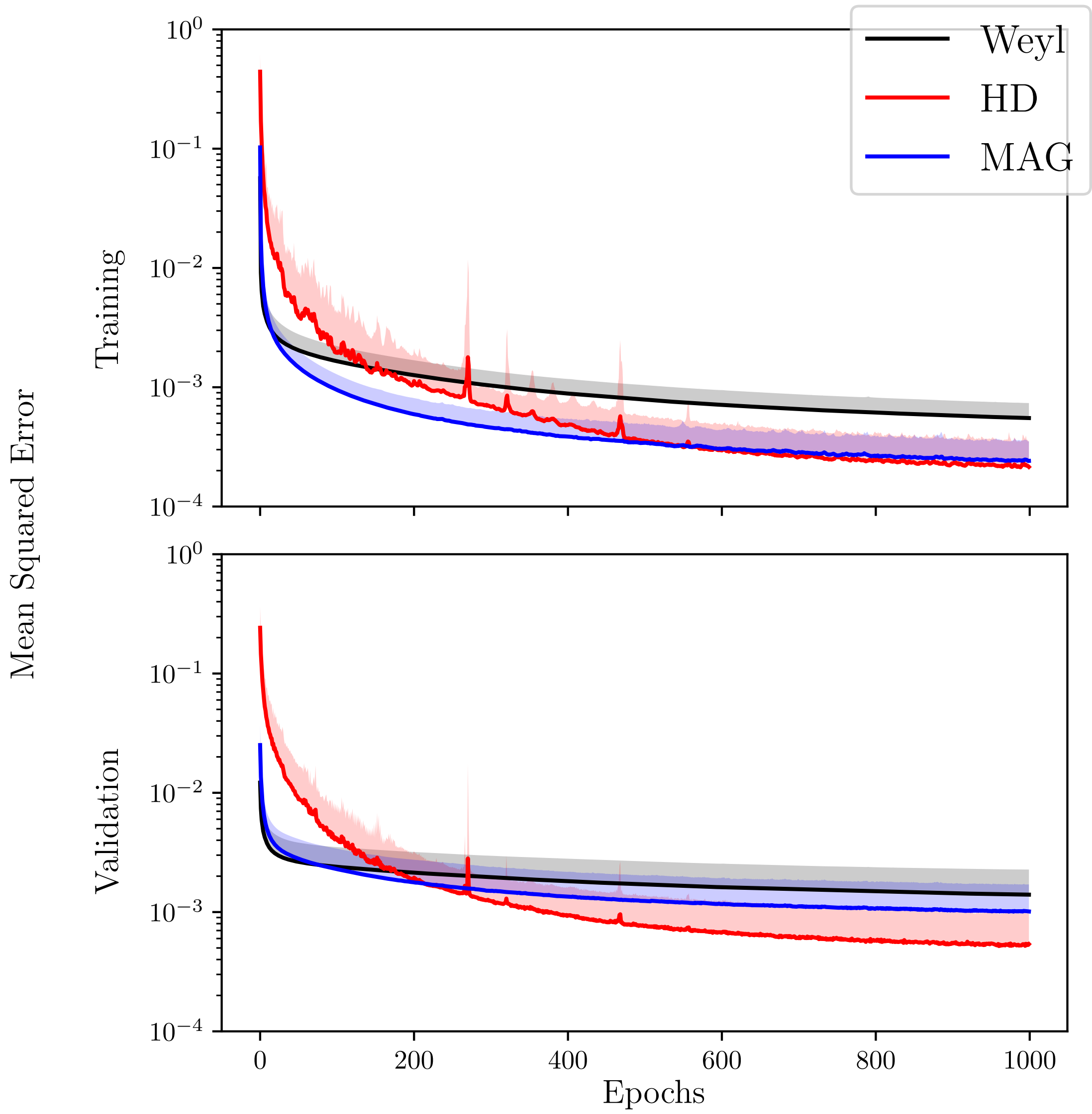

Figure 6.1 Learning comparison of Hironaka decomposition, minimal algebra generators, and Weyl networks applied to polynomial regression for the case of $n = 4, d = 4, S_2 \times S_2$. Averaged from applying the networks to 100 randomly generated polynomials. L_2 regularisation was applied with $\lambda = 10^{-5}$. The number of parameters was ~ 2500, the number of hidden layers was 2, the activation function of the hidden layers was $\tanh(x)$, the polynomial degree used was 10, and the size of the data set was 5000 points. The shaded regions denote a variation of 1σ.

where we treat these networks as Bayesian models. Here, we tested networks with 500, 1000, and 2500 parameters.[4]

We can see that there is no notable improvement in performance for MAG over Weyl networks as the number of parameters increases. But for HD networks, the efficiency gain increases with the number of parameters. This is probably due to the point made in Chapter 5, that HD networks need to have a width comparable to the number of secondaries for their full power to kick in.

Activation functions

Activation functions play a major role in the performance of networks. We tested some of the most popular activation functions

$$\begin{aligned}
&\tanh(x) \\
&\mathrm{ReLU}(x) = \max(x, 0) \\
&\text{Leaky ReLU}(x) = \max(x, 0) + \max(0, -cx)
\end{aligned}$$

where c is some small positive constant. Since we are doing a regression task, the activation function of the output layer will always be a linear activation function. What we vary are the activation functions of the hidden layers.

A common problem which arises with the ReLU function is called "ReLU death". This occurs when the network parameters get pushed by gradient descent to the region where the ReLU function is zero and so cannot leave (since the gradient there is also zero). The Leaky ReLU tries to combat this issue by adding a small negative contribution in that region.

We can see that in all cases in Figure 6.3, HD networks prefer $\tanh(x)$ as an activation function. We can also see that they perform poorly with ReLU activation functions. More testing is required to understand why this happens exactly, but it might be due to the final secondaries layer being more susceptible to ReLU death.

Depth vs. Width of networks

We also tested how the depth of the networks affects performance. Generically, the touted mantra is that shallower networks (with fewer hidden layers) learn faster but

[4]The actual number of parameters used in testing does not always exactly match these values. The exact number depends on the number of inputs, the number of layers, and the width of the layers. We constructed the networks to get the actual number of parameters as close as possible to these values.

deeper networks (with more hidden layers) learn more complicated structures in the data. To see if this applies here, we tested networks with $1, 2$, and 3 hidden layers.

From Figure 6.4, we can see that the width vs. depth ratio makes little difference. For HD networks, this is true as long as the width of the networks is comparable to the number of secondaries. As we will see in Section 6.2, if given the choice between narrow and deep vs. wide and shallow networks, one should always choose the wide and shallow ones to allow the networks to better approximate the polynomials $h_k(x)$ in $r_{sym} = \sum_k \eta_k h_k(x)$.

Size of data set

Another interesting hyperparameter to test was the size of the data set. That is, how much data does each type of network need to learn? This is very important when one is trying to probe properties of the system which either occur rarely (and so appear less frequently in the data) or when one does not have much data about the system because it is expensive or difficult to obtain. We carried out the testing using 5000, 7000, and $10,000$ data points. As expected, training on a larger data set resulted in lower overall training loss and validation loss. But, we found no additional significant improvement in any one network over the others.

Error of data set

Finally, it was also interesting to see how noise in the data set affects performance. That is, are some architectures more robust to noise than others? We trained the networks on data sets with 0% error (*i.e.* "perfect" data), 1% error, 5% error, and for an extreme case 20% error. As expected, the noisier data sets resulted in worse performance, but, as with the size of the data set, there was no notable reduction of the performance of one network over the others.

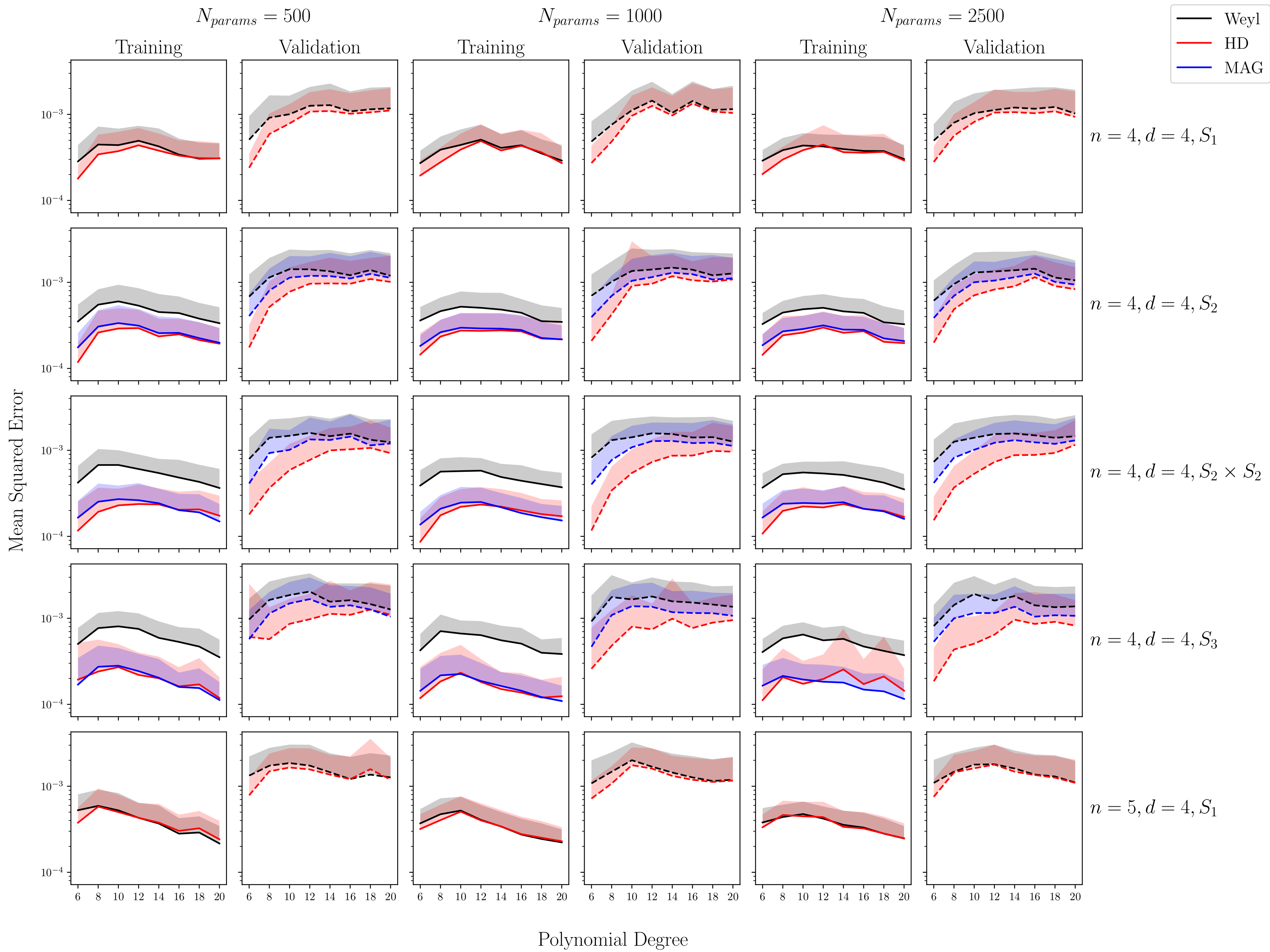

Figure 6.2 A grid showing the performance of Hironaka decomposition, minimal algebra generators, and Weyl networks applied to polynomial regression with increasing polynomial degree. Averaged from 100 runs on randomly generated polynomials at each degree. L_2 regularisation was applied with $\lambda = 10^{-5}$. The activation function of the hidden layers was $\tanh(x)$ and the number of hidden layers was 2. The shaded regions denote a variation of 1σ.

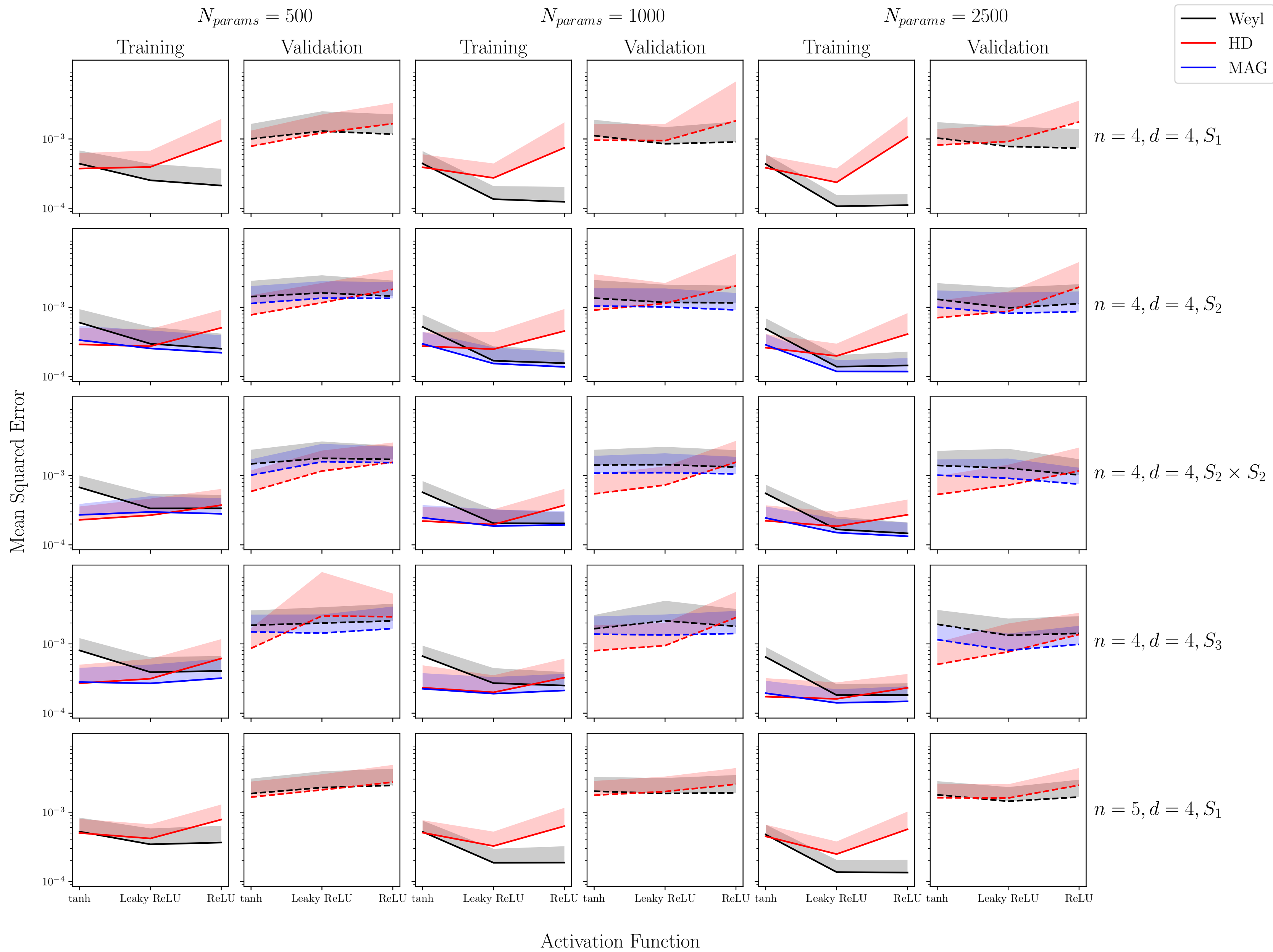

Figure 6.3 A grid showing the performance of Hironaka decomposition, minimal algebra generators, and Weyl networks for some invariant algebras applied to polynomial regression with different activation functions. Averaged from applying the networks to 100 randomly generated polynomials for each activation function. L_2 regularisation was applied with $\lambda = 10^{-5}$. The polynomial degree used was 10 and the number of hidden layers was 2. The shaded regions denote a variation of 1σ.

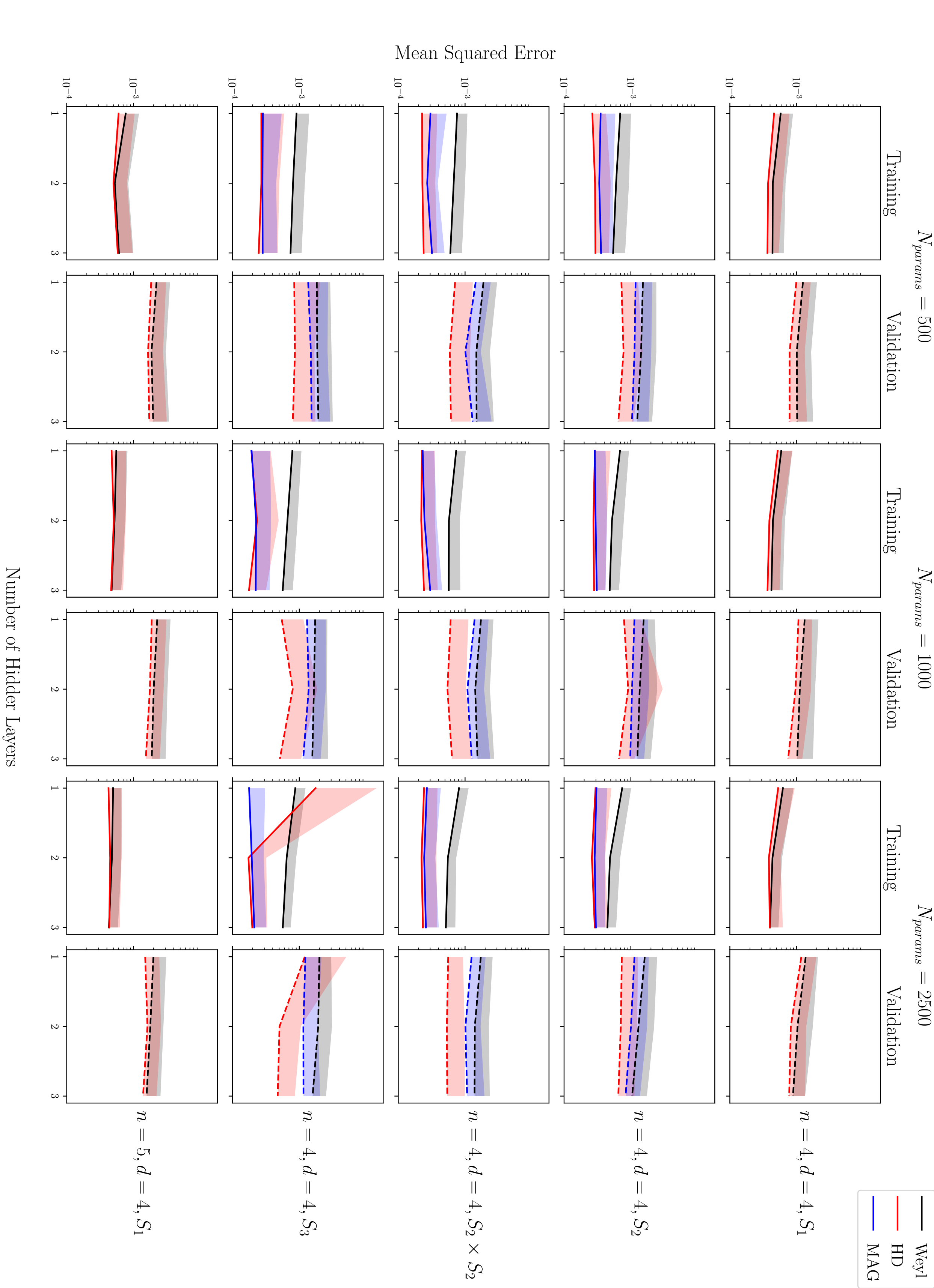

Figure 6.4 A grid showing the performance of Hironaka decomposition, minimal algebra generators, and Weyl networks for some invariant algebras applied to polynomial regression with increasing number of hidden layers. Averaged from applying the networks to 100 randomly generated polynomials for each activation function. L_2 regularisation was applied with $\lambda = 10^{-5}$. The activation function of the hidden layers was $\tanh(x)$ and the polynomial degree used was 10. The shaded regions denote a variation of 1σ.

6.2 A Bayesian's-eye view

Let us now take a step back and re-examine what comparing neural networks actually means. A neural network is essentially a glorified model, which we denote by M, that depends on some parameters $\alpha = \{\alpha_i\}$. "Training" a neural network amounts to tuning these parameters, using some gradient descent algorithm, to fit the data "best", where goodness-of-fit is judged by some loss function. More precisely, this process can be viewed as the *maximum likelihood estimation* (MLE) of a Bayesian inference problem on some data set D. Let $P(D|\alpha, M)$ denote the likelihood[5] of the data given some parameters α and let $P(\alpha|M)$ denote the prior[6] on these parameters. Then, what is usually termed "training" or "learning" is the process of finding the MLE of the posterior

$$P(\alpha|D, M) \propto P(D|\alpha, M)P(\alpha|M). \tag{6.1}$$

A glaring disadvantage of thinking solely of the MLE as the end goal is the fact that we gain no insight of how confident our model (neural network) is performing when applied to unseen data. Two different points in parameter space might both result in a high posterior (*i.e.* perform well on training data), but correspond to models whose performance is very different on unseen data. For a toy example, consider Figure 6.5. Although the network fits the training data very well for the two points in parameter space, the behaviour on unseen data is very different. Furthermore, both of these choices of parameters perform poorly on unseen data.

Therefore, instead of thinking of the MLE as the ultimate goal, a model scientist[7] should think of neural networks as Bayesian models. Viewed as such, the predictions of Bayesian neural networks, obtained by averaging the predictions of a regular network at all points of the parameter space weighted by the posterior (corresponding to goodness-of-fit) as

$$\hat{f}_{\text{avg}}(x) = \int \hat{f}(x|\alpha)P(\alpha|D, M)d\alpha, \tag{6.2}$$

[5]The likelihood must be a positive semi-definite, strictly decreasing, function of the loss.

[6]The prior corresponds to using *regularisation* in conventional training approaches. A gaussian prior on the parameters corresponds to using L_2-regularisation.

[7]Pardon the pun.

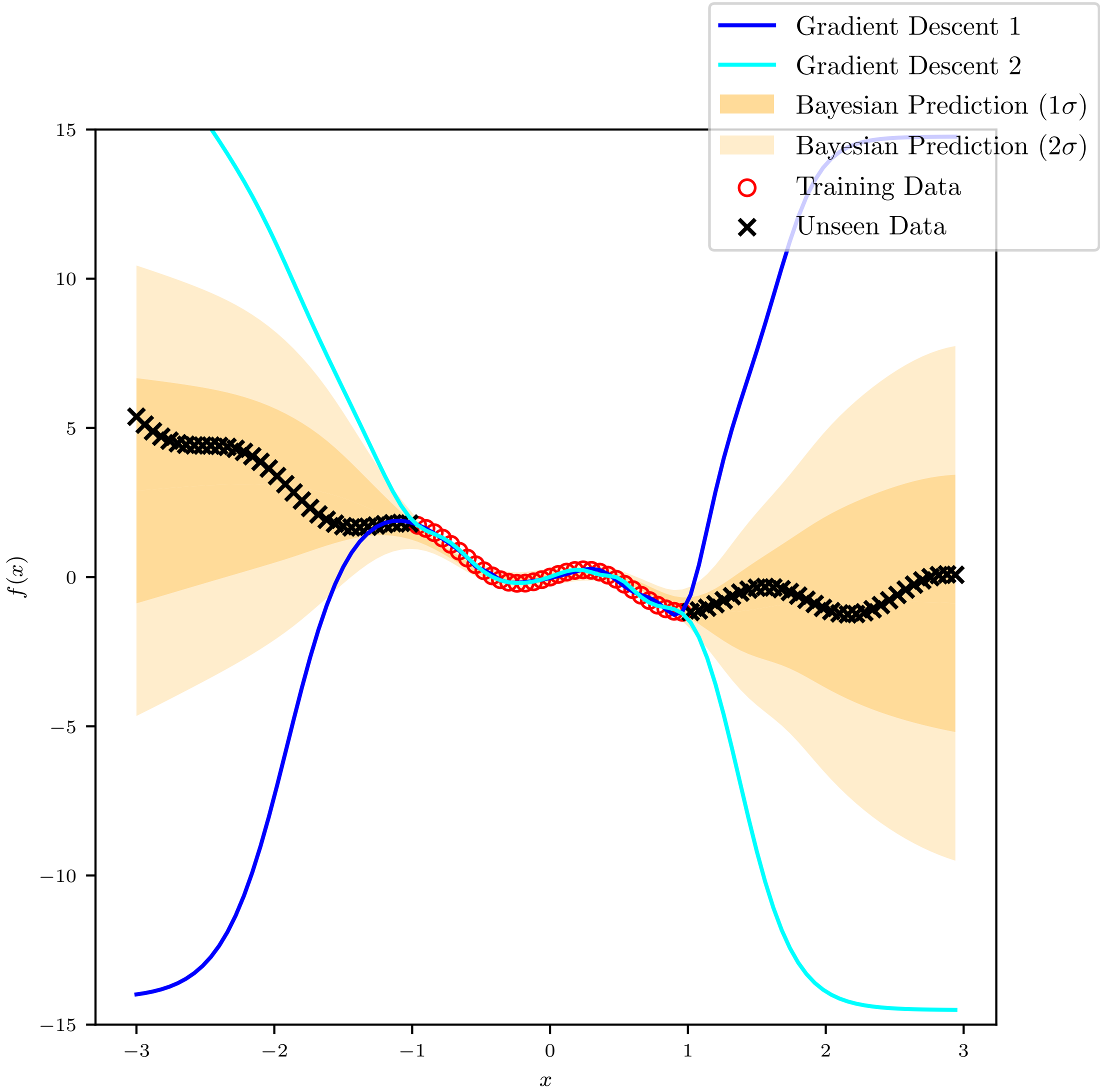

Figure 6.5 Neural networks applied to a regression problem. The predictions from two networks trained using the conventional approaches of gradient descent fit the training data well but perform poorly on unseen data. The Bayesian approach on the other hand provides meaningful and modest predictions on unseen data, while still performing well on training data.

where $\hat{f}(x|\alpha)$ is the output of the network for some set of parameters α, are now much more meaningful.[8] The advantage of this procedure can be seen in Figure 6.5, where the Bayesian network gives better predictions, but more importantly is modest where it is uncertain (*i.e.* on the unseen data). Alas, this approach is not without its limitations. For a thorough discussion of the issues that arise, see [25]. Here, we will highlight two immediate problems that come to mind.

The first, and major, problem is twofold and can be traced back to the ever-looming curse of dimensionality. When working with models with more than a few hundred parameters, the integral in Equation 6.2 quickly becomes intractable. On one hand, high dimensional integrals are difficult to perform, but also, computing the required constant of proportionality of Equation 6.1

$$Z(M) = P(M|D) = \int P(D|\alpha, M)P(\alpha|M)d\alpha,$$

termed the *evidence*, is equally as difficult. The usual way around both of these problems is to use *Markov chain Monte Carlo* (MCMC) methods which produce a representative sample of points from the posterior and avoid computing $Z(M)$. The integral in Equation 6.2 then reduces to a manageable sum and one can obtain an estimate of the function $\hat{f}_{\text{avg}}(x)$. But since these methods have no access to $Z(M)$, they miss out on the opportunity to compare different models (in our case networks utilising different input methods). This is because, given the evidences, $Z(M_i)$, and priors, $P(M_i)$, of a set of models $\{M_i\}$, the posterior (or probability) of a model is given by

$$P(M_i|D) = \frac{P(D|M_i)P(M_i)}{P(D)} = \frac{Z(M_i)P(M_i)}{\sum_i Z(M_i)P(M_i)}.$$

Therefore, having access to the evidence of models gives us an excellent tool to choose, with good confidence, the best performing model (or even better, average the predictions from all the models weighted by their respective posteriors). One of the cutting-edge methods used to compute the evidence is called *nested sampling*. Its basic premise is to transform the multi-dimensional integral in Equation 6.2 to a one-dimensional one which is much easier to compute. We give a brief exposition of nested sampling in Appendix C.

The second obstacle that we face when applying this Bayesian approach to neural networks is that unlike usual models where one knows what the likelihood is, here

[8] One can also quantify the uncertainty here as well by computing the variance $\text{Var}(\hat{f}(x)) = \int \hat{f}(x|\alpha)^2 P(\alpha|D, M)d\alpha - \hat{f}_{\text{avg}}(x)^2$.

we have the freedom to choose the likelihood. This extra freedom seems to throw us back to square one, where we are forced to choose a preferred likelihood function over any other function.[9] In the case of a mean-square-error loss function, this is slightly remedied by the fact that we can make an educated choice and construct a gaussian looking function of the loss

$$P(D|\alpha, M) \propto \exp\left(-\frac{\sum_i (y_i - \hat{f}(x_i))^2}{2N\sigma^2}\right),$$

to act as the likelihood. There is still the freedom of choosing the parameter σ^2 though, but since we are only concerned with performing this analysis as a proof-of-concept, we will make the simplifying choice of setting $\sigma^2 = 1$.[10]

We will brush this second problem under the rug for the rest of this work and focus on obtaining evidence estimates for our networks using the semi-arbitrary gaussian choice of the likelihood. To perform this analysis, we used the nested sampling package PolyChord [21, 22].[11] In line with the usual L_2-regularisation techniques used in gradient descent algorithms, we choose a gaussian prior on the parameters

$$P(\alpha|M) \propto \exp\left(-\frac{\sum_i \alpha_i^2}{2\sigma_\alpha^2}\right)$$

where $\sigma_\alpha^2 = 0.01$. Unfortunately, the curse of dimensionality, although mitigated by nested sampling, still haunts us. We therefore only perform this analysis on networks of sizes 100, 200, and 300 parameters with a single hidden layer.

From Figure 6.6, we can see that the performance of the HD networks depends highly on the number of parameters. As discussed previously, and further confirmed here, this is because for the HD networks to make full use of the nice structure of a HD, they must have enough nodes ($\sim O$(number of secondaries)) in the networks to approximate the polynomials in the primaries multiplying the secondary generators. After a certain threshold, their performance significantly overtakes that of MAG networks which in turn perform better than Weyl networks. These results, combined with those found in Section 6.1, seem to confirm what was found in [33], namely that

[9] There have been some recent atempts to perform likelihood-free inference [12] but we will not discuss these here.

[10] The Bayesian way to deal with this issue is to make σ^2 itself a parameter of the model which one then infers from the data. The process of inferring hyperparameters from the data is called *Hierarchical modelling*.

[11] Available as PolyChordLite on `github`: https://github.com/PolyChord/PolyChordLite

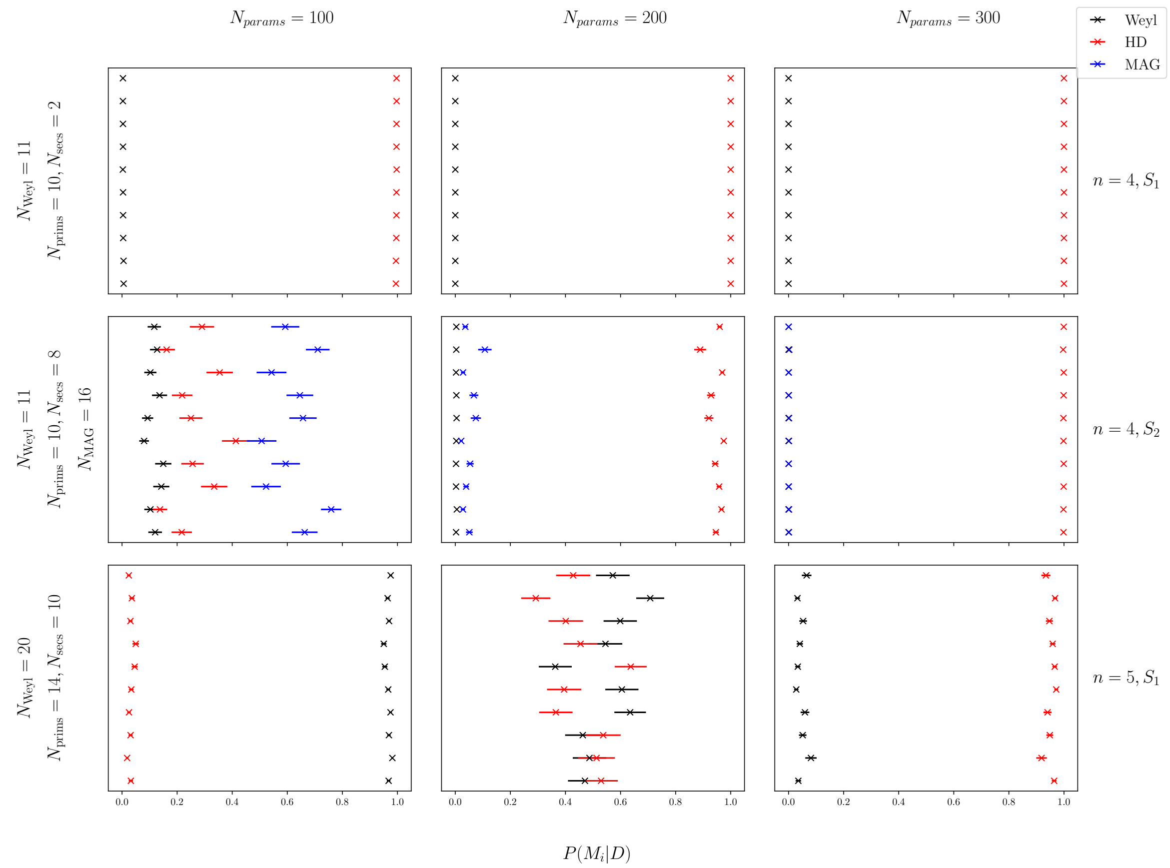

Figure 6.6 A grid showing the evidence of Hironaka decomposition, minimal algebra generators, and Weyl networks with a single layer and an increasing number of generators applied to polynomial regression. The evidence of each network can be used as a measure of its generalisability as observed by MacKay [33]. Using this as a proxy, we can see that the performance of Hironaka decomposition networks depends highly on the number of parameters, but where large enough networks can be constructed, they outperform minimal algebra generators and Weyl networks. The evidence of each network was computed 10 times. The activation function of the hidden layer was $\tanh(x)$ and the polynomial degree used was 10. The bars denote a variation of 1σ.

networks with a higher evidence tend to perform better on unseen validation data in general.

Although their evidence and validation loss does not rival that of HD networks, MAG networks might actually be preferred over HD networks in some cases. Because of their smaller number of generators and the relative ease with which they can be obtained, MAG can be a middle ground compromise between the basic Weyl generators and the difficult to find, but more efficient, HDs.

7

Conclusion and outlook

In this thesis, we attempted to address the problem of incorporating our knowledge of the symmetries of certain systems (particle physics and physical chemistry) into the machine learning methods we use to describe them. To do this, we first restricted our attention from all invariant functions that could describe our systems to the invariant polynomials, which allowed us to make use of the powerful tools of commutative algebra and invariant theory. By building on two results of Weyl, we characterised the algebras of polynomials invariant under the Euclidean group (rotations and translations) and then generalised the results to the algebras invariant under the combined Lorentz (orthogonal) and permutation groups and the Euclidean and permutation groups. The invariant algebras when permutations are included are much more complicated and the tools of invariant theory gave us two approaches to characterise them.

In the first, we developed a systematic method which produces sets of minimal algebra generators for the Lorentz- (or Euclidean-) and permutation-invariant polynomials. Our method results in manageable sets of generators for phenomenologically-relevant examples, at least when the number of particles is sufficiently small, and we hope that the results will prove to be useful in future phenomenological analyses. This approach has some shortcomings though. One is that it is computationally intractable to apply to the case where parity is not a symmetry and the second is that the resulting generators are not algebraically independent and contain non-trivial redundancies which may frustrate analyses.

In the second approach, we addressed the problem of redundancies. Instead of providing a set of generators (FFT) and the relations between them (SFT), we observed that one may provide (via the theorem of Hochster and Roberts) a more direct characterization in terms of a Hironaka decomposition, that is as a free, finitely-generated module over a polynomial subalgebra. This approach has the added advantage that it readily generalises to the case where parity is not a symmetry. In cases where $n \leq d+1$, we gave an explicit solution (for an arbitrary permutation symmetry) to the "hard" part of finding such a decomposition, namely the identification of a homogeneous system of parameters (HSOP). The "easy" part of finding a decomposition, namely the identification of suitable secondary generators, reduces to a linear algebra algorithm, but is nonetheless inefficient. We provided Hironaka decompositions in the examples of $(n, d) = (5, 4)$ with $P = 1$ and $(n, d) = (3, 2)$ with $P = S_3$ and a conjecture in the general case of $n = d + 1$ with no permutations.

To arrive back at our original aim of approximating invariant functions in machine learning, we then presented and extended some approximation theorems which utilise the generators of invariant polynomials to approximate invariant functions in machine learning tasks in general and in neural networks specifically. We showed that invariant polynomials can indeed approximate continuous functions invariant under compact groups or linearly reductive groups defined over $\mathbb{C}$. Finally, to put our results to the test, we implemented the invariant generators in neural networks applied to polynomial regression and tested their performance under a variety of hyperparameters. We found an overall reduction in training loss and a more significant reduction in validation loss compared to ordinary approaches. These improvements were dependant on the maximal polynomial degree and the activation functions used. We also performed a Bayesian analysis of these networks and found that, beyond a minimum size, Hironaka decomposition networks always outperform minimal algebra generator networks and Weyl networks.

As is always the case, there are some drawbacks to our general approach. The first is that computing generators of invariant algebras in either flavour is in general hard. Although we provided a method to construct minimal algebra generators and solved the problem of finding HSOPs of Hironaka decompositions for the invariant algebras, the task of finding the secondaries in both approaches is often computationally prohibitive due to their large number and high degrees. The description via Hironaka decompositions requires one to find all the secondaries and so might not always be the preferred choice. In such cases, a description via a set of minimal algebra generators (which can be significantly smaller since one can make away with some generators) might be a better alternative in either theoretical work or for efficient computational use. In both cases, the silver-lining is that the costly procedure of finding explicit generators has to be done only once and then one can just look up these generators from a database.

Another problem is that our generators are not able to fully separate the orbits, which is certainly a useful thing to do from a physicist's point of view (for example in searching for parity violating LHC signals, as explored in [31]). To give a somewhat trivial example, the invariant $p \cdot p$ is unable to separate the orbits with $p \cdot p = 0$ and with either $p = 0$ or $p \neq 0$. For a slightly non-trivial example, consider the case of $n = 3, d = 2$ with no permutation symmetry given in Subsection 4.2.3. Here, the Hironaka decomposition fails to separate the orbits $\{p = (p, p), q = (q, q), r = (r, r)\}$ and $\{p = (0, 0), q = (0, 0), r = (0, 0)\}$, on which all the primaries and secondaries (except the trivial secondary 1) vanish. Separating invariants is even useful from a purely mathematical perspective. As can be seen in Appendix B, the crucial criterion for an algebra of invariant polynomials to approximate the algebra of continuous invariant functions is that the algebra separates the same orbits as the functions. If this is the case, then one only needs to find the generators of the separating subalgebra of the invariant polynomials (which is usually a smaller set) to approximate the invariant functions. The drawback here is that the subject of separating invariants is still young and finding these separating sets is not as straightforward. Furthermore, separating sets lack a nice structure (compared to Hironaka decompositions) in general.

Finally, there are a few avenues for future work on both the theoretical and computational sides. First, an interesting open problem to attack is the restriction of the Stone-Weierstrass theorem to the case of group invariant functions. Although proved for complex-valued functions invariant under linearly reductive groups over $\mathbb{C}$ and real-valued functions invariant under compact groups over $\mathbb{R}$, the case of a general group is yet to be solved. This has important consequences even for systems which are only invariant under the Lorentz group, which is linearly reductive over $\mathbb{R}$ but not compact. Weyl's theorem for Lorentz-invariant polynomial generators is taken for granted when considering non-polynomial functions because of the assumption that we can approximate any continuous Lorentz-invariant function by the generators of the invariant polynomials. But as we have seen, this is not guaranteed.

On the computational side, it would be interesting to see how these networks perform in different machine learning settings such as classification problems (in neural networks or otherwise) and unsupervised learning problems. Furthermore, it would be interesting to see whether the results found using the Bayesian approach still hold if likelihood-free inference methods are used to circumvent the need to choose a likelihood function. A final problem to address comes up when one is dealing with systems with a variable number of vectors.[1] In such cases, a method to input the invariant generators of these different algebras into the machine learning algorithms in an "invariant" manner is still lacking.

[1]For example, when one is trying to probe the production of some exotic particle which can be produced through different processes involving a different number of final state particles. For another example, consider training a neural network to approximate the solubility of molecules with different numbers of atoms. Thanks to Emil-Nicolae Nichita for first bringing this to my attention.

References

[1] Andoni, A., Panigrahy, R., Valiant, G., and Zhang, L. (2014). Learning polynomials with neural networks. 32(2):1908–1916.

[2] Aslaksen, H., Chan, S. P., and Gulliksen, T. (1996). Invariants of S_4 and the shape of sets of vectors. *Applicable Algebra in Engineering, Communication and Computing*, 7(1):53–57.

[3] Atiyah, M. and MacDonald, I. (1994). *Introduction to commutative algebra*. Avalon Publishing, New York, NY, US.

[4] Beecken, M., Mittmann, J., and Saxena, N. (2013). Algebraic independence and blackbox identity testing. *Information and Computation*, 222:2 – 19.

[5] Bogatskiy, A., Anderson, B., Offermann, J. T., Roussi, M., Miller, D. W., and Kondor, R. (2020). Lorentz Group Equivariant Neural Network for Particle Physics.

[6] Bruns, W. and Herzog, H. J. (1998). *Cohen-Macaulay rings*. Cambridge University Press, Cambridge, UK.

[7] Buchberger, B. (1976). A theoretical basis for the reduction of polynomials to canonical forms. *SIGSAM Bull.*, 10(3):19–29.

[8] Butter, A., Kasieczka, G., Plehn, T., and Russell, M. (2018). Deep-learned top tagging with a lorentz layer. *SciPost Physics*, 5(3):28.

[9] Chong, K. F. E. (2020). A closer look at the approximation capabilities of neural networks.

[10] Cox, D., Little, J., and O'Shea, D. (2008). *Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra*. Springer, New York, NY, US.

[11] Derksen, H. and Kemper, G. (2002). *Computational invariant theory*. Springer, Berlin/Heidelberg, DE.

[12] Didelot, X., Everitt, R. G., Johansen, A. M., and Lawson, D. J. (2011). Likelihood-free estimation of model evidence. *Bayesian Analysis*, 6(1).

[13] Domokos, M. and Hegedus, P. (2000). Noether's bound for polynomial invariants of finite groups. *Archiv der Mathematik*, 74(3):161–167.

[14] Ehrenborg, R. and Rota, G. (1993). Apolarity and canonical forms for homogeneous polynomials. *European Journal of Combinatorics*, 14(3):157 – 181.

[15] Fulton, W., Harris, W., and Harris, J. (1991). *Representation theory: a first course*. Springer, New York, NY, US.

[16] Gauss, C. F. (2011). *Werke*, volume 3. Cambridge University Press, Cambridge, UK.

[17] Grayson, D. R. and Stillman, M. E. (1992). Macaulay2, a software system for research in algebraic geometry. Available at http://www.math.uiuc.edu/Macaulay2/.

[18] Gripaios, B., Haddadin, W., and Lester, C. G. (2021). Lorentz- and permutation-invariants of particles. *Journal of Physics A: Mathematical and Theoretical*, 54(15):155201.

[19] Gripaios, B., Müller, T., Parker, M. A., and Sutherland, D. (2014). Search strategies for top partners in composite higgs models. *Journal of High Energy Physics*, 2014(8):171.

[20] Haddadin, W. (2021). Invariant polynomials and machine learning.

[21] Handley, W. J., Hobson, M. P., and Lasenby, A. N. (2015a). polychord: nested sampling for cosmology. *Monthly Notices of the Royal Astronomical Society: Letters*, 450(1):L61–L65.

[22] Handley, W. J., Hobson, M. P., and Lasenby, A. N. (2015b). polychord: next-generation nested sampling. *Monthly Notices of the Royal Astronomical Society*, 453(4):4385–4399.

[23] Henning, B., Lu, X., Melia, T., and Murayama, H. (2017). Operator bases, S-matrices, and their partition functions. *Journal of High Energy Physics*, 10:199.

[24] Hochster, M. and Roberts, J. L. (1974). Rings of invariants of reductive groups acting on regular rings are cohen-macaulay. *Advances in Mathematics*, 13(2):115 – 175.

[25] Javid, K., Handley, W., Hobson, M., and Lasenby, A. (2020). Compromise-free bayesian neural networks.

[26] Kleiman, S. and Altman, A. (2013). *A term of commutative algebra*. Worldwide Center of Mathematics, LLC, Cambridge, MA, US.

[27] Komiske, P. T., Metodiev, E. M., and Thaler, J. (2020). Cutting multiparticle correlators down to size. *Phys. Rev. D*, 101:036019.

[28] Kotikov, A. (1991). Differential equations method: New technique for massive feynman diagram calculation. *Physics Letters B*, 254(1):158 – 164.

[29] Kraft, H. and Procesi, C. (1996). Classical invariant theory: a primer.

[30] Lester, C. G., Haddadin, W., and Gripaios, B. (2020). Lorentz and permutation invariants of particles iii: constraining non-standard sources of parity violation.

[31] Lester, C. G. and Schott, M. (2019). Testing non-standard sources of parity violation in jets at the lhc, trialled with cms open data. *Journal of High Energy Physics*, 2019(12).

[32] Liptrap, J. (2010). On translation invariant symmetric polynomials and haldane's conjecture.

[33] MacKay, D. J. C. (1992). A practical bayesian framework for backpropagation networks. *Neural Comput.*, 4(3):448–472.

[34] Pajer, E. (2021). Building a boostless bootstrap for the bispectrum. *Journal of Cosmology and Astroparticle Physics*, 2021(01):023–023.

[35] Paule, P. and Sturmfels, B. (2008). *Algorithms in invariant theory.* Springer, Vienna, AT.

[36] Pinkus, A. (1999). Approximation theory of the mlp model in neural networks. *Acta Numerica*, 8:143–195.

[37] Skilling, J. (2006). Nested sampling for general bayesian computation. *Bayesian Analysis*, 1(4).

[38] Stone, M. H. (1937). Applications of the theory of boolean rings to general topology. *Transactions of the American Mathematical Society*, 41(3):375–481.

[39] Stone, M. H. (1948). The generalized weierstrass approximation theorem. *Mathematics Magazine*, 21(4):167–184.

[40] Tkachov, F. (1981). A theorem on analytical calculability of 4-loop renormalization group functions. *Physics Letters B*, 100(1):65 – 68.

[41] Voigtlaender, F. (2020). The universal approximation theorem for complex-valued neural networks. *arXiv e-prints*.

[42] Weyl, H. (1966). *The classical groups: their invariants and representations.* Princeton University Press, Princeton, NJ, US.

[43] Yarotsky, D. (2018). Universal approximations of invariant maps by neural networks.

Appendix A

Hilbert Series

In this Appendix, we describe how to compute Hilbert series of invariant algebras.

A.1 Hilbert Series for finite groups

The Hilbert series of algebras invariant under finite groups are straightforward to compute and can be done using Molien's formula.

$$H(\mathbb{C}[V]^G, t) = \frac{1}{|G|} \sum_{g \in G} \frac{1}{\det_V(1 - t \cdot \rho_V)},$$

where $\rho_V : G \to GL(V)$ denotes the representation of G carried by V. Since the integrand is constant on a conjugacy class of G, one can simply compute the terms of this sum for a representative element from each conjugacy class and weight it by its respective multiplicity.

A.2 Hilbert Series for $O(d, \mathbb{C}) \times P$ and $E(d, \mathbb{C}) \times P$

Here, we describe how to compute Hilbert series of invariant algebras under the combined (complexified) Lorentz and permutation groups in dimension $2 \leq d \leq 4$. Using the isomorphism established in Chapter 3, the computations described below can also be used to obtain the Hilbert series of the Euclidean- and permutation-invariant algebras (although care has to be taken when $P = S_n$).

To do so, we use a generalisation of Molien's formula valid for a reductive group G, whereby the Hilbert series of an invariant algebra $\mathbb{C}[V]^G$ is given by [11]

$$H(\mathbb{C}[V]^G, t) = \int_C \frac{d\mu}{\det_V(1 - t \cdot \rho_V)},$$

where C is a maximal compact subgroup of G, $d\mu$ is a Haar measure on C normalised such that $\int_C d\mu = 1$, and $\rho_V : C \to GL(V)$ denotes the representation of C carried by V. Again, for what follows, it is useful to note that the integrand is constant within a conjugacy class of G.

We now consider in turn the cases of $d = 2, 3$, and 4, with an arbitrary number of momenta n and an arbitrary permutation group, $P \subset S_n$ acting on those momenta. The complexification of the Lorentz group when parity is a symmetry means that the groups we consider are of the form $G = O(d, \mathbb{C}) \times P$. For completeness, we also discuss the case where parity is not a symmetry, *i.e.* when $G = SO(d, \mathbb{C}) \times P$.

The case of $O(2, \mathbb{C}) \times P$

We start by considering the invariant algebra $\mathbb{C}[V]^G$ in the case of n momenta in 2 dimensions with no permutation symmetry which corresponds to $G = O(2, \mathbb{C})$ and $V \cong \mathbb{C}^{2n}$. The group $O(2, \mathbb{C})$ has maximal compact subgroup $O(2, \mathbb{R}) \cong U(1) \rtimes \mathbb{Z}_2$ and its action on $\mathbb{C}^2$ may be written as[1]

$$M^+(z) = \begin{pmatrix} z & 0 \\ 0 & z^{-1} \end{pmatrix}, M^-(z) = \begin{pmatrix} 0 & z^{-1} \\ z & 0 \end{pmatrix},$$

where $z \in \mathbb{C}$ such that $|z| = 1$ and where M^+ corresponds to the component connected to the identity and M^- corresponds to the other connected component. When acting on n copies of $\mathbb{C}^2$ (corresponding to n particles), we have

$$M_V^\pm = \begin{pmatrix} M^\pm & & & \\ & M^\pm & & \\ & & \ddots & \\ & & & M^\pm \end{pmatrix}.$$

The normalised Haar measure is given by $\frac{1}{2}\frac{1}{2\pi i}\frac{dz}{z}$ on each component (which is half the Haar measure for the group $U(1)$ and so takes into account the 2 disconnected

[1] If we consider $O(2, \mathbb{R}) \subset O(2, \mathbb{C})$ as acting on the real components of the momenta, then the isomorphism $O(2, \mathbb{R}) \cong U(1) \rtimes \mathbb{Z}_2$ corresponds to the linear map $(p_0, p_1) \in \mathbb{C}^2 \mapsto (p_0 + ip_1, p_0 - ip_1)$.

components). The Hilbert series is thus given by

$$H(\mathbb{C}[V^n]^{O(2)},t)=\frac{1}{2}\frac{1}{2\pi i}\oint_{|z|=1}\frac{dz}{z}\left(\frac{1}{\det_V(1-t\cdot M_V^+)}+\frac{1}{\det_V(1-t\cdot M_V^-)}\right).$$

For our example of $n=3$ with $P=1$, the integral becomes

$$\begin{aligned}H(\mathbb{C}[V^3]^{O(2)},t)&=\frac{1}{2}\frac{1}{2\pi i}\oint_{|z|=1}\frac{dz}{z}\left(\frac{1}{(1-tz)^3(1-t/z)^3}+\frac{1}{(1-t^2)^3}\right)\\&=\frac{1}{2}\left(\frac{1+4t^2+t^4}{(1-t^2)^5}+\frac{1}{(1-t^2)^3}\right)=\frac{1+t^2+t^4}{(1-t^2)^5}.\end{aligned}$$

where the integrals have been carried out using the residue theorem of contour integration.

We now include some permutation group $P\subseteq S_n$ acting on the n momenta so that the combined group becomes $G=O(2,\mathbb{C})\times P$ and its maximal compact subgroup is just $O(2,\mathbb{R})\times P$. Here, one must additionally average over the permutation group P, where the action of P simply permutes the n particles, ergo the n copies of $\mathbb{C}^2$. Since the integrand is constant within conjugacy classes, it suffices to pick one representative element from each class, and weight accordingly. The Haar measure is rescaled by $1/|P|$ so that it is still properly normalised.

For our example of $n=3$ with $P=S_3$, we have 3 conjugacy classes: the identity with multiplicity 1, $(\cdot\cdot)$ with multiplicity 3, and $(\cdot\cdot\cdot)$ with multiplicity 2. We use the following representative elements from each permutation conjugacy class

$$\begin{pmatrix}M^\pm&0&0\\0&M^\pm&0\\0&0&M^\pm\end{pmatrix},\begin{pmatrix}0&M^\pm&0\\M^\pm&0&0\\0&0&M^\pm\end{pmatrix},\begin{pmatrix}0&M^\pm&0\\0&0&M^\pm\\M^\pm&0&0\end{pmatrix}.$$

The contribution of the component connected to the identity then becomes

$$\begin{aligned}&H^+(\mathbb{C}[V^3]^{O(2)\times S_3},t)\\&=\frac{1}{6}\frac{1}{2}\frac{1}{2\pi i}\oint_{|z|=1}\frac{dz}{z}\left(\frac{1}{(1-tz)^3(1-t/z)^3}+\frac{3}{(1-tz)(1-t/z)(1-(tz)^2)(1-(t/z)^2)}+\frac{2}{(1-(tz)^3)(1-(t/z)^3)}\right)\\&=\frac{1}{2}\frac{1+3t^4+4t^6+3t^8+t^{12}}{(1-t^2)^2(1-t^4)^2(1-t^6)}.\end{aligned}$$

Similarly, the contribution of the other connected component is

$$H^-(\mathbb{C}[V^3]^{O(2)\times S_3},t)=\frac{1}{2}\frac{1+t^4}{(1-t^2)^2(1-t^6)},$$

and so finally we obtain

$$\begin{aligned} H(\mathbb{C}[V^3]^{O(2)\times S_3}, t) &= H^+(\mathbb{C}[V]^{O(2)\times S_3}, t) + H^-(\mathbb{C}[V]^{O(2)\times S_3}, t) \\ &= \frac{1 + t^4 + 2t^6 + t^8 + t^{12}}{(1-t^2)^2(1-t^4)^2(1-t^6)}. \end{aligned}$$

Notice that we also get the Hilbert series for the case $G = SO(2,\mathbb{C}) \times P$, corresponding to when parity is not a symmetry, for free, by just considering the component connected to the identity

$$\begin{aligned} H(\mathbb{C}[V^3]^{SO(2)}, t) &= \frac{1 + 4t^2 + t^4}{(1-t^2)^5}, \\ H(\mathbb{C}[V^3]^{SO(2)\times S_3}, t) &= \frac{1 + 3t^4 + 4t^6 + 3t^8 + t^{12}}{(1-t^2)^2(1-t^4)^2(1-t^6)}. \end{aligned}$$

The case of $O(3,\mathbb{C}) \times P$

In $d = 3$, the group $O(3,\mathbb{C})$ has maximal compact subgroup $O(3,\mathbb{R}) \cong (SU(2)/\mathbb{Z}_2)\times\mathbb{Z}_2$. Since the integrand is constant on the conjugacy classes, we need consider only the maximal torus of $SU(2)$ with elements

$$\begin{pmatrix} z & 0 \\ 0 & z^{-1} \end{pmatrix},$$

where $|z| = 1$[2] acting on $\mathbb{C}^3$ as[3]

$$M^+(z) = \begin{pmatrix} z^2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & z^{-2} \end{pmatrix}, M^-(z) = \begin{pmatrix} -z^2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -z^{-2} \end{pmatrix},$$

where $\pm$ again distinguishes the 2 connected components. The normalised Haar measure on each component is $\frac{1}{2}\frac{1}{2\pi i}\frac{(1-z^2)dz}{z}$ (which is just half of the usual normalised Haar measure for $SU(2)$). The Hilbert series with n particles is then given by

$$H(\mathbb{C}[V^n]^{O(3)}, t) = \frac{1}{2}\frac{1}{2\pi i}\oint_{|z|=1} \frac{(1-z^2)dz}{z}\left(\frac{1}{\det_V(1 - t\cdot M_V^+)} + \frac{1}{\det_V(1 - t\cdot M_V^-)}\right).$$

[2]Strictly speaking, one should consider only half of the unit circle, since z and $-z$ yield the same element in $SU(2)/\mathbb{Z}_2$. But since the integral will turn out to be symmetric under $z \to -z$, we can get away with integrating over the whole circle.

[3]Here, the isomorphism $O(3,\mathbb{R}) \cong (SU(2)/\mathbb{Z}_2) \times \mathbb{Z}_2$ corresponds to the linear map $(p_0, p_1, p_2) \in \mathbb{C}^3 \mapsto (p_0 - ip_1, p_2, p_0 + ip_1)$.

For example, with $n = 4$ the integral becomes

$$H(\mathbb{C}[V^4]^{O(3)}, t) =$$
$$\frac{1}{2}\frac{1}{2\pi i}\oint_{|z|=1} \frac{(1-z^2)dz}{z}\left(\frac{1}{(1-t)^4\left(1-\frac{t}{z^2}\right)^4(1-tz^2)^4} + \frac{1}{(1+t)^4\left(1+\frac{t}{z^2}\right)^4(1+tz^2)^4}\right),$$

which we evaluate using the residue theorem, obtaining

$$H(\mathbb{C}[V^4]^{O(3)}, t) = \frac{1}{2}\left(\frac{1+t^2+4t^3+t^4+t^6}{(1-t^2)^9} + \frac{1+t^2-4t^3+t^4+t^6}{(1-t^2)^9}\right) = \frac{1+t^2+t^4+t^6}{(1-t^2)^9}.$$

We also obtain the Hilbert series for when $G = SO(3, \mathbb{C})$ for free by only considering the component connected to the identity

$$H(\mathbb{C}[V^4]^{SO(3)}, t) = \frac{1+t^2+4t^3+t^4+t^6}{(1-t^2)^9}.$$

To include an arbitrary permutation group $P \subseteq S_n$ acting on the n momenta, one needs to average over the conjugacy classes of P as discussed previously.

The case of $O(4, \mathbb{C}) \times P$

In $d = 4$, $O(4, \mathbb{C})$ has maximal compact subgroup $O(4, \mathbb{R}) \cong ((SU(2) \times SU(2))/\mathbb{Z}_2) \rtimes \mathbb{Z}_2$, where the automorphism in the semi-direct product corresponds to interchanging the 2 $SU(2)$ factors. Since the integrand is constant on the conjugacy classes, we need consider only the maximal torus with elements

$$\left(\begin{pmatrix} z & 0 \\ 0 & z^{-1} \end{pmatrix}, \begin{pmatrix} w & 0 \\ 0 & w^{-1} \end{pmatrix}\right),$$

where $|z| = |w| = 1$.[4] The action on $\mathbb{C}^4$ is given by[5][6]

$$M^+(z,w) = \begin{pmatrix} zw & 0 & 0 & 0 \\ 0 & zw^{-1} & 0 & 0 \\ 0 & 0 & wz^{-1} & 0 \\ 0 & 0 & 0 & (zw)^{-1} \end{pmatrix}, M^-(z) = \begin{pmatrix} z & 0 & 0 & 0 \\ 0 & 0 & z & 0 \\ 0 & z^{-1} & 0 & 0 \\ 0 & 0 & 0 & z^{-1} \end{pmatrix},$$

The normalised Haar measure on the component connected to the identity is $\frac{1}{2}\frac{1}{(2\pi i)^2}\frac{(1-z^2)dz}{z}\frac{(1-w^2)dw}{w}$ and the Haar measure on the disconnected component is $\frac{1}{2}\frac{1}{2\pi i}\frac{(1-z^2)dz}{z}$. The Hilbert series with n particles is then given by

$$\begin{aligned} H(\mathbb{C}[V^n]^{O(4)}, t) =& \frac{1}{2}\frac{1}{(2\pi i)^2}\oint_{|z|=|w|=1} \frac{(1-z^2)(1-w^2)dzdw}{zw}\frac{1}{\det_V(1-t\cdot M_V^+)} \\ &+ \frac{1}{2}\frac{1}{2\pi i}\oint_{|z|=1}\frac{(1-z^2)dz}{z}\frac{1}{\det_V(1-t\cdot M_V^-)}. \end{aligned}$$

In our example of $n = 5$ with $P = 1$, the integral becomes

$$\begin{aligned} H(\mathbb{C}[V^5]^{O(4)}, t) =& \frac{1}{2}\frac{1}{(2\pi i)^2}\oint_{|z|=|w|=1}\frac{dzdw}{zw}\frac{(1-z^2)(1-w^2)}{(1-t/(wz))^5(1-(tw)/z)^5(1-(tz)/w)^5(1-twz)^5} \\ &+ \frac{1}{2}\frac{1}{2\pi i}\oint_{|z|=1}\frac{dz}{z}\frac{(1-z^2)}{(1-t^2)^5(1-t/z)^5(1-tz)^5}, \end{aligned}$$

which we evaluate using the residue theorem, obtaining

$$H(\mathbb{C}[V^5]^{O(4)}, t) = \frac{1}{2}\left(\frac{1+t^2+6t^4+t^6+t^8}{(1-t^2)^{14}} + \frac{1+3t^2+t^4}{(1-t^2)^{12}}\right) = \frac{1+t^2+t^4+t^6+t^8}{(1-t^2)^{14}}.$$

We also obtain the Hilbert series for when $G = SO(4,\mathbb{C})$ for free by only considering the component connected to the identity

$$H(\mathbb{C}[V^5]^{SO(4)}, t) = \frac{1+t^2+6t^4+t^6+t^8}{(1-t^2)^{14}}.$$

To include an arbitrary permutation group $P \subseteq S_n$ acting on the n momenta, one again needs to average over the conjugacy classes of P as discussed previously.

[4] As in $d = 3$, there is no need to take care in projecting to $(SU(2) \times SU(2))/\mathbb{Z}_2$.

[5] The asymmetry in the formulæ arises from the fact that the conjugacy classes in the disconnected component can be parameterized by a single $U(1)$; for details see [19].

[6] Here, the isomorphism $O(4,\mathbb{R}) \cong ((SU(2) \times SU(2))/\mathbb{Z}_2) \rtimes \mathbb{Z}_2$ corresponds to the linear map $(p_0, p_1, p_2, p_4) \in \mathbb{C}^4 \mapsto (p_0 + ip_3, p_1 + ip_2, p_1 - ip_2, p_0 - ip_3)$.

A.3 Explicit list of Hilbert Series

Algebra	Hilbert Series
$\mathbb{C}[y_{ij}, i < j \leq 4]$	$\frac{1}{(1-t)^6}$
$\mathbb{C}[y_{ij}, i < j \leq 4]^{S_2 \times S_2}$	$\frac{1+t^3}{(1-t)^3(1-t^2)^3}$
$\mathbb{C}[y_{ij}, i < j \leq 5]$	$\frac{1}{(1-t)^{10}}$
$\mathbb{C}[y_{ij}, i < j \leq 5]^{S_2 \times S_3}$	$\frac{1+t^2+6t^3+8t^4+6t^5+12t^6+14t^7+9t^8+8t^9+5t^{10}+2t^{11}}{(1-t)^3(1-t^2)^4(1-t^3)^2(1-t^6)}$
$\mathbb{C}[y_{ij}, i \leq j \leq 4]^{S_2}$	$\frac{1+3t^2}{(1-t)^7(1-t^2)^3}$
$\mathbb{C}[y_{ij}, i \leq j \leq 4]^{S_2 \times S_2}$	$\frac{1+2t^2+4t^3+t^4}{(1-t)^5(1-t^2)^5}$
$\mathbb{C}[y_{ij}, i \leq j \leq 4]^{S_3}$	$\frac{1+3t^2+7t^3+6t^4+6t^5+10t^6+3t^7}{(1-t)^4(1-t^2)^3(1-t^3)^3)}$
$\mathbb{C}[V^4]^{SO(4)}$	$\frac{1+t^4}{(1-t^2)^{10}}$
$\mathbb{C}[V^4]^{SO(4) \times S_2}$	$\frac{1+3t^4+3t^6+t^{10}}{(1-t^2)^7(1-t^4)^3}$
$\mathbb{C}[V^4]^{SO(4) \times S_2 \times S_2}$	$\frac{1+2t^4+5t^6+5t^8+2t^{10}+t^{14}}{(1-t^2)^5(1-t^4)^5}$
$\mathbb{C}[V^4]^{SO(4) \times S_3}$	$\frac{1+3t^4+7t^6+9t^8+16t^{10}+16t^{12}+9t^{14}+7t^{16}+3t^{18}+t^{22}}{(1-t^2)^4(1-t^4)^3(1-t^6)^3}$
$\mathbb{C}[V^5]^{SO(4)}$	$\frac{1+t^2+6t^4+t^6+t^8}{(1-t^2)^{14}}$

Table A.1 The Hilbert series of some relevant invariant algebras.

Appendix B

Restriction of the Stone-Weierstrass theorem

In this Appendix, we discuss two approaches to a full proof of the restriction of the Stone-Weierstrass theorem to continuous invariant functions.

B.1 Via a "Reynolds operator"

The first method proceeds by a direct extension of the proof of Theorem 5.1.2. The crucial step in that proof is the integral over a maximal compact subgroup using a Haar measure. For algebras of invariant polynomials, this integral is essentially the projection by a Reynolds operator, which we know to exist for algebras invariant under any linearly reductive group. Therefore a possible generalisation of this approach can be achieved by establishing the existence of an equivalent projection operator for the $C^\star$-algebra of continuous complex-valued functions $\pi : C(X) \twoheadrightarrow C^G(X)$ (whose restricted action on the subalgebra of invariant polynomials is as required $\pi| = \mathcal{R}_G : \mathbb{C}[X] \twoheadrightarrow \mathbb{C}[X]^G$).

The main bottleneck here is that it is not entirely clear whether the equivalent of the Reynolds operator exists. The arguments used to show the existence of the Reynolds operator for algebras of invariant polynomials relies on a special decomposition of the induced representation on these polynomials which is not readily extended to the $C^\star$-algebra of continuous functions.

B.2 Via separating subalgebras

The second method leverages the all-powerful Stone-Weierstrass theorem to obtain a proof. The general idea (sketched below in more detail) relies on reducing the problem from the consideration of invariant functions defined on X to all functions defined on the orbit space X/G, thus allowing us to apply the Stone-Weierstrass directly to establish the density of $\mathbb{C}[X]^G$ in $C^G(X)$. The author would like to thank Gerald Folland for his help and insights with the following discussion.

Our starting point will be the full Stone-Weierstrass theorem.

Theorem B.2.1 (Stone-Weierstrass theorem)**.** *Let X be a compact Hausdorff space and denote by $C(X)$ the $C^\star$ Banach algebra of continuous complex-valued functions, $f : X \to \mathbb{C}$ where $C(X)$ separates X. Let $A \subset C(X)$ be a subalgebra (closed under the unit and algebra multiplication operations on $C(X)$). A subalgebra inclusion $A \subset C(X)$ is dense if and only if it also separates X.*

There is a lot to unravel here. We will give a concise overview of the required background mathematical definitions to understand the intricacies present.

Definition B.2.1 (Topological Space)**.** *A topological space X is a set X equipped with a set of subsets $U \subset X$ called the open sets which are closed under finite intersections and arbitrary unions.*

A consequence of this is that the empty set $\emptyset \subset U \subset X$.

Definition B.2.2 (Open cover)**.** *An open cover of a topological space X is a collection $\{U_i \subset X\}$ of open subsets of X whose union equals X: $\cup_i U_i = X$.*

Definition B.2.3 (Compact topological space)**.** *A topological space is called compact if every open cover has a finite subcover.*

Definition B.2.4 (Closed set)**.** *Let X be a topological space. A subset $S \subset X$ is called a closed set if its complement $X \setminus S$ is an open subset.*

Definition B.2.5 (Open neighbourhood)**.** *Let X be a topological space and $x \in X$ a point. Then*

- *A subset $U \subset X$ is a neighbourhood of x if there exists an open subset $O \subset X$ such that $x \in O$ and $O \subset U$.*
- *U is an open neighbourhood if it is also open.*

Definition B.2.6 (Hausdorff space). *Let X be a topological space. X is called Hausdorff if given $x, y \in X$, with $x \neq y$, there exists open neighbourhoods U of x and V of y in X that are disjoint.*

Definition B.2.7 (Banach space). *A Banach space is a vector space V equipped with a complete*[1] *norm $|| - || : V \to \mathbb{R}$ such that*

- $||0|| \leq 0$
- $||rv|| = |r| \; ||v||, \forall r \in \mathbb{R}, v \in V$
- $||v + w|| \leq ||v|| + ||w||, \forall v, w \in V$

Definition B.2.8 (Banach algebra). *A Banach algebra is a Banach space A equipped with a bilinear (associative) multiplication map $m : A \times A \to A$, such that for any $a, b \in A$, $||m(a, b)|| \leq ||a|| \; ||b||$.*

Definition B.2.9 ($C^\star$-algebra). *A $C^\star$-algebra is a Banach algebra $(A, || - ||)$ over $\mathbb{C}$ equipped with an involution $(-)^\star$ ($a^{\star\star} = (a^\star)^\star = a$) compatible with complex conjugation where $(a + b)^\star = a^\star + b^\star$ and $(ab)^\star = b^\star a^\star$ for all $a, b \in A$ that satisfies the $C^\star$-identity*

$$||aa^\star|| = ||a||^2, \forall a \in A$$

In the SW theorem, the complete norm of the $C^\star$ Banach algebra $C(X)$ is the sup-norm $||f|| = \sup_{x \in X} |f(x)|$ and so from now on $||f||$ will denote the sup-norm of f.

Definition B.2.10 (Separating algebra). *We say that an algebra A separates a space X if for any distinct $x, y \in X$, there exists an $f \in A$ such that $f(x) \neq f(y)$.*

Now we move onto considering a group G acting regularly on a compact Hausdorff space X, $G \times X \to X$.

Definition B.2.11 (Orbit). *For an $x \in X$, the set $Gx = \{g \circ x | g \in G\}$ is called the G-orbit of x. We denote the set of all orbits by X/G.*

Definition B.2.12 (Orbit space). *Let $P : X \to X/G$ be the map that takes each $x \in X$ to its orbit. The topology on X/G is the quotient topology: a set in X/G is open if its inverse image under P is open in X. With this topology, P is a continuous map which is furthermore surjective by definition.*

[1] A norm is complete if given any infinite sequence $(v_1, v_2, \dots)$ such that $\lim_{m,n\to\infty} || \sum_{i=m}^{m+n} v_i || = 0$, there exists a sum S such that $\lim_{n\to\infty} ||S - \sum_{i=1}^{n} v_i|| = 0$; we write $S = \sum_{i=1}^{\infty} v_i$.

Consider the $C^\star$ Banach (sub)algebra of invariant continuous complex-valued functions $C^G(X), f : X \to \mathbb{C}$ and consider the $C^\star$ Banach algebra of continuous complex-valued functions $C(X/G), f_G : X/G \to \mathbb{C}$. We wish to show that there is a one-to-one correspondence between these two algebras, or more precisely an isomorphism in the category **C⋆Alg**. That is, a isomorphism that preserves the $C^\star$ structure.

Proposition B.2.1. *The $C^\star$ algebras $C^G(X)$ and $C(X/G)$ are isomorphic with isomorphism $i : C(X/G) \xrightarrow{\sim} C^G(X) : f_G \mapsto f_G \circ P$.*

Proof. First, we have that every invariant continuous function $f \in C^G(X)$ descends to a unique continuous function in $f_G \in C(X/G)$. To show existence, consider $x, y \in X$ where $x = g \circ y$. We have that $f(x) = f(y)$ because f is invariant. Define $f_G : X/G \to \mathbb{C}$ to be the map that takes $Gx \mapsto f(x)$. To show continuity, let $U \subset \mathbb{C}$ be an open subset. Then $(f_G \circ P)^{-1}(U) = f^{-1}(U)$ which is open. Therefore, $P^{-1}(f_G^{-1}(U))$ is open and so because P is an open map, $f_G^{-1}(U)$ is open and therefore f_G is continuous. To show uniqueness, let f_G be a continuous map such that $f = f_G \circ P$. For some $x \in X$, $f_G \circ P(x) = f_G(Gx) = f(x)$, which specifies the map completely.

Hence, for each $f \in C^G(X)$ there is a unique $f_G \in C(X/G)$ such that the following triangle diagram commutes.

$$\begin{tikzcd} X \arrow[r, "P"] \arrow[dr, "f"'] & X/G \arrow[d, "f_G"] \\ & \mathbb{C} \end{tikzcd}$$

Going the other way is easy: every function $f_G \in C(X/G)$ can be composed with P to give a function $f \in C^G(X)$. Hence, we have an isomorphism of sets. Finally, we need to show that we indeed have an isomorphism in the category **C⋆Alg**, *i.e.* that the isomorphism preserves the $C^\star$-identity. For any $f_G \in C(X/G)$, the $C^\star$-identity tells us that $||f_G f_G^\star|| = ||f_G||^2$. The image of f_G under the isomorphism is $f = f_G \circ P$. So, we have that $||ff^\star|| = \sup_{x\in X}|f(x)f^\star(x)| = \sup_{x\in X}|f_G(Gx)f_G^\star(Gx)| = (\sup_{x\in X}|f_G(Gx)|)^2 = (\sup_{x\in X}|f(x)|)^2 = ||f||^2$, where in the last three steps we used the fact that f_G obeys the $C^\star$ identity, that $f = f_G \circ P$, and the definition of the norm. Hence, the isomorphism preserves the $C^\star$ identity. $\square$

We now move onto discussing the obit space X/G.

Proposition B.2.2. *Let X (henceforth) be a compact Hausdorff space and G be a group acting on X, $G \times X \to X$. The orbit space X/G is also compact.*

Proof. X/G is the image of a compact space X under a continuous map P so it is compact. □

In general, even if X is Hausdorff (and G is a topological group that is also Hausdorff), the orbit space X/G might not be Hausdorff. But, if we have that $C(X/G)$ separates X/G, then we can prove it as follows

Proposition B.2.3. *Consider $C^\star$-algebra of continuous complex-valued functions $C(Y)$. If $C(Y)$ separates Y, then Y is Hausdorff.*

Proof. Since $C(Y)$ separates Y, for any two points y_1, y_2 one can always find some $f \in C(Y)$ where $f(y_1) \neq f(y_2)$. But, if $f \in C(Y)$ takes different values $c_1, c_2 \in \mathbb{C}$ on points $y_1, y_2 \in Y$, and D_1, D_2 are disjoint open sets in $\mathbb{C}$ containing c_1, c_2 respectively (which is always possible because $\mathbb{C}$ is Hausdorff), then, since f is also continuous, $f^{-1}(I_1)$ and $f^{-1}(I_2)$ are disjoint open sets in Y containing y_1 and y_2. Hence, Y is Hausdorff. □

Therefore, if $C(X/G)$ separates X/G, then X/G is Hausdorff. But, in general it might not be separating. In such cases we construct a coarser quotient space as follows

Definition B.2.13. *Consider the orbit space X/G and the $C^\star$ Banach algebra $C(X/G)$. Define two orbits $O_1, O_2 \in X/G$ to be equivalent, $O_1 \sim O_2$ if for all $f \in C(X/G)$, $f(O_1) = f(O_2)$. Denote the space of equivalence classes by $Y = X/G/\sim$ giving it the quotient topology and we let the continuous projection map that takes each orbit to its equivalence class by $q : X/G \to Y$.*

Proposition B.2.4. *Y is compact Hausdorff.*

Proof: Y is the image of a compact space X/G under a continuous map q so it is compact. Furthermore, $C(Y)$ separates Y by definition and so by the proposition above, Y is also Hausdorff. □

Proposition B.2.5. *The $C^\star$ algebras $C^G(X)$ and $C(Y)$ are isomorphic with isomorphism $\bar{i} : C(Y) \xrightarrow{\sim} C^G(X) : f_Y \mapsto f_Y \circ q \circ P$.*

Proof. To do this, we first establish an isomorphism between $C(Y)$ and $C(X/G)$ with isomorphism $j : C(Y) \xrightarrow{\sim} C(X/G) : f_Y \mapsto f_Y \circ q$. This follows in exactly the same way as in Proposition B.2.1. Then, we compose this isomorphism with the isomorphism $i : C(X/G) \xrightarrow{\sim} C^G(X) : f_G \mapsto f_G \circ P$ to finally get $\bar{i} = i \circ j : C(Y) \xrightarrow{\sim} C^G(X)$ as required. □

Now, all we need to do is to apply the Stone-Weierstrass theorem to the space Y. This gives us

Theorem B.2.2. *Let X be a compact Hausdorff space and G be a group with a group action on X, $G \times X \to X$. Denote by $C^G(X)$ the $C^\star$ Banach algebra of invariant continuous complex-valued functions, $f : X \to \mathbb{C}$. Let $A \subset C^G(X)$ be a subalgebra (closed under the unit and algebra multiplication operations on $C^G(X)$). A subalgebra inclusion $A \subset C^G(X)$ is dense if and only if it separates $Y = X/G/\sim$.*

All that remains is to characterise the space of equivalence classes Y and show that the subalgebra of invariant polynomials $\mathbb{C}[X]^G$ separates Y.

There are two main bottlenecks to seeing this proof through to completion. First, it is difficult to check whether Y is isomorphic to X/G. In other words, does $C(X/G)$ separate X/G? But say that one can establish which space is the correct one to consider. The second bottleneck is then showing that the subalgebra of invariant polynomials $\mathbb{C}[X]^G$ separates X/G or Y. This is in general hard. Furthermore there are known examples where the subalgebra of invariant polynomials fails to separate the orbits.

We hope that this is a motivating example, where a generalisation of the Stone-Weierstrass theorem is necessary, that may encourage the expert (and interested) reader to investigate further.

Appendix C

Nested sampling

In this Appendix, we give a brief summary of nested sampling. Consider a model M with parameters α. Denote the likelihood of the data given the parameters by $L(\alpha) = P(D|\alpha, M)$ and the prior on these parameters by $\pi(\alpha) = P(\alpha|M)$. The posterior is then $P(\alpha|D, M) \propto L(\alpha)\pi(\alpha)$.

The constant of proportionality of this relation is called the evidence, $Z(M)$. Computing $Z(M)$ amounts to evaluating the (usually) very high dimensional integral

$$Z(M) = \int L(\alpha)\pi(\alpha)d\alpha.$$

The idea of nested sampling is to re-express this multi-dimensional integral as a one-dimensional one using a change of variables. Define $X(\lambda) = \int_{L(\alpha)>\lambda} \pi(\alpha)d\alpha$ to be the fraction of the prior (or prior volume) contained within an iso-likelihood contour $L(\alpha) = \lambda$. This is a decreasing positive semi-definite function. One may now write $Z(M)$, using a change of variable $L(\alpha)$ to $L(X)$, as

$$Z(M) = \int L(X)dX.$$

Nested sampling evaluates this integral by sampling a set of n_{live} *live points* from the prior, $\{\alpha_i^{\text{live}}\}$, each with a likelihood $L_i = L(\alpha_i^{\text{live}})$ and $X_i = X(L_i)$. These points are then iteratively modified to converge exponentially onto the peaks of $L(X)$. The evidence can then be computed as a simple Riemann sum

$$Z(M) = \sum_{i=1}^{n_{\text{live}}-1} (X_{i+1} - X_i)L_i.$$

For excellent overviews which deal with the many details that we have skimmed over, refer to the original nested sampling paper [37] and to the PolyChord package papers [21, 22].

Index