BPS states, conserved charges and centres of symmetric group algebras

Garreth Kemp[a] **and Sanjaye Ramgoolam**[b,c]

[a]*Department of Mathematics and Applied Mathematics,*
University of Johannesburg, Auckland Park, 2006, South Africa

[b]*Centre for Research in String Theory, School of Physics and Astronomy,*
Queen Mary University of London, Mile End Road, London E1 4NS, U.K.

[c]*National Institute for Theoretical Physics,*
School of Physics and Mandelstam Institute for Theoretical Physics,
University of the Witwatersrand, Wits, 2050, South Africa

E-mail: gkemp@uj.ac.za, s.ramgoolam@qmul.ac.uk

ABSTRACT: In $\mathcal{N} = 4$ SYM with U(N) gauge symmetry, the multiplicity of half-BPS states with fixed dimension can be labelled by Young diagrams and can be distinguished using conserved charges corresponding to Casimirs of U(N). The information theoretic study of LLM geometries and superstars in the dual $AdS_5 \times S^5$ background has raised a number of questions about the distinguishability of Young diagrams when a finite set of Casimirs are known. Using Schur-Weyl duality relations between unitary groups and symmetric groups, these questions translate into structural questions about the centres of symmetric group algebras. We obtain algebraic and computational results about these structural properties and related Shannon entropies, and generate associated number sequences. A characterization of Young diagrams in terms of content distribution functions relates these number sequences to diophantine equations. These content distribution functions can be visualized as connected, segmented, open strings in content space.

https://doi.org/10.1007/JHEP01(2020)146

Contents

1 Introduction

One of the best understood instances of the AdS/CFT correspondence [1] is the duality between $\mathcal{N} = 4$ SYM with U(N) gauge group and type IIB superstring theory on $AdS_5 \times S^5$ with N units of five-form flux. In particular, in the half-BPS sector, giant gravitons [2] were identified as important non-perturbative objects in the string theory which demonstrate remarkable sensitivity to finite N effects, notably the stringy exclusion principle [3], in their classical properties. Sub-determinant operators in the CFT were identified as duals for an interesting class of giant gravitons [4]. The construction of CFT duals of general giant gravitons was obtained by using Young diagrams to organize a finite N orthogonal basis of CFT operators [5]. An underlying free-fermion description of this sector was identified [5, 6].

Recent computations have been able to reproduce giant graviton correlators from the CFT using calculations in space-time [7–10]

Large numbers of giant gravitons back-react on the space-time to produce new geometries with $AdS_5 \times S^5$ asymptotics. The general smooth half-BPS solutions with these asymptotics were foundby Lin,Lunin and Maldacena (LLM) [11]. The LLM picture can also be used to describe singular geometries called superstars and their relation to giant gravitons [13–15]. An important feature of the half-BPS sector is the existence of conserved charges, related to the Casimirs of $u(N)$ [5]. The interpretation of these charges in the space-time picture of LLM geometries and superstars was discussed in [12]. They were found to correspond to multi-pole moments of the gravitational solutions. It was also argued that measuring all the N independent Casimir eigenvalues needed to determine a geometry requires measurement of the variations of the metric over distances shorter than the Planck distance. This was proposed as an important mechanism of information loss, where the semiclassical observer fails to identify the exact quantum state corresponding to a given geometry. A precise and general mathematical understanding of the limitations of information accessible to the semi-classical observer in gravity can be expected to have deep implications in black hole physics. More recent discussions on this theme include [16–19].

The generalization from half-BPS to sectors with lower symmetry is a problem of multi-matrix quantum field theory combinatorics. Constructions of multi-matrix gauge invariant operators with multiple Young diagrams were given in [20, 21]. Bases for multi-matrix gauge invariant operators, which diagonalise the free field theory inner products, were found in [22–26], by using symmetric groups to organize the gauge invariants. The key reason for symmetric groups entering these questions about U(N) gauge invariants is Schur-Weyl duality, which relates the action of U(N) on the n-fold tensor product of the fundamental $V^{\otimes n}$ to the action of S_n - the symmetric group of all permutations of n distinct objects. These multi-matrix orthogonal bases included bases covariant under the global symmetry group, e.g. U(2) for the quarter-BPS and U(3) for the eighth-BPS sector [23, 24], as well as restricted Schur bases with quantum numbers associated with subgroups of S_n [25, 26]. It was found that the different bases amounted to diagonalizing different sets of "generalized Casimir operators" associated with different kinds of enhanced symmetries in the free field limit [27]. The goal of further understanding the structure of the diagonal bases for multi-matrix systems, and associated expressions for correlators, prompted a systematic study of the structure of permutation algebras related to these bases. The Wedderburn-Artin decomposition of semi-simple algebras into matrix blocks [29, 30] was found to illuminate the properties of the permutation algebras underlying multi-matrix orthogonal bases and correlators. These semi-simple algebras are also related to the formulation of the combinatorics of multi-matrix gauge invariants under classical Lie group gauge symmetries in terms of two dimensional topological field theories [31, 32]. This matrix block structure of permutation algebras underlying the construction of gauge invariants is also applicable to gauge invariants in tensor models [36, 37]. Closely related discussions of tensor model correlators are in [38–42]. In particular, Casimir operators have also been studied in the guise of cut-and-join operators in matrix and tensor models, see for example [41] and references therein.

In this paper, motivated by the discussion in [12] and the subsequent developments in the mathematics of the space of gauge invariant operators - particularly the relevance of the structure theory of permutation algebras - we initiate a systematic study of the quantitative characterisation of the uncertainty in the determination of Young diagram operators in the half-BPS sector, when a finite set of Casimirs is specified. In section 2 we review some key elements of the connections between BPS operators of dimension n, Casimir operators of U(N), and the symmetric group S_n of all permutations of n distinct objects. The group algebra $\mathbb{C}(S_n)$ of formal linear combinations of S_n group elements with complex coefficients plays an important role, along with the subspace of this group algebra which commutes with all $\mathbb{C}(S_n)$. This subspace is a commutative sub-algebra called the centre of $\mathbb{C}(S_n)$, or the central algebra, and denoted $\mathcal{Z}(\mathbb{C}(S_n))$. The eigenvalues of Casimirs of U(N) are related to the normalized characters of central elements in $\mathcal{Z}(\mathbb{C}(S_n))$.

In section 3, we consider two linear bases for $\mathcal{Z}(\mathbb{C}(S_n))$: one corresponds to conjugacy classes of S_n and another to irreducible representations. As is well-known the conjugacy classes correspond to cycle structures of permutations. Thus $\mathcal{Z}(\mathbb{C}(S_n))$ is a vector space of dimension equal to the number of partitions of n, denoted $p(n)$, with a commutative and associative product. A distinguished set of conjugacy classes correspond to permutations have a single cycle of length k: the corresponding central element is denoted T_k. We prove that for any n, the set $\mathcal{G}_n = \{T_2, T_3, \cdots, T_n\}$ form a generating subspace of the central algebra. This means that by taking linear combinations of these elements and their products, we can get any element of $\mathcal{Z}(\mathbb{C}(S_n))$. In fact, for a fixed n, we generically only need a subset of $\mathcal{G}_n$ to generate the central algebra. The connection between cycle structures and irreps, which may be viewed as a Fourier transform, leads to a formulation of the distinguishability of Young diagrams in terms of minimal generating subspaces $\mathcal{Z}(\mathbb{C}(S_n))$. A simple inspection of normalized characters of the cycle operators in irreps shows, for example, that for n up to 5 and $n = 7$, but not $n = 6$, T_2 alone suffices to generate the centre: in other words, T_2 and its powers form a linear basis for $\mathcal{Z}(\mathbb{C}(S_n))$. This is demonstrated directly by writing out the powers of T_2 in terms of linear combinations of central elements corresponding to the conjugacy classes.

In section 4, we investigate the dimensions of the subspaces of $\mathcal{Z}(\mathbb{C}(S_n))$ generated by T_2, by T_3 and by the pair $\{T_2, T_3\}$. These dimensions, as shown in section 3, are given respectively by the number of distinct normalized characters $\frac{\chi_R(T_2)}{d_R}$, the number of distinct $\frac{\chi_R(T_3)}{d_R}$, and the number of distinct pairs $\{\frac{\chi_R(T_2)}{d_R}, \frac{\chi_R(T_3)}{d_R}\}$, as R runs over the set of Young diagrams. In each case, for small enough n, there are no degeneracies as R runs over all the Young diagrams. However as n increases, one or more R give the same normalized character, or list of normalized characters. The distribution of degeneracies can be used to define a probability distribution over the space of possible normalized characters. For each fixed value or list of values, the Shannon entropy - which is the logarithm of the multiplicity - gives a measure of the uncertainty associated with having knowledge of the value or value sets but not the exact identity of the Young diagram. Depending on a choice of probability distribution over the spectrum of values we can get an expectation value for this uncertainty associated with multiplicities. We study two natural ways of averaging

this entropy and discuss the data measuring these entropy averages. This gives involves developing an interesting AdS/CFT-based information theoretic physical perspective on mathematical data of fundamental interest, namely normalized characters of specified sets of conjugacy classes in S_n.

In section 5 we define and study a number sequence $n_*(k)$: for a given k, $n_*(k)$ is the smallest n, where the normalised characters of $\{T_2, \cdots, T_k\}$ or equivalently the Casimirs $C_2, \cdots, C_k$ fail to distinguish all the Young diagrams. The mathematics literature [43, 44] contains elegant formulae for the normalised characters in terms of *content polynomials* which are explained in this section. The transformations between Casimirs, normalized characters, and content polynomials have a useful triangularity property which allows us to compute $n_*(k)$ in terms of content polynomials which are efficiently programmable in Mathematica. For $k = 6$, we find that $n_* = 80$. Our present computational approach becomes prohibitively inefficient beyond $n = 80$, so the determination of $n_*(7)$ is an interesting computational challenge. As a first step in the direction of developing analytic approaches to the determination of $n_*(k)$, for large k, we introduce a notion of *content distribution functions* which are shown to uniquely characterise Young diagrams: the content polynomials are moments of the content distribution functions. We express our earlier result about $\mathcal{G}_n$ forming a generating set in terms of the content distribution functions and observe that at $n = n_*(k)$, a set of vanishing moment equations are satisfied by the differences of content distribution functions. These content distribution functions can be visualized as segmented, connected, open strings in content space - which may be useful in the future as a tool to develop new techniques to determine the properties of $n_*(k)$.

We conclude with a summary and discussion of future research directions.

2 Casimirs, charges and matrix invariants

In this section we recall the definition of the Schur polynomial basis for the half-BPS sector [5], where half-BPS operators are labelled by Young diagrams of U(N) and are linear combinations of multi-traces of one complex matrix Z. Multi-traces with scaling dimension n, where each Z has dimension 1, are parametrized by permutations which control the contraction of U(N) indices. We review how the action of the $U(N)$ Casimirs on the multi traces can be expressed in terms of central elements of $\mathcal{Z}(\mathbb{C}(S_n))$ acting on the permutations labels [5, 27]. We explain a diagrammatic algorithm for finding the map between the Casimirs and the central elements. We then show that knowledge of the Casimirs $\{C_2, C_3, \cdots, C_k\}$ is equivalent to knowing the normalized characters for $\{T_2, T_3, \cdots T_k\}$.

2.1 The map from Casimirs to central elements for 1-matrix problem

The half-BPS operators in $N = 4$ SYM with U(N) gauge group are gauge invariant functions of one complex matrix Z transforming in the adjoint of the gauge group

$$Z \to UZU^\dagger. \tag{2.1}$$

Z is a quantum field with scaling dimension one. The gauge invariant functions are traces and products of traces. By the operator-state correspondence of CFT, these correspond

to quantum states in CFT and hence quantum states in the AdS. For scaling dimension $n \leq N$, the linearly independent gauge invariants correspond to partitions of n. The scaling dimension corresponds to the energy operator for translation along the time direction of global coordinates in AdS [28]. For example at $n = 3$, we have the following basis for gauge invariants

$$\text{tr}Z^3, \text{tr}Z^2\text{tr}Z, (\text{tr}Z)^3 \tag{2.2}$$

General multi-trace operators of degree n can be parametrized by permutations σ in S_n, the symmetric group of all permutations of $\{1, 2, \cdots, n\}$.

$$\mathcal{O}_\sigma(Z) = \sum_{i_1, \cdots, i_n} Z^{i_1}_{i_{\sigma(1)}} Z^{i_2}_{i_{\sigma(2)}} \cdots Z^{i_n}_{i_{\sigma(n)}} = \text{tr}\left(Z^{\otimes n} \sigma\right) \tag{2.3}$$

In the second expression $Z^{\otimes n}$ and σ are both being viewed as linear operators on the n-fold tensor product $V_N^{\otimes n}$ of the fundamental representation V_N of U(N).

The matrix differential operators E^i_m

$$E^i_m = \sum_{j=1}^N Z^i_j \frac{\partial}{\partial Z^m_j} \tag{2.4}$$

obey the commutation relations of the $gl(N)$ Lie algebra.

$$[E^i_j, E^k_l] = \delta^k_j E^i_l - \delta^i_l E^k_j \tag{2.5}$$

Appropriate anti-hermitian linear combinations generate the $u(N)$ Lie algebra. The Casimirs $\tilde{C}_k$ generate the centre of the enveloping algebra $U(u(N))$.

$$\tilde{C}_k = \sum_{i_1, i_2, \cdots, i_k = 1}^N E^{i_1}_{i_2} E^{i_2}_{i_3} \cdots E^{i_k}_{i_1} \tag{2.6}$$

The commutators of E^i_m with Z^p_q are

$$[E^i_m, Z^p_q] = \delta^p_m E^i_q \tag{2.7}$$

The lower q index is left invariant, while the upper index transforms as the fundamental representation V_N. The commutator of E^i_k with a product is

$$[E^i_m, Z^{p_1}_{q_1} Z^{p_2}_{q_2} \cdots Z^{p_n}_{q_n}] = \sum_{j=1}^n \delta^{p_j}_m Z^i_{q_j} \prod_{l \neq j} Z^{p_l}_{q_l} \tag{2.8}$$

The upper indices $\{p_1, p_2, \cdots, p_n\}$ transform as $V_N^{\otimes n}$.

These equations can be used to show that the Casimirs (2.6) act on the operators (2.3) through left multiplication by elements $\hat{C}_k$ in the central sub-algebra $\mathcal{Z}(\mathbb{C}(S_n))$

$$\tilde{C}_k \mathcal{O}_\sigma = \mathcal{O}_{C_k \sigma} \tag{2.9}$$

This is explained in more detail in section 3.1 and appendix A.1 of the paper [27], where the Casimirs are related to Noether charges for an enhanced $U(N) \times U(N)$ symmetry in the free field limit of $\mathcal{N} = 4$ SYM. It is shown how the quadratic Casimir of $U(N)$, expressed as a second order matrix differential operator, relates to T_2, the central element of $\mathcal{Z}(\mathbb{C}(S_n))$ which is related to permutations having one non-trivial cycle of length 2.

This map between C_k, viewed as operators on $V_N^{\otimes n}$ and central elements of the group algebra $\mathbb{C}(S_n)$ has been studied systematically in the context of 2d Yang Mills theory [45, 46]. For example, we have the results

$$
\begin{aligned}
C_2 &= \sum_{r \neq s} (rs) + Nn = 2T_2 + Nn \\
C_3 &= 3T_3 + 4NT_2 + N^2 n + n(n-1) \\
C_4 &= 4T_4 + 9NT_3 + (6N^2 + (6n - 10))T_2 + 3Nn(n-1) + N^3 n \\
C_5 &= 5T_5 + 16T_{(2,2)} + 16NT_4 + (18N^2 + 12(n-3) + 15)T_3 + (4N^3 + 24N(n-2))T_2 \\
&\quad + (nN^4 + 3N^2 n(n-1) + 2n(n-1)(n-2))
\end{aligned}
\tag{2.10}
$$

Note the appearance of $T_{(2,2)}$ in C_5. For C_6, the central elements $T_{(2,2)}$ and $T_{(3,2)}$ will appear.

An orthogonal basis in the free field inner product is parametrised by Young diagrams [5]

$$
\begin{aligned}
\mathcal{O}_R &= \frac{1}{n!} \sum_{\sigma \in S_n} \chi_R(\sigma) \mathcal{O}_\sigma(Z) \\
&= \sum_{p \vdash n} \frac{1}{\text{Sym}(p)} \prod_i (\text{tr} Z^i)^{p_i}
\end{aligned}
\tag{2.11}
$$

These are referred to as Schur Polynomial operators of the half-BPS sector. The commutation relations of the Casimirs with the gauge invariant operators can be read off from the Casimir to central elements transformations. For example

$$
\begin{aligned}
[C_2, \mathcal{O}_R] &= \left(2 \frac{\chi_R}{d_R}(T_2) + Nn \right) \mathcal{O}_R \\
[C_3, \mathcal{O}_R] &= \left(3 \frac{\chi_R(T_3)}{d_R} + 4N \frac{\chi_R}{d_R}(T_2) + N^2 n + n(n-1) \right) \mathcal{O}_R
\end{aligned}
\tag{2.12}
$$

In the above examples, we see that knowing the degree n and the normalized character of T_2 is equivalent to knowing C_2. Knowing n along with the normalized characters of T_2, T_3 is equivalent to knowing C_2, C_3. We will prove the following theorem.

Theorem 1. *The Casimir operators $C_2, \cdots, C_k$ in $V_N^{\otimes n}$ can be expressed in terms of $T_2, T_3, \cdots, T_k$. This relies on the form of the transformations between the Casimirs and the central elements.*

To prove this theorem express the Casimir C_k, acting on $V_N^{\otimes n}$ in the form

$$
C_k = \sum_{r_1, r_2, \cdots r_k = 1}^{n} \rho_{r_1} \left(E_{i_2}^{i_1} \right) \rho_{r_2} \left(E_{i_3}^{i_2} \right) \cdots \rho_{r_k} \left(E_{i_1}^{i_k} \right).
\tag{2.13}
$$

where the r_i label the different factors of $V_N^{\otimes n}$, and $\rho_{r_i}(E^{j_1}_{j_2})$ is the linear operator of $E^{j_1}_{j_2}$ acting on the r_i'th factor. There is a diagrammatic algorithm for converting the generating Casimirs to central class operators [45, 46]. We will review this algorithm and use it to prove the theorem. An immediate corollary given the above discussion is that the eigenvalues of the Casimir operators $C_2(R), \cdots, C_k(R)$ on the Young diagram operators $\mathcal{O}_R$ is determined in terms of the normalised characters $\frac{\chi_R(T_2)}{d_R}, \cdots, \frac{\chi_R(T_k)}{d_R}$.

We now describe the diagrammatic algorithm for C_k. Draw a circle with k crosses, labeled 1 to k with orientation as shown in figure 1. The sum over the r indices can be separated into different coincidences between the $\{r_1, r_2, r_3, \cdots, r_k\}$. This is a sum over set partitions: partitions of the set $\{1, 2, 3, \cdots, k\}$ into collections of subsets [47]. Thus, a particular contribution to C_k will be a partition of the set $\{1, 2, 3, \cdots, k\}$ into p subsets, where $p \leq k$. The total number of set partitions of k elements into p subsets is given by Stirling's number of the second kind [48]. All the crosses labeled by the elements in a particular subset are joined with a chord or a line. To each of the lines, apply the following procedure. Thicken the line, let the cross with the smaller label disappear, and let the cross with the largest number slide along the graph in the direction of the orientation to join the edge of the thickening - this will be illustrated in examples shortly. It is a diagrammatic translation of the multiplication of E operators acting on the same V_N: two crosses correspond to two E operators, their multiplication produces a single E and a δ function which results in a reconnection of index lines. After this operation is applied to all the chords, the effect of this is to separate the graph into a set of loops. In general a graph will separate into a loop with k'_1 crosses, a loop with k'_2 crosses and so on, which we denote by $D_{k'_1, k'_2, \cdots}$. The central element is obtained from $D_{k'_1, k'_2, \cdots}$ by retaining all the $k'_i > 1$. Relabel the $k'_i > 1$ with k_i and drop the k'_i which are equal to 0 or 1. We thus obtain $T_{(k_1, k_2, \cdots)}$. Let n_0 be the number of loops with zero crosses and let n_1 be the number of loops with one cross. There is a factor of N^{n_0} and a multiplicative factor of

$$\prod_{i=0}^{n_1-1} (n - (k_1 + k_2 + \cdots) - i).$$

(2.14)

There is one last symmetry factor generated. Consider $T_{(k_1, k_2, \cdots)}$. If $m_\mu (\mu > 1)$ is the number of k_i values equal to μ, we obtain a multiplicative factor of

$$\mathrm{Symm}(\vec{k}) = \prod_{\mu > 1} \mu^{m_\mu} m_\mu!$$

(2.15)

which accounts for the cyclic variations and permutations of all the r_i leaving the overall contribution to C_k invariant. Thus the formula for C_k takes the form

$$C_k = \sum_{\text{set partitions of } \{1, 2, \cdots, k\}} N^{n_0} \left(\prod_{i=0}^{n_1-1} (n - (k_1 + k_2 + \cdots) - i) \right) \mathrm{Symm}(\vec{k}) \, T_{k_1, k_2, \cdots}$$

(2.16)

We study some examples now. For $k = 2$, we are computing C_2. There are only two partitions of $\{1, 2\}$ to sum over. They are $\{1, 2\}$ and $\{1|2\}$. The first is a partition
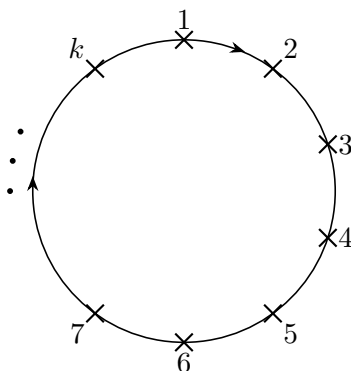
Figure 1. To compute C_k, draw a circle with k crosses labeled 1 to k. Each cross represents a E in equation (2.13).

of the set into one subset and the second is a partition of the set into two subsets. The first corresponds to the case where $r_1 = r_2$. Here crosses 1 and 2 are joined by a single chord. The second partition corresponds to the case where $r_1 \neq r_2$ and the crosses are not joined. Thicken the chord in the first graph, erase the cross labelled 1 and slide the cross labelled 2 along the graph in the direction of the orientation. The result is two separate loops - one with $k'_1 = 1$ and the other with $k'_2 = 0$. Thus, we get a $D_{1,0}$. Furthermore, $n_0 = 1$ and $n_1 = 1$ for this graph. If none of the k' are bigger than 1, which is the case here, we write the central element labeled by the identity T_1. Since $n_0 = 1$, there is a factor of N. Since $n_1 = 1$ and $k_i = 0$, we obtain the factor n. The overall result of this graph is nNT_1. This is shown in figure 2. The second graph gives a D_2, where $k'_1 = 2$. Both $n_0 = n_1 = 0$ meaning that we have a single loop with two crosses and no factors of N or n. Since $k'_1 = 2$, we relabel to $k_1 = 2$ and we finally obtain $2T_2$. The factor of 2 comes from applying formula (2.15). For $k = 5$, there are 52 set partitions in total to sum over, given by the Bell number B_5. One example is $\{1|234|5\}$. This corresponds to $r_1 \neq r_2 = r_3 = r_4 \neq r_5$. Crosses 2,3 and 4 are joined by a single chord. After thickening the chord, erasing crosses 2 and 3 and sliding 4 along the graph toward 5, the graph splits into 3 loops where $k'_1 = 3, k'_2 = 0$ and $k'_3 = 0$. Thus, we get $D_{3,0,0}$, which leads to $T_{(3)}$. Here, $n_0 = 2$ and $n_1 = 0$. This contributes a factor of N^2. We finally obtain a contribution of $3NT_3$ to C_5 from this partition. The factor of 3 is again obtained from (2.15). This example is illustrated in figure 3.

The key fact that we make use of is the following. When we sum over the different set partitions of $\{1, 2, \cdots, k\}$, the case $r_1 \neq r_2 \cdots \neq r_k$ leads to kT_k as described by the diagrammatic rules. This has branching number $k - 1$, where branching number is defined in equations (3.3) and (3.3). Any other set partition produces something of lower branching number. Suppose the set $\{r_1, \cdots, r_k\}$ is divided into p disjoint subsets. Within each subset we have r's which coincide. Setting aside the case where we get kT_k, we have $p < k$. After we multiply out the E's within a subset, we get a single E. The total number of E's left is p which is also equal to the sum of all the k'_i values. Some of the k'_i could equal 1 which corresponds to an E^i_i, which is the identity. These are not included when writing the T
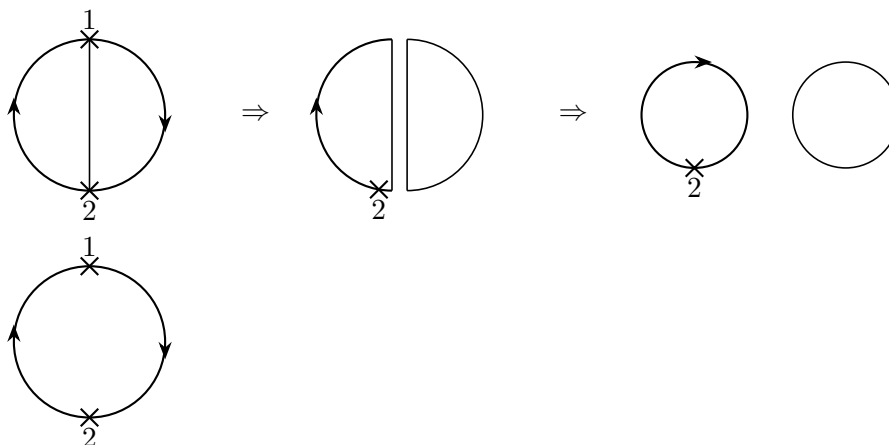
Figure 2. Computing C_2. The above graphs correspond to the two possible set partitions for $\{1,2\}$. The first graph depicts the case when r_1 and r_2 coincide. Thus, these two crosses are joined. After thickening the chord and erasing the cross with the smaller label, the graph splits into a loop with one cross labeled 2 and a loop with no crosses. The second graph depicts the case when $r_1 \neq r_2$. Here, there is no joining of crosses and the graph just remains as is.
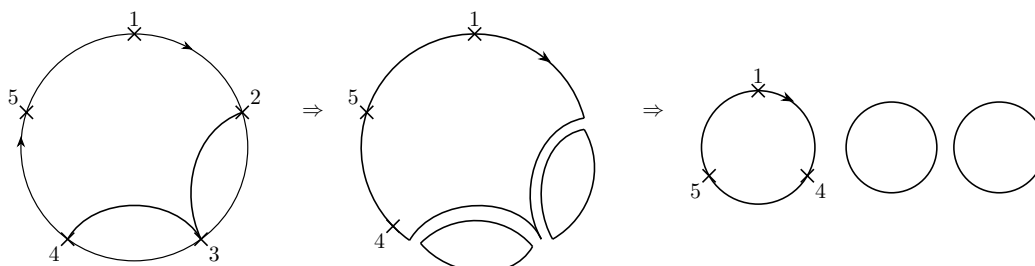
Figure 3. Computing the contribution of $r_1 \neq r_2 = r_3 = r_4 \neq r_5$ to C_5. Crosses 2,3 and 4 are joined as shown on the left. After thickening the chord joining these crosses, erasing the smaller labels 2 and 3 and sliding 4 along toward 5, this diagram splits into three disconnected pieces. Following the recipe converts this term in the Casimir sum into $3N^2 T_{(3)}$.

operators. The conjugacy class we get is generally of the form $T_{k_1,k_2,\cdots,k_l}$ where $k_1,\cdots k_l$ are positive integers larger than one. The remaining copies of E are collected into l cyclic collections. We thus have

$$p \geq k_1 + k_2 + \cdots k_l. \tag{2.17}$$

The branching number is $k_1 + k_2 + \cdots + k_l - l \leq p - l < k - 1$. But $T_2, \cdots, T_{k-1}$ generate all the T operators for branching number less than $k-1$, a result (theorem 2) we prove in section 3.2. Thus C_k can be expressed in terms of T_k along with products involving T_i for $i \leq k-1$. This means that knowing the normalized characters of $T_2, \cdots, T_k$ for any Young diagram R is equivalent to knowing the Casimir eigenvalue of $C_2, \cdots, C_k$ for the Young diagram R.

3 $\mathcal{Z}(\mathbb{C}(S_n))$: the centre of $\mathbb{C}(S_n)$

In this section, we consider the centre of the group algebra $\mathbb{C}(S_n)$, denoted as $\mathcal{Z}(\mathbb{C}(S_n))$. First, we identify a basis for $\mathcal{Z}(\mathbb{C}(S_n))$ from conjugacy classes labeled by partitions of n. Next, we show that a certain subset $\mathcal{G}_n$ of these basis elements is capable of generating $\mathcal{Z}(\mathbb{C}(S_n))$. However, at a given n not all of the elements of $\mathcal{G}_n$ are needed to generate $\mathcal{Z}(\mathbb{C}(S_n))$. Another useful basis for the centre comes from projectors associated with irreducible representations (irreps) of S_n. In the irrep basis we develop criteria for when elements of $\mathcal{G}_n$ generate $\mathcal{Z}(\mathbb{C}(S_n))$. Lastly, we present an explicit non-trivial example where a single element of $\mathcal{G}_n$ generates the centre.

3.1 A linear basis for $\mathcal{Z}(\mathbb{C}(S_n))$ from conjugacy classes

It is well-known that the partitions of n label the conjugacy classes in S_n. In particular $\lambda \vdash n$ labels the conjugacy class of permutations with cycle structure λ. Identify T_λ with the formal sum over all elements of the conjugacy class λ with equal coefficient. The elements T_λ form a basis for the centre $\mathcal{Z}(\mathbb{C}(S_n))$. Consider cycle structures of the form $[k, 1^{n-k}]$, with one cycle of length k and remaining cycles of length 1. Denote the sum of permutations, in the group algebra $\mathbb{C}(S_n)$, with this cycle structure as T_k. So, for example

$$
\begin{aligned}
T_2 &= \sum_{i<j=1}^{n} (ij) \\
T_3 &= \sum_{i<j<k} (ijk) + (ikj).
\end{aligned}
\tag{3.1}
$$

These cycle operators play an important role in this paper.

3.2 Proving that $T_2, T_3, \cdots T_n$ generate the centre of $\mathcal{Z}(\mathbb{C}(S_n))$

In this section, we prove that the central elements $\{T_2, \cdots T_n\}$ generates the centre of the group algebra $\mathbb{C}(S_n)$.

It is convenient to first define the branching number for the permutation σ corresponding to modified cycle type λ. Say $\sigma \in S_n$ has a cycle type $\rho = (\rho_1, \rho_2, \cdots, \rho_k)$, i.e. k cycles of length $\rho_1, \rho_2 \cdots \rho_k$ such that

$$
\sum_{i=1}^{k} \rho_i = n
\tag{3.2}
$$

which is denoted as $\rho \vdash n$. The modified cycle type of ρ is defined as $\lambda = (\rho_1 - 1, \rho_2 - 1, \cdots, \rho_k - 1)$ [54]. Then the branching number is defined by

$$
\begin{aligned}
B(\sigma) &= \sum_{\text{cycles}} ((\text{Cycle Length}) - 1) = n - k \\
&= \sum_{i} \lambda_i = |\lambda|
\end{aligned}
\tag{3.3}
$$

Define C_λ to be the set of all permutations ω whose modified cycle type is the partition λ. For each partition λ, let $\mathcal{C}_\lambda$ denote the sum of all $\omega \in S_n$ whose modified cycle type is λ. For example, take $\mathcal{C}_{(2)}$. We have $\mathcal{C}_{(2)} = \sum_{\omega \in S_n \cap C_{(2)}} \omega$. So, if $n = 10$,

$$\mathcal{C}_{(2)} = (1,2,3) + (1,3,2) + \cdots + (8,9,10). \tag{3.4}$$

Thus $\mathcal{C}_{(2)}$ is equal to T_3; but it is convenient here to work with a notation that uses the reduced cycle type. The set $\{T_2, \cdots, T_n\}$ has branching numbers $B = 1, \cdots, n-1$. Thus, the branching number of the T_i can be read off from the labels of the corresponding $\mathcal{C}_{i-1}$.

Theorem 2. *Given the set of central elements in $\mathcal{Z}(\mathbb{C}(S_n))$, $\mathcal{G}_k = \left\{ \mathcal{C}_{(1)}, \mathcal{C}_{(2)}, \cdots \mathcal{C}_{(k)} \right\}$, any $\mathcal{C}_\lambda$, where λ is any partition such that $|\lambda| \leq k$ can be written in terms of linear combinations of products of elements in the set.*

The statement that $\{T_2, \cdots, T_n\}$ generate $\mathcal{Z}(\mathbb{C}(S_n))$ is an immediate corollary.

We make use of the following result from [54] about the product of central elements

$$\mathcal{C}_\lambda \cdot \mathcal{C}_\mu = \sum_\nu a_{\lambda\mu}^\nu \mathcal{C}_\nu, \tag{3.5}$$

where the coefficients $a_{\lambda\mu}^\nu = 0$ unless $|\nu| \leq |\lambda| + |\mu|$.[1] In what follows, we frequently consider the case, $\mathcal{C}_\lambda \cdot \mathcal{C}_{(r)}$, $r \geq 1$ and $|\lambda| + r = |\nu|$. We have the conditions [54]

$$a_{\lambda(r)}^\nu = 0 \qquad \text{unless } \nu \geq \lambda \bigcup (r), \tag{3.6}$$

$$a_{\lambda(r)}^{\lambda \bigcup (r)} > 0. \tag{3.7}$$

In the first point, the so-called natural ordering of partitions is being used. A partition λ can be described by the sequence of its parts, listed in weakly decreasing order $(\lambda_1, \lambda_2, \cdots, \lambda_r)$:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r. \tag{3.8}$$

Given two partitions $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_r)$ and $\mu = (\mu_1, \mu_2, \cdots, \mu_s)$, the natural ordering is defined by saying that $\mu \geq \lambda$ if

$$\mu_1 + \cdots + \mu_k \geq \lambda_1 + \cdots + \lambda_k \text{ for all } k \geq 1. \tag{3.9}$$

In this definition partitions are extended by zero parts if necessary. This is also called dominance order. Taking the transpose of partitions reverses the dominance order. In other words, if $\mu \geq \lambda$ then $\lambda^T \geq \mu^T$. We now present examples for small values of k:

---

[1]Note that the interpretation of permutations in terms of branched covers which plays an important role in the string theory of 2D Yang Mills [33] allows a physical interpretation of this inequality. μ, λ describe the branching over two branch points. If we let the two branch points collide to have a single branching described by ν, the change in branching number due to the collision $|\lambda| + |\mu| - |\nu|$ must be non-negative since this is accounted by the formation of a number, positive or zero, of collapsed handle singularities as a result of the collision. In the process of collision the Euler characteristic of the covering surface does not change, but contributions to the Riemann Hurwitz formula from branch points are traded for contributions form collapsed handles.

- When $k = 1$, we have $\mathcal{G}_k = \{\mathcal{C}_{(1)}\}$. All possible $\mathcal{C}$'s are already in our generating set.

- For $k = 2$, we have $\mathcal{G}_k = \{\mathcal{C}_{(1)}, \mathcal{C}_{(2)}\}$. The partitions ν with $|\nu| = 2$ are (2) and $(1,1)$. $\mathcal{C}_{(2)}$ is already in our generating set. Let's check that $\mathcal{C}_{(1,1)}$ is generated by $\mathcal{G}_2$.

$$\mathcal{C}_{(1,1)} = \sum_{\omega \in S_{10} \bigcap C_{(1,1)}} \omega = (1,2)\,(3,4) + (1,3)\,(2,4) + \cdots + (7,8)\,(9,10). \quad (3.10)$$

Using equation (3.5), we have

$$\mathcal{C}_{(1)} \cdot \mathcal{C}_{(1)} = \sum_{\nu} a^{\nu}_{(1)(1)} \mathcal{C}_{\nu}, \qquad |\nu| \le 2. \quad (3.11)$$

The terms for which $|\nu| < 2$ are already generated by $\mathcal{G}_1$ which is contained in $\mathcal{G}_2$. When $|\nu| = 2$, we need to sum over the ν for which $\nu \ge \lambda \bigcup (r)$. Here $\lambda = (1)$ and $(r) = (1)$. So $\lambda \bigcup (r) = (1,1)$. We have:

$$(2) \ge (1,1) \quad (3.12)$$

$$(1,1) = (1) \bigcup (1) \quad (3.13)$$

where the possible ν's are on the left hand side. Thus

$$\mathcal{C}_{(1)} \cdot \mathcal{C}_{(1)} = a\,\mathcal{C}_{(2)} + b\,\mathcal{C}_{(1,1)} + \sum_{\nu\,(|\nu| \le 1)} x_{\nu}\,\mathcal{C}_{\nu}. \quad (3.14)$$

$$(3.15)$$

Checking this explicitly for $n = 10$

$$\mathcal{C}_{(1)} \cdot \mathcal{C}_{(1)} = 3\mathcal{C}_{(2)} + 2\,\mathcal{C}_{(1,1)} + 45\mathcal{C}_{(0)}. \quad (3.16)$$

The coefficient of $\mathcal{C}_{(1,1)}$ is always non-zero from equation (3.7). This can also be seen from the diagrammatic algorithm described earlier. The algorithm for converting C_k into $T_{k_1,k_2\cdots}$ can also be used to multiply the T's. The term $\mathcal{C}_{(1,1)}$ results, in the diagrammatic algorithm, from the diagram with zero lines joining crosses from one $\mathcal{C}_{(1,1)}$ to the other.

- $k = 3$, $\mathcal{G}_k = \{\mathcal{C}_{(1)}, \mathcal{C}_{(2)}, \mathcal{C}_{(3)}\}$. The partitions with $|\nu| = 3$ are $\{(3), (2,1), (1,1,1)\}$. We need to check that $\mathcal{C}$ labeled by each of these partitions is generated by $\mathcal{G}_3$. All terms for which $|\nu| \le 2$ are generated by $\mathcal{G}_2$ which is contained in $\mathcal{G}_3$. Now, $\mathcal{C}_{(3)} \in \mathcal{G}_3$ already. According to natural ordering

$$(3) \ge (2,1) \ge (1,1,1). \quad (3.17)$$

Thus, the next largest from (3) is $(2,1)$. From equations (3.5)–(3.7), we consider $\mathcal{C}_{(2)} \cdot \mathcal{C}_{(1)}$. Both $\mathcal{C}_{(1)}$ and $\mathcal{C}_{(2)}$ are contained in $\mathcal{G}_2$. We see that $\mathcal{C}_{(2)} \cdot \mathcal{C}_{(1)}$ will contain $\mathcal{C}_{(3)}$ and $\mathcal{C}_{(2,1)}$. The only other terms will be $\mathcal{C}_{\nu}$ such that $|\nu| \le 2$, which are generated by $\mathcal{G}_2$. The next largest partition is $(1,1,1)$. Thus, we consider $\mathcal{C}_{(1,1)} \cdot \mathcal{C}_{(1)}$. Again, both $\mathcal{C}_{(1)}$ and $\mathcal{C}_{(1,1)}$ are generated by $\mathcal{G}_2$. From the ordering in (3.17) and from equations (3.5)–(3.7), $\mathcal{C}_{(1,1)} \cdot \mathcal{C}_{(1)}$ will contain $\mathcal{C}_{(1,1,1)}$ and may contain $\mathcal{C}_{(3)}$, $\mathcal{C}_{(2,1)}$ along with $\mathcal{C}_{\nu}$ for which $|\nu| \le 2$. For $n = 10$,

$$\mathcal{C}_{(1,1)} \cdot \mathcal{C}_{(1)} = 2\,\mathcal{C}_{(3)} + 3\,\mathcal{C}_{(2,1)} + 3\,\mathcal{C}_{(1,1,1)} + 28\,\mathcal{C}_{(1)}. \quad (3.18)$$

- For $k = 4$, we have $\mathcal{G}_k = \{\mathcal{C}_{(1)}, \mathcal{C}_{(2)}, \mathcal{C}_{(3)}, \mathcal{C}_{(4)}\}$. The natural ordering for partitions of 4 is a total ordering:

$$(4) \geq (3,1) \geq (2,2) \geq (2,1,1) \geq (1,1,1,1). \tag{3.19}$$

Again, we need to check that $\mathcal{C}$ labeled by each of these partitions is generated by $\mathcal{G}_4$. All terms for which $|\nu| \leq 3$ are generated by $\mathcal{G}_3$ which is contained in $\mathcal{G}_4$. Now, $\mathcal{C}_{(4)}$ is already contained in our generating set. We proceed to the next largest partition $(3,1)$. We get $\mathcal{C}_{(3,1)}$ from $\mathcal{C}_{(3)} \cdot \mathcal{C}_{(1)}$. Both $\mathcal{C}_{(3)}$ and $\mathcal{C}_{(1)}$ are contained in $\mathcal{G}_3$. $\mathcal{C}_{(3)} \cdot \mathcal{C}_{(1)}$ will contain $\mathcal{C}_{(3,1)}$ and may contain $\mathcal{C}_{(4)}$ as well as $\mathcal{C}_\nu$ for which $|\nu| \leq 3$. For $n = 10$,

$$\mathcal{C}_{(3)} \cdot \mathcal{C}_{(1)} = \mathcal{C}_{(3,1)} + 5\,\mathcal{C}_{(4)} + 4\,\mathcal{C}_{(1,1)} + 21\,\mathcal{C}_{(2)}. \tag{3.20}$$

We continue in this way down the order in (3.19) generating the $\mathcal{C}$ operators labeled by each these partitions using $\mathcal{C}$ operators generated by $\mathcal{G}_3$. For $n = 10$,

$$\mathcal{C}_{(2)} \cdot \mathcal{C}_{(2)} = 240\,\mathcal{C}_{(0)} + 22\,\mathcal{C}_{(2)} + 8\,\mathcal{C}_{(1,1)} + 5\,\mathcal{C}_{(4)} + 2\,\mathcal{C}_{(2,2)}. \tag{3.21}$$

$$\mathcal{C}_{(2,1)} \cdot \mathcal{C}_{(1)} = 21\,\mathcal{C}_{(2)} + 24\,\mathcal{C}_{(1,1)} + 5\,\mathcal{C}_{(4)} + 4\,\mathcal{C}_{(3,1)} + 6\,\mathcal{C}_{(2,2)} + 2\,\mathcal{C}_{(2,1,1)}. \tag{3.22}$$

$$\mathcal{C}_{(1,1,1)} \cdot \mathcal{C}_{(1)} = 15\,\mathcal{C}_{(1,1)} + 2\,\mathcal{C}_{(3,1)} + 3\,\mathcal{C}_{(2,1,1)} + 4\,\mathcal{C}_{(1,1,1,1)}. \tag{3.23}$$

- For $k = 5$, natural ordering is still a total ordering,

$$(5) \geq (4,1) \geq (3,2) \geq (3,1,1) \geq (2,2,1) \geq (2,1,1,1) \geq (1,1,1,1,1). \tag{3.24}$$

For $n = 10$, we have

$$\mathcal{C}_{(4)} \cdot \mathcal{C}_{(1)} = 6\,\mathcal{C}_{(5)} + \mathcal{C}_{(4,1)} + 24\,\mathcal{C}_{(3)} + 6\,\mathcal{C}_{(2,1)} \tag{3.25}$$

$$\mathcal{C}_{(3)} \cdot \mathcal{C}_{(2)} = 6\,\mathcal{C}_{(5)} + \mathcal{C}_{(3,2)} + 28\,\mathcal{C}_{(3)} + 12\,\mathcal{C}_{(2,1)} + 112\,\mathcal{C}_{(1)} \tag{3.26}$$

$$\mathcal{C}_{(3,1)} \cdot \mathcal{C}_{(1)} = 6\,\mathcal{C}_{(5)} + 5\,\mathcal{C}_{(4,1)} + 3\,\mathcal{C}_{(3,2)} + 2\,\mathcal{C}_{(3,1,1)} + 15\,\mathcal{C}_{(3)} + 15\,\mathcal{C}_{(2,1)} \tag{3.27}$$

$$+ 12\,\mathcal{C}_{(1,1,1)} \tag{3.28}$$

$$\mathcal{C}_{(2,2)} \cdot \mathcal{C}_{(1)} = 3\,\mathcal{C}_{(5)} + 4\,\mathcal{C}_{(3,2)} + \mathcal{C}_{(2,2,1)} + 10\,\mathcal{C}_{(2,1)} \tag{3.29}$$

$$\mathcal{C}_{(2,1,1)} \cdot \mathcal{C}_{(1)} = 5\,\mathcal{C}_{(4,1)} + 2\,\mathcal{C}_{(3,2)} + 4\,\mathcal{C}_{(3,1,1)} + 6\,\mathcal{C}_{(2,2,1)} + 4\,\mathcal{C}_{(2,1,1,1)} \tag{3.30}$$

$$+ 10\,\mathcal{C}_{(2,1)} + 24\,\mathcal{C}_{(1,1,1)} \tag{3.31}$$

$$\mathcal{C}_{(1,1,1,1)} \cdot \mathcal{C}_{(1)} = 2\,\mathcal{C}_{(3,1,1)} + 3\,\mathcal{C}_{(2,1,1,1)} + 5\,\mathcal{C}_{(1,1,1,1,1)} + 6\,\mathcal{C}_{(1,1,1)} \tag{3.32}$$

- For $k = 6$, natural ordering is no longer a total ordering. We have

$$(6) \geq (5,1) \geq (4,2) \geq \{(4,1,1),(3,3)\} \geq (3,2,1)$$
$$\geq \{(3,1^3),(2^3)\} \geq (2,2,1,1) \geq (2,1^4) \geq (1^6) \tag{3.33}$$

The partitions $(4,1,1)$ and $(3,3)$ are incomparable according to natural ordering. Thus, $(3,3)$ will not appear in the product $\mathcal{C}_{(4,1)} \cdot \mathcal{C}_{(1)}$, and $(4,1,1)$ will not appear

in the product $\mathcal{C}_{(3)} \cdot \mathcal{C}_{(3)}$. Similarly for the incomparable $(3, 1^3)$ and (2^3), which are conjugates of $(4, 1, 1)$ and $(3, 3)$. Calculating these products for $n = 10$, we find

$$\begin{aligned}
\mathcal{C}_{(4,1)} \cdot \mathcal{C}_{(1)} &= 7\,\mathcal{C}_{(6)} + 6\,\mathcal{C}_{(5,1)} + 3\,\mathcal{C}_{(4,2)} + 2\,\mathcal{C}_{(4,1,1)} + 10\,\mathcal{C}_{(4)} \\
&\quad + 15\,\mathcal{C}_{(3,1)} + 12\,\mathcal{C}_{(2,1,1)}.
\end{aligned} \tag{3.34}$$

$$\begin{aligned}
\mathcal{C}_{(3)} \cdot \mathcal{C}_{(3)} &= 7\,\mathcal{C}_{(6)} + 2\,\mathcal{C}_{(3,3)} + 40\,\mathcal{C}_{(4)} + 16\,\mathcal{C}_{(3,1)} + 27\,\mathcal{C}_{(2,2)} + 147\,\mathcal{C}_{(2)} \\
&\quad + 98\,\mathcal{C}_{(1,1)} + 1260\,\mathcal{C}_{(0)}
\end{aligned} \tag{3.35}$$

$$\begin{aligned}
\mathcal{C}_{(3,1,1)} \cdot \mathcal{C}_{(1)} &= 6\,\mathcal{C}_{(5,1)} + 5\,\mathcal{C}_{(4,1,1)} + 4\,\mathcal{C}_{(3,3)} + 3\,\mathcal{C}_{(3,2,1)} + 3\,\mathcal{C}_{(3,1,1,1)} + 6\,\mathcal{C}_{(3,1)} \\
&\quad + 9\,\mathcal{C}_{(2,1,1)} + 24\,\mathcal{C}_{(1,1,1,1)}
\end{aligned} \tag{3.36}$$

$$\begin{aligned}
\mathcal{C}_{(2,2)} \cdot \mathcal{C}_{(2)} &= 7\,\mathcal{C}_{(6)} + 5\,\mathcal{C}_{(4,2)} + 3\,\mathcal{C}_{(2,2,2)} + 25\,\mathcal{C}_{(4)} + 8\,\mathcal{C}_{(3,1)} + 26\,\mathcal{C}_{(2,2)} \\
&\quad + 8\,\mathcal{C}_{(2,1,1)} + 70\,\mathcal{C}_{(2)}.
\end{aligned} \tag{3.37}$$

- Assume that all $\mathcal{C}$ labeled by partitions λ for which $|\lambda| \leq k - 1$ can be generated by $\mathcal{G}_{k-1} = \{\mathcal{C}_{(1)}, \mathcal{C}_{(2)}, \cdots, \mathcal{C}_{(k-1)}\}$

- Now consider the generating set $\mathcal{G}_k$ and all partitions λ such that $|\lambda| = k$. We show that all $\mathcal{C}$ labeled by partitions of k can be generated by $\mathcal{G}_k$. According to natural ordering

$$(k) \geq (k - 1, 1) \geq (k - 2, 2) \geq (k - 2, 1, 1), \cdots, (2, 1^{k-2}) \geq (1^k). \tag{3.38}$$

We consider $\mathcal{C}_\lambda \cdot \mathcal{C}_{(r)}$ where both $\mathcal{C}_\lambda$ and $\mathcal{C}_{(r)}$ have $|\lambda| \leq k - 1$ and $r \leq k - 1$ and are thus generated by $\mathcal{G}_{k-1}$. According to equations (3.5)–(3.7), we sum over partitions ν such that $\nu \geq \lambda \bigcup (r)$ and $a_{\lambda(r)}^{\lambda \bigcup (r)} > 0$. Now, $\mathcal{C}_{(k)}$ is already contained in $\mathcal{G}_k$. To generate $\mathcal{C}_{(k-1,1)}$, we multiply $\mathcal{C}_{(k-1)} \cdot \mathcal{C}_{(1)}$. The term $\mathcal{C}_{(k-1,1)}$ will appear with non-zero coefficient and all partitions larger than $(k - 1, 1)$ may also appear. This only includes $\mathcal{C}_{(k)}$, which is already contained in $\mathcal{G}_k$. Next, to generate $\mathcal{C}_{(k-2,2)}$, we multiply $\mathcal{C}_{(k-2)} \cdot \mathcal{C}_{(2)}$. The term $\mathcal{C}_{(k-2,2)}$ will appear and partitions larger than $(k - 2, 2)$, which are (k) and $(k - 1, 1)$, may also appear. However, $\mathcal{C}_{(k)}$ is already contained in $\mathcal{G}_k$, and $\mathcal{C}_{(k-1,1)}$ is generated by $\mathcal{G}_k$ as we have seen above. To generate $\mathcal{C}_{(k-2,1,1)}$, we multiply $\mathcal{C}_{(k-2,1)} \cdot \mathcal{C}_{(1)}$. The term $\mathcal{C}_{(k-2,1,1)}$ is sure to appear, while the partitions larger than $(k - 2, 1, 1)$, i.e., $(k), (k - 1, 1)$ and $(k - 2, 2)$ may also appear. However, each one of these have already been shown to be generated by $\mathcal{G}_k$. We may continue in this way proceeding one by one down the chain of partitions in (3.38). When we arrive at the smallest partition (1^k), we compute $\mathcal{C}_{(1^{k-1})} \cdot \mathcal{C}_1$. $\mathcal{C}_{(1^k)}$ will be generated as well as all partitions larger may also appear. But each of these have, in turn, been shown to be generated by $\mathcal{G}_k$ in the same way as described above.

A small comment is in order. In proceeding down the list of partitions, we will arrive at a set of partitions that are mutually, or pairwise, incomparable according to natural ordering. To generate any one of these partitions, we still form a product of the form (3.5). Since we need to sum strictly over partitions that are larger than the one in question all incomparable partitions will not appear in the result. As an example, see $k = 6$.

We have successfully shown that $\mathcal{G}_k$ can generate any $\mathcal{C}$ labeled by partition λ whose $|\lambda| \leq k$. In terms of the T operators, this result means that $\{T_2, \cdots, T_{k+1}\}$ is capable of generating any T labeled by a partition λ with branching number k or less. This means that $\{T_2, \cdots, T_n\}$ generates any central element T_λ with branching number $n - 1$ or less. However, the $T_\lambda \in \mathcal{Z}(\mathbb{C}(S_n))$ with the largest branching number is T_n with $B = n - 1$, which is already contained in our generating set. Thus, $\{T_2, \cdots, T_n\}$ can generate the centre of $\mathbb{C}(S_n)$.

Remark 1. *Lastly, it is worth noting that T_n can be expressed in terms of $\{T_2, \cdots, T_{n-1}\}$. Using (3.5), the calculation of $\mathcal{C}_{n-2} \cdot \mathcal{C}_1$ yields $\mathcal{C}_{n-1}$ (which is T_n), $\mathcal{C}_{(n-2,1)}$, (which is $T_{(n-1,2)}$) and then T_λ with a lower branching number than $n - 1$, which can all be generated $\{T_2, \cdots, T_{n-1}\}$. However, we exclude $T_{(n-1,2)}$ since it is labeled by a permutation of $n + 1$. Thus, T_n can be expressed in terms the set $T_2, \cdots, T_{n-1}$.*

3.3 Generating sets for $\mathcal{Z}(\mathbb{C}(S_n))$ from irreducible representations

A basis for the centre $\mathcal{Z}(\mathbb{C}(S_n)$ is given by projectors (orthogonal idempotents)

$$P_R = \frac{d_R}{n!} \sum_{\sigma \in S_n} \chi_R(\sigma)\ \sigma \tag{3.39}$$

where $\chi_R(\sigma)$ is the character in the irreduciblel representation R of the permutation σ. The R correspond to Young diagrams. This is general fact about group algebras of finite groups, see e.g. [34]. These obey

$$P_R P_S = \delta_{RS} P_R$$
$$\sum_R P_R = 1 \tag{3.40}$$

where 1 is the identity permutation and the identity in the associative algebra $\mathbb{C}(S_n)$. The number of these projectors is $p(n)$, the number of partitions of n. Taking the trace in an irrep R, and using orthogonality of characters, we have

$$\chi_S(P_R) = \delta_{RS} d_R \tag{3.41}$$

We make use of the following fact (lemma 2.1 in [35]):

Lemma 1. ?? *A linear combination $T = \sum_R a_R P_R$ with all a_R distinct generate the centre. To see this, we observe that*

$$P_S = \prod_{S_1 \neq S} \frac{(T - a_{S_1})}{(a_S - a_{S_1}))} \tag{3.42}$$

which follows by expanding the right hand side and using the projector equations. This is essentially a fact about the algebra of diagonal matrices: the algebra of diagonal matrices is generated by any diagonal matrix with distinct entries.

3.4 Generating sets of cycle structures from lists of normalised characters

We can expand any central element such as T_2 in terms of these projectors

$$T_2 = \sum_R (a_2)^R P_R \tag{3.43}$$

Take a trace in the representation S on both sides

$$\chi_S(T_2) = (a_2)^S d_S \tag{3.44}$$

So we have

$$T_2 = \sum_R \frac{\chi_R(T_2)}{d_R} P_R \tag{3.45}$$

From lemma **??**, this means that if the list of $\frac{\chi_R(T_2)}{d_R}$ has no repetitions, then T_2 generates the centre.

If we have a list of irreps R, where the normalized characters of T_2 are equal, then we can use another central element such as T_3. Within this block, we apply the lemma **??** again. So to find out whether T_2, T_3 generate the centre, we just need to look at the matrix $2 \times p(n)$ matrix

$$\left(\frac{\chi_R(T_2)}{d_R}, \ \frac{\chi_R(T_3)}{d_R} \right) \tag{3.46}$$

If no two rows are identical, then T_2, T_3 generate the centre. We may apply lemma **??** iteratively. More generally, we would like to consider the $k \times p(n)$ matrix. The central elements $\{T_2, T_3, \cdots, T_k\}$ may each be expanded in terms of the projectors P_R, with their respective normalized characters being the expansion coefficients. If this list of normalized characters is distinct for each R then no two rows in

$$\left(\frac{\chi_R(T_2)}{d_R}, \ \frac{\chi_R(T_3)}{d_R} \cdots, \ \frac{\chi_R(T_k)}{d_R} \right) \tag{3.47}$$

are identical. According to lemma **??**, the $\{T_2, T_3, \cdots, T_k\}$ will generate the centre. For example if for R_1 and R_2, the list of normalized characters for $\{T_2, T_3, \cdots, T_{k-1}\}$ are identical then this list of T_i and their respective powers are no longer linearly independent. They no longer generate the subspace of the centre spanned by P_{R_1} and P_{R_2}. We now include one more central element T_k. If $\frac{\chi_R(T_k)}{d_R}$ is distinct for R_1 and R_2 then the set $\{T_2, T_3, \cdots, T_k\}$ are linearly independent and, according to lemma **??**, this list generates $\mathcal{Z}(\mathbb{C}(S_n))$. It is interesting to study the sequence of the smallest n values where the $\{T_2, T_3, \cdots, T_k\}$ fail to generate the centre of $\mathbb{C}(S_n)$. Denote this sequence by $n_*(k)$. The problem of finding $n_*(k)$ is a matter of understanding the degeneracies in the characters of S_n, using nice formulae for these characters available from the mathematics literature [43, 44]. Using the discussion in section 2 $n_*(k)$ is also the smallest value of n where knowledge of the all the Casimir eigenvalues $C_2, C_3, \cdots, C_k$ does not suffice to distinguish all the Young diagrams.

3.5 Generating the centre for $n = 5$ with T_2

The list of normalized characters for T_2 is $\{10, 5, 2, 0, -2, -5, -10\}$. In this list, there are no degeneracies. Hence by the above argument, we expect T_2 to generate the centre. Below, we show this explicitly.

We generate the following equations in Mathematica by taking successive powers of T_2.

$$
\begin{aligned}
(T_2)^2 &= 10 + 3T_3 + 2T_{(2,2)} \\
(T_2)^3 &= 34T_2 + 16T_4 + 9T_{(3,2)} \\
(T_2)^4 &= 340 + 207T_3 + 125T_5 + 168T_{(2,2)} \\
(T_2)^5 &= 2086T_2 + 1664T_4 + 1461T_{(3,2)} \\
(T_2)^6 &= 20860 + 17703T_3 + 15625T_5 + 16672T_{(2,2)} \\
(T_2)^7 &= 177094T_2 + 166656T_4 + 161469T_{(3,2)} \\
(T_2)^8 &= 1770940 + 1692687T_3 + 1640625T_5 + 1666688T_{(2,2)}.
\end{aligned}
\tag{3.48}
$$

For example, after computing $(T_2)^2$, we count that the identity element T_1 appears ten times, T_3 appears 3 times and $T_{(2,2)}$, the formal sum of permutations having disjoint two-cycles, appearing twice. We may now invert this system of equations to solve for the other T quantities. Solving for T_4 and $T_{(3,2)}$ in terms of T_2, we find

$$
T_4 = -\frac{2603}{2800}(T_2)^3 + \frac{12987}{280000}(T_2)^5 - \frac{103}{280000}(T_2)^7 \tag{3.49}
$$

$$
T_{(3,2)} = \frac{331}{525}(T_2)^3 - \frac{3541}{105000}(T_2)^5 + \frac{29}{105000}(T_2)^7 \tag{3.50}
$$

Next, we solve for $T_1 = 1, T_3, T_5$ and $T_{(2,2)}$ in terms of T_2

$$
1 = \frac{62273}{208320}(T_2)^2 - \frac{45819}{3472000}(T_2)^4 + \frac{6863}{27776000}(T_2)^6 - \frac{1}{666624}(T_2)^8 \tag{3.51}
$$

$$
T_3 = -\frac{541}{420}(T_2)^2 + \frac{25}{336}(T_2)^4 - \frac{1}{1680}(T_2)^6 \tag{3.52}
$$

$$
T_5 = \frac{625}{10416}(T_2)^2 - \frac{125}{6944}(T_2)^4 + \frac{43}{55552}(T_2)^6 - \frac{1}{166656}(T_2)^8, \tag{3.53}
$$

$$
T_{(2,2)} = \frac{195299}{208320}(T_2)^2 - \frac{31681}{694400}(T_2)^4 - \frac{1903}{5555200}(T_2)^6 + \frac{5}{666624}(T_2)^8. \tag{3.54}
$$

This shows that for $n = 5$ each T_μ corresponding to a cycle-type μ, may be written in terms of T_2. Thus, T_2 generates the centre $\mathcal{Z}(\mathbb{C}(S_5))$.

4 Dimensions and entropies associated with low order cycle structures

From the previous section, we concluded that not all elements in $\mathcal{G}_n$ are needed to generate $\mathcal{Z}(\mathbb{C}(S_n))$. Indeed for some n, all we need is T_2. It is interesting to study this operator's ability to generate the centre as a function of n. In this section we present some data concerning T_2 as well as the combination of T_2 and T_3. We first discuss the codimension for the subspaces generated by these two central elements, and then we discuss the entropy related to the degeneracies of their normalized characters.

4.1 Co-dimensions as measures of uncertainty

We have proved that T_2 generates the centre of S_n if there are no repetitions or degeneracies in the list of its normalized character. In this section, we refer to the normalized characters frequently and thus we define

$$\frac{\chi_R(T_k)}{d_R} \equiv \widehat{\chi}_k^R. \tag{4.1}$$

For a given n, if there are degeneracies in $\widehat{\chi}_2^R$ then we include the normalized character of T_3. If the list $\{\widehat{\chi}_2^R, \widehat{\chi}_3^R\}$ has no repetitions then T_2 and T_3 generate the centre for that particular n.

For $n = 1, 2, 3, 4, 5$, there are no repetitions in $\widehat{\chi}_2^R$, so T_2 does indeed generate $\mathcal{Z}(\mathbb{C}(S_n))$. For $n = 6$, we encounter our first degeneracy. The list of normalized characters for T_2 is

$$\{15, 9, 5, 3, 3, 0, -3, -3, -5, -9, -15\} \tag{4.2}$$

There are two repetitions in this list. Thus, T_2 no longer generates $\mathcal{Z}(\mathbb{C}(S_n))$. T_2 instead only generates a subspace of $\mathcal{Z}(\mathbb{C}(S_n))$. The codimension of a subspace is the difference between the dimension of the full space and the dimension of the subspace. We define $\text{codim}_{T_k}(n)$ to be the difference of the dimension of $\mathcal{Z}(\mathbb{C}(S_n))$ and the dimension of the subspace generated by T_k. The codimension of a subspace generated by a T_k or a collection of T_k's is a measure of how close the central element or collection of central elements is, to generating the centre. For T_2 at $n = 6$, the number of distinct $\widehat{\chi}_2^R$ is 9, which is 2 less than $p(6) = 11$, so the codimension of the subspace generated is equal to two. Interestingly, T_2 once again generates $\mathcal{Z}(\mathbb{C}(S_n))$ for $n = 7$, producing a codimension of zero. From $n = 8$ however, T_2 fails to generate $\mathcal{Z}(\mathbb{C}(S_n))$. The codimension data for T_2 for $n = 2$ to $n = 70$ is shown in table 1. These co-dimensions can be viewed as a measure of the uncertainty in the determination of Young diagrams with n boxes when we only know the normalized character of T_2: equivalently when we only know the second Casimir. From this data, we can calculate the relative dimension for T_2

$$R_{T_2} = \frac{p(n) - \text{codim}_{T_2}(n)}{p(n)}, \tag{4.3}$$

where $p(n)$ is the dimension of $\mathcal{Z}(\mathbb{C}(S_n))$. The plot for this is shown in figure 4.

We now discuss the codimension of the subspace generated by both T_2 and T_3. For $n = 6$, the list of normalized characters of T_2 and T_3 is

$$\Big\{\{15, 55\}, \{9, 31\}, \{5, 15\}, \{3, 19\}, \{3, 7\}, \{0, 10\},$$
$$\{-3, 19\}, \{-3, 7\}, \{-5, 15\}, \{-9, 31\}, \{-15, 55\}\Big\}. \tag{4.4}$$

Since there are no repetitions in this list (i.e. each pair of numbers is unique in this list), these two central elements generate $\mathcal{Z}(\mathbb{C}(S_n))$. Thus, the codimension for this subspace is zero. We find a zero codimension for all cases of n up until $n = 15$. For $n = 15$ the subspace generated by $\{T_2, T_3\}$ has a codimension of 3. See table 1 for data for $n = 2$ to

n	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
T_2 codim	0	0	0	0	2	0	3	5	11	9	32	26	56	89	122	156
T_3 codim	1	1	2	3	5	7	12	17	24	33	49	64	90	120	164	214
$\{T_2,T_3\}$ codim	0	0	0	0	0	0	0	0	0	0	0	0	0	3	4	4
n	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
T_2 codim	244	305	434	571	755	964	1280	1613	2059	2599	3277	4064	5097	6267	7742	9488
T_3 codim	285	367	485	619	801	1013	1298	1637	2052	2578	3214	3978	4945	6110	7492	9181
$\{T_2,T_3\}$ codim	8	1	17	20	29	36	89	98	111	171	253	292	460	556	856	1043
n	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49
T_2 codim	11607	14138	17192	20776	25108	30214	36329	43486	52053	62044	73908	87807	104155	123263	145750	171880
T_3 codim	11224	13668	16606	20149	24278	29315	35258	42262	50636	60452	72049	85751	101917	120694	142862	168775
$\{T_2,T_3\}$ codim	1343	1797	2467	3088	3891	4964	6638	8295	10537	13254	17110	20958	26528	32990	41070	51358
n	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65
T_2 codim	202537	238158	279738	327980	384160	449171	524666	611871	712853	829335	963938	1118836	1297411	1502656	1738693	2009491
T_3 codim	199031	234300	275590	323265	379108	443666	518646	605320	705914	821576	955785	1109954	1287924	1492280	1727867	1997539
$\{T_2,T_3\}$ codim	62802	77099	95467	117359	141406	173538	211477	254289	309073	374251	444426	538166	643109	766460	914920	1088525
n	66	67	68	69	70											
T_2 codim	2320397	2676422	3084370	3550860	4084383											
T_3 codim	2307932	2662970	3070080	3535437	4068446											
$\{T_2,T_3\}$ codim	1286732	1530746	1807711	2130136	2507976											

Table 1. Table showing the codimension data for T_2, for T_3, and then for $\{T_2,T_3\}$ for $n = 2$ to $n = 70$.
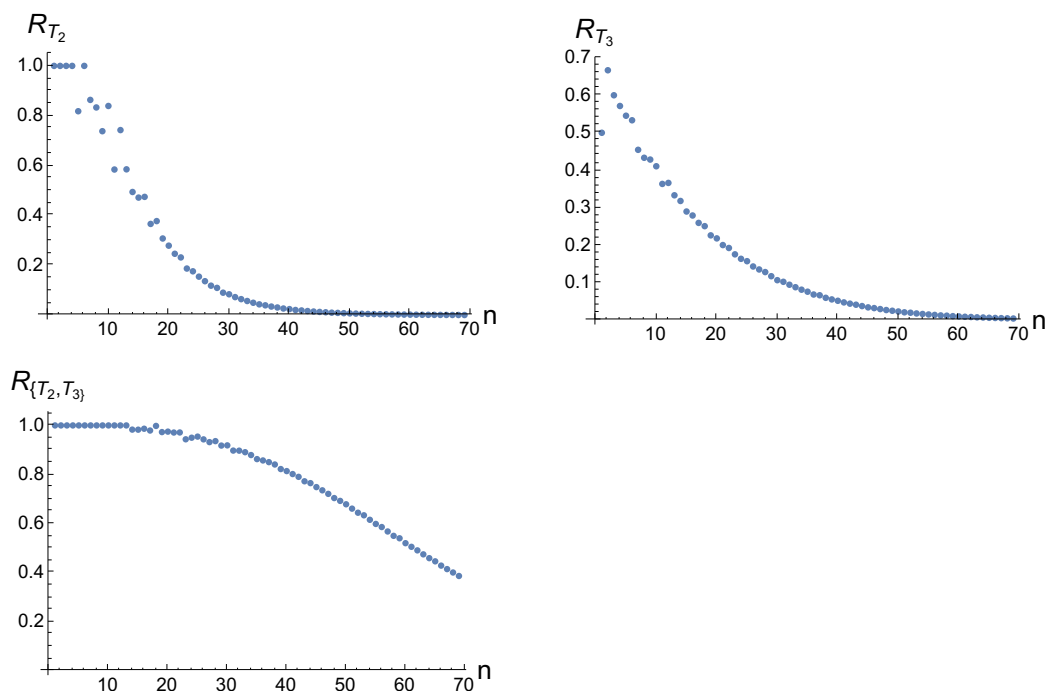
Figure 4. Figures showing relative dimension plots for T_2, T_3 and for $\{T_2,T_3\}$.

$n = 70$. The co-dimensions for $\{T_2,T_3\}$ can again be viewed as a measure of the uncertainty in the determination of Young diagrams with n boxes when we only know the normalized characters of T_2 and T_3: equivalently when we only know the second and third Casimir. Figure 4 shows the relative dimension for $\{T_2,T_3\}$. We also plot the relative dimension for $\{T_2,T_3\}$ in figure 4. Note that the codimension of the space generated by $\{T_2, \cdots, T_n\}$ is zero for all n. Furthermore, we note that the relative dimension plot for $\{T_2, \cdots, T_n\}$ would just be a straight line at $R_{\{T_2,\cdots,T_n\}} = 1$.

It is natural to expect that the uncertainty of identifying the Young diagram is smaller when we know both $\widehat{\chi}_2^R$ and $\widehat{\chi}_3^R$ compared to when we only know one of these quantities. Furthermore, the higher the codimension of T_k (or of some subset of T's), the higher this uncertainty becomes. Thus, we expect

$$
\begin{aligned}
\mathrm{codim}_{\{T_2,T_3\}} &< \mathrm{codim}_{T_2} \\
\mathrm{codim}_{\{T_2,T_3\}} &< \mathrm{codim}_{T_3}.
\end{aligned}
\tag{4.5}
$$

This is indeed reflected in the data of table 1 for all values of n listed there. The relative sizes of codim_{T_2} and codim_{T_3} may also be studied. From the data, we see that

$$
\begin{aligned}
\mathrm{codim}_{T_2} &< \mathrm{codim}_{T_3} \text{ for } n \leq 25 \\
\mathrm{codim}_{T_3} &< \mathrm{codim}_{T_2} \text{ for } 26 \leq n \leq 70.
\end{aligned}
\tag{4.6}
$$

The behaviour of these codimensions for n larger than 70 is an interesting problem. It is natural to conjecture that $\mathrm{codim}_{T_3} < \mathrm{codim}_{T_2}$ persists for all n higher than 26.

Note that AdS/CFT motivates the study of finite N versions of this codimension problem. There is a finite N truncation of $\mathcal{Z}(\mathbb{C}(S_n))$, where we set to zero all the projectors P_R with height $l(R)$ constrained to satisfy $l(R) > N$. This subspace, which we denote $\mathcal{Z}_N(\mathbb{C}(S_n))$, is a proper subspace when $N < n$ and forms a sub-algebra. Consider the set of generators $\{T_2, \cdots, T_k\}$ for some $k < n$. The particular case of $k \sim N^{1/4}, n \sim N^2$ in the limit of large N is of particular interest in connection with the information loss discussion of [12]. $n \sim N^2$ is the dimension of CFT operators which produce non-trivial deformations of the AdS space-time. $k \sim N^{1/4}$ corresponds to the Planck scale cutoff. Calculating the codimensions in this regime of k, n is a very interesting problem for the future.

4.2 Average entropies for uniform probability distribution over values of charges

Consider measuring the normalized character for the operator T_2. Given the discussion in section 2, in particular equation (2.10) this is equivalent to knowing the quadratic Casimir charge. There will be a list of normalized characters generated over the Young diagrams. The value v_2 contained in this list occurs with multiplicity $M(v_2)$. Assuming that we have no knowledge about the half-BPS state beyond the dimension n and the quadratic Casimir, this means that a total of $M(v_2)$ Young diagrams are equally likely. We thus have a uniform distribution over this subset of Young diagrams. We also have

$$
\sum_{v_2} M(v_2) = p(n).
\tag{4.7}
$$

The Shannon entropy associated with this value of v_2, and the uniform distribution, is

$$
S_{T_2}(v_2) = \log(M(v_2)).
\tag{4.8}
$$

This may be viewed as a measure of the uncertainty in our knowledge of the state R when we only know $\widehat{\chi}_2^R = v_2$. We may also take an average of these entropy values

$$
S_{T_2}^{ave} = \frac{1}{N_2} \sum_{v_2} \log(M(v_2)),
\tag{4.9}
$$

n	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$S^{ave}_{T_2}$	0.	0.	0.	0.	0.154033	0.	0.109444	0.138629	0.208835	0.13273	0.409244	0.205769	0.394343	0.529039	0.536582
$S^{ave}_{T_3}$	0.693147	0.346574	0.462098	0.51986	0.577623	0.606504	0.664379	0.755526	0.738199	0.776171	0.864573	0.863525	0.928273	0.980178	1.06048
$S^{ave}_{\{T_2,T_3\}}$	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.0120199	0.012214
n	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
$S^{ave}_{T_2}$	0.563526	0.730224	0.743094	0.854613	0.946289	1.01928	1.10351	1.21204	1.29252	1.32994	1.44694	1.52004	1.59453	1.69658	1.78313
$S^{ave}_{T_3}$	1.07416	1.11325	1.17569	1.23406	1.28993	1.33327	1.35432	1.43858	1.51161	1.50538	1.61116	1.62438	1.66037	1.73318	1.82496
$S^{ave}_{\{T_2,T_3\}}$	0.00946276	0.0147087	0.00141748	0.0193172	0.0175845	0.0200677	0.0204703	0.0405462	0.0355927	0.0315061	0.040534	0.0482671	0.046559	0.0567626	0.058324
n	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46
$S^{ave}_{T_2}$	1.83615	1.93773	1.98932	2.13505	2.15106	2.24552	2.29611	2.37268	2.44947	2.51518	2.60154	2.66398	2.72379	2.82513	2.85408
$S^{ave}_{T_3}$	1.83197	1.90237	1.94616	2.01054	2.05769	2.12388	2.12992	2.22336	2.26533	2.31421	2.39071	2.39148	2.45929	2.51554	2.58844
$S^{ave}_{\{T_2,T_3\}}$	0.0741798	0.0738499	0.0792513	0.0865217	0.0978702	0.10465	0.109309	0.115949	0.128827	0.138693	0.14455	0.1575	0.17054	0.174216	0.190026
n	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61
$S^{ave}_{T_2}$	2.93773	3.01755	3.04966	3.12983	3.18908	3.25141	3.33456	3.41051	3.48155	3.56756	3.58782	3.63649	3.70292	3.78007	3.83382
$S^{ave}_{T_3}$	2.61117	2.67504	2.71647	2.79268	2.81239	2.89125	2.90253	2.9925	3.01555	3.09754	3.12063	3.18674	3.21088	3.29807	3.31192
$S^{ave}_{\{T_2,T_3\}}$	0.202603	0.209931	0.229586	0.236442	0.249809	0.263787	0.282991	0.289623	0.309702	0.323097	0.336236	0.356257	0.376416	0.381171	0.406974
n	62	63	64	65	66	67	68	69	70						
$S^{ave}_{T_2}$	3.89228	3.97566	4.02229	4.07997	4.16048	4.22011	4.25737	4.32714	4.37174						
$S^{ave}_{T_3}$	3.39403	3.41817	3.48096	3.51733	3.58923	3.61489	3.67565	3.71282	3.77678						
$S^{ave}_{\{T_2,T_3\}}$	0.423566	0.440401	0.457433	0.47581	0.494219	0.515644	0.533373	0.553679	0.573106						

Table 2. Table showing the average entropy values $S^{ave}_{T_2}$, $S^{ave}_{T_3}$ and $S^{ave}_{\{T_2,T_3\}}$.
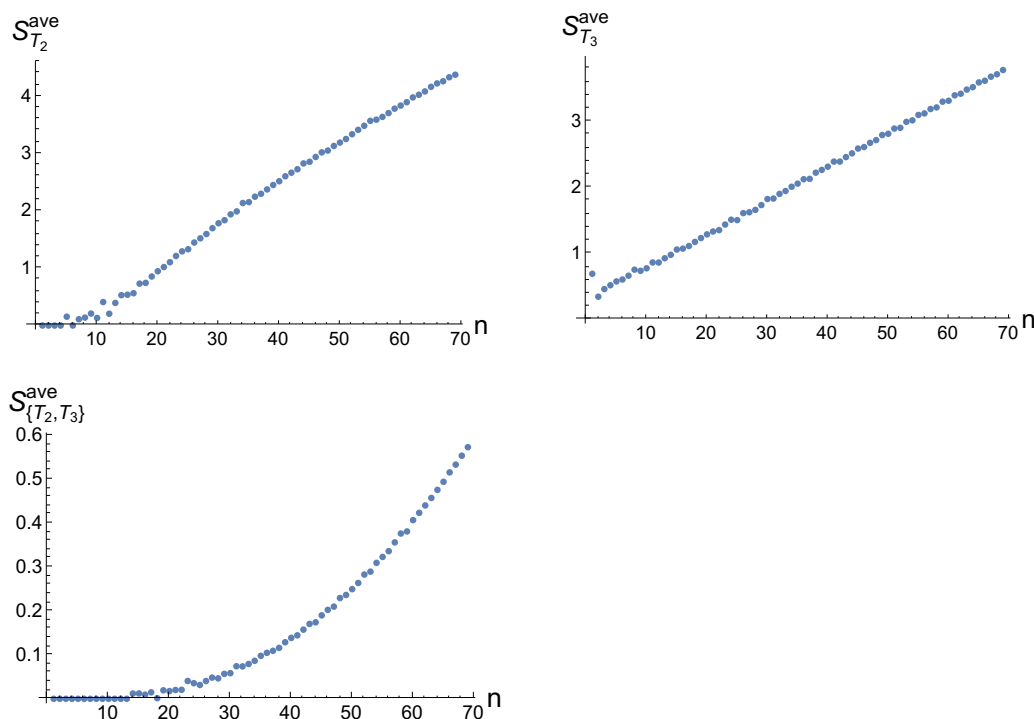
Figure 5. Figures showing the average entropy values $S^{ave}_{T_2}$, $S^{ave}_{T_3}$ and $S^{ave}_{\{T_2,T_3\}}$.

where the sum is over all distinct values of $\widehat{\chi}^R_2 = v_2$ and N_2 is the total number of distinct $\widehat{\chi}^R_2$ values. This quantity may be viewed as a measure of the average uncertainty of the Young diagram when we only know the values of the T_2 normalized characters. We present data for $S^{ave}_{T_2}$ in table 2 for $n = 2$ to $n = 70$. We also present a plot of this data in figure 5. Similar data for T_3 is presented there as well. Note that the average entropy (4.9) can also be viewed as an expectation value in a probability distribution over the values v_2, where all these values are equally probable, in other words, the uniform distribution over v_2.

We may also consider the list of values $\{v_2, v_3\}$ for $\widehat{\chi}_2^R$ and $\widehat{\chi}_3^R$. Denote the multiplicity of the pair (v_2, v_3) by $M(v_2, v_3)$. We have

$$\sum_{v_3} M(v_2, v_3) = M(v_2), \qquad \sum_{v_2} M(v_2, v_3) = M(v_3), \qquad (4.10)$$

where we sum over distinct values of $\widehat{\chi}_3^R = v_3$ in the first equation and over distinct values of $\widehat{\chi}_2^R = v_2$ in the second. $\log(M(v_2, v_3))$ is the entropy associated with the values (v_2, v_3) and the uniform distribution over the subset of R corresponding to these values for $(\widehat{\chi}_2^R, \widehat{\chi}_3^R)$. Again we may compute the average of this entropy

$$S_{\{T_2, T_3\}}^{\mathrm{ave}} = \frac{1}{N_{(2,3)}} \sum_{v_2, v_3} \log(M(v_2, v_3)) \qquad (4.11)$$

where $N_{(2,3)}$ is the total number of distinct values for $\{\widehat{\chi}_2^R, \widehat{\chi}_3^R\}$. This entropy may be viewed as a measure of the uncertainty in identifying the exact Young diagram when we only know the values of the T_2 and T_3 normalized characters.

We expect that there should be a smaller uncertainty when we know both $\{\widehat{\chi}_2^R, \widehat{\chi}_3^R\}$ compared to knowing just one of the normalized characters. In other words, we expect these average entropies to obey

$$S_{\{T_2, T_3\}}^{\mathrm{ave}} < S_{\{T_2\}}^{\mathrm{ave}}$$
$$S_{\{T_2, T_3\}}^{\mathrm{ave}} < S_{\{T_3\}}^{\mathrm{ave}} \qquad (4.12)$$

This is indeed compatible with the results in the table 5 and is also the trend we see in the comparison of the codimensions in (4.5). It is not a priori clear which $S_{\{T_2\}}^{\mathrm{ave}}$ or $S_{\{T_3\}}^{\mathrm{ave}}$ should be larger. The data shows that

$$S_{\{T_2\}}^{\mathrm{ave}} < S_{\{T_3\}}^{\mathrm{ave}} \text{ for } n \le 31$$
$$S_{\{T_3\}}^{\mathrm{ave}} < S_{\{T_2\}}^{\mathrm{ave}} \text{ for } 32 \le n \le 70 \qquad (4.13)$$

It is natural to conjecture, and would be interesting to prove that the trend visible for $32 \le n \le 70$ extends for all $n \ge 32$.

We can define finite N versions of these entropies by considering only Young diagrams with height no larger than N. In these finite N ensembles of Young diagrams, we can define multiplicities of Young diagrams and derive entropies for specified values of n. It will be interesting to obtain estimates of the finite N entropies for $n \sim N^2$, since this corresponds to classical solutions of supergravity. Also of particular interest, given the discussion [12], are the average entropies for the sets $\{\hat{\chi}_2^R, \hat{\chi}_3^R, \cdots, \hat{\chi}_k^R\}$ where $k \sim N^{1/4}$.

4.3 Average entropies for multiplicity-weighted probability distributions over values of charges

In the discussion of the average entropy over a set of known charges, we used a uniform distribution over the spectrum of charges. We found that the average entropies thus calculated satisfied inequalities which were similar to those satisfied by the co-dimensions,

lending support to the idea that both the co-dimensions and the average entropies are sensible measures of the information available from knowing a set of charges - in particular, the inequalities reflect the fact that knowing more charges reduces the uncertainty. The comparisons of the information available by knowing just one charge. e.g. T_2 versus T_3 depends on how one measures the information, whether it is through codimensions or average entropies. The data is compatible with the conjecture that at large n, a definite pattern emerges: there is more information in knowing T_3 rather than T_2, whether this is measured by looking at the codimensions or average entropies.

Given a set of charges, and an associated multiplicity of Young diagrams states, there is yet another interesting way to measure the information or conversely the uncertainty associated with that probability distribution. Suppose we take the set of values of $\widehat{\chi}_2^R$ at a given n. Let these values form the set $\mathcal{V}_2$. The multiplicity of a given value $v_2 \in \mathcal{V}_2$ is

$$M(v_2; n) = \sum_{R \vdash n} \delta(v_2, \widehat{\chi}_2^R). \tag{4.14}$$

Let N_2 be the size of the set $\mathcal{V}_2$. The probability of having value v_2 is

$$P(v_2; n) = \frac{M(v_2; n)}{p(n)} \tag{4.15}$$

since the total number of Young diagrams is $p(n)$.

$$\sum_{v_2 \in \mathcal{V}_2} P(v_2; n) = 1 \tag{4.16}$$

Now consider the Shannon entropy for this probability distribution

$$\begin{aligned} S(T_2; n) &= -\sum_{v_2} P(v_2; n) \log P(v_2; n) \\ &= -\frac{1}{p(n)} \sum_{v_2} M(v_2; n) \log M(v_2; n) + \frac{1}{p(n)} \sum_{v_2} M(v_2; n) \log p(n) \\ &= \log(p(n)) - \sum_{v_2} \frac{M(v_2; n)}{p(n)} \log M(v_2; n) \end{aligned} \tag{4.17}$$

This entropy has an interesting interpretation in a quantum information setting involving quantum measurement and classical communication. Suppose we have a density matrix for the Hilbert space of Young diagram states, where the diagonal part of the density matrix is

$$\rho_{\text{diag}} = \sum_{n=n_0}^{n_1} \frac{1}{(n_1 - n_0 + 1)} \sum_{R \vdash n} \frac{1}{p(n)} |R\rangle\langle R| \tag{4.18}$$

for a uniform probability distribution over Young diagram states with energies between n_0 and n_1. The states $|R\rangle$ have unit norm. Suppose observer A measures the energy n and the set of charges $\{T_2, \cdots, T_n\}$ to determine the exact R. Given the form of (4.18) we have a uniform probability distribution over R. The Observer A communicates the information to Observer B of the energy n, but to observer C the more detailed information of n, along

n	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
S_{T_2}	0.693147	1.09861	1.60944	1.94591	2.14584	2.70805	2.902	3.17015	3.32476	3.80255	3.66872	4.21163	4.2155	4.29012	4.46491
S_{T_3}	0	0.636514	1.05492	1.35178	1.76776	2.06111	2.16129	2.47089	2.76364	3.00383	3.16641	3.45826	3.62378	3.85848	4.02057
$S_{\{T_2,T_3\}}$	0.693147	1.09861	1.60944	1.94591	2.3979	2.70805	3.09104	3.4012	3.73767	4.02535	4.34381	4.61512	4.90527	5.14685	5.41841
n	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
S_{T_2}	4.76269	4.6738	4.99981	4.94975	5.09142	5.15615	5.36429	5.27195	5.47018	5.482	5.58016	5.59657	5.75089	5.6739	5.84508
S_{T_3}	4.21406	4.35544	4.6068	4.69534	4.92155	5.02822	5.19344	5.31327	5.47841	5.60027	5.74219	5.84809	5.99283	6.07895	6.21713
$S_{\{T_2,T_3\}}$	5.67506	5.92444	6.19158	6.40336	6.63889	6.86859	7.09512	7.28201	7.50869	7.73209	7.92885	8.12253	8.33612	8.50813	8.71314
n	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46
S_{T_2}	5.82958	5.91412	5.93302	6.02386	5.99058	6.11159	6.10852	6.17636	6.17289	6.26647	6.23725	6.33206	6.32741	6.37935	6.39015
S_{T_3}	6.2982	6.42812	6.51941	6.62493	6.71789	6.79649	6.89425	6.98144	7.05408	7.15678	7.2181	7.28986	7.37349	7.4435	7.49904
$S_{\{T_2,T_3\}}$	8.87932	9.07243	9.25758	9.42605	9.58561	9.76661	9.93823	10.1018	10.2462	10.4154	10.5627	10.7234	10.859	11.0096	11.152
n	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61
S_{T_2}	6.46051	6.44134	6.5165	6.51129	6.56246	6.56774	6.62732	6.61593	6.67486	6.67503	6.71835	6.72465	6.77441	6.76538	6.81789
S_{T_3}	7.58222	7.64015	7.69895	7.76954	7.82713	7.87246	7.94326	7.99421	8.03952	8.10671	8.14905	8.18788	8.25452	8.29205	8.33362
$S_{\{T_2,T_3\}}$	11.2967	11.4233	11.5622	11.6938	11.8266	11.9438	12.0728	12.1975	12.3155	12.4226	12.5447	12.6535	12.7626	12.8684	12.973
n	62	63	64	65	66	67	68	69	70						
S_{T_2}	6.81972	6.85732	6.86305	6.90441	6.90174	6.94541	6.94837	6.98194	6.98709						
S_{T_3}	8.3895	8.43081	8.46249	8.52008	8.55551	8.58719	8.63838	8.67634	8.70224						
$S_{\{T_2,T_3\}}$	13.076	13.1789	13.2702	13.3668	13.4665	13.5534	13.6395	13.7326	13.816						

Table 3. Table showing the entropy data for T_2, T_3 and $\{T_2, T_3\}$.

with the eigenvalues of C_2, or equivalently the normalized character of T_2. The first term in (4.17) is a measure of the uncertainty open to B, who knows that the Young diagram is one of $p(n)$ but has no further information.[2] The second term

$$\sum_{v_2} \frac{M(v_2; n)}{p(n)} \log M(v_2; n) \tag{4.19}$$

is a measure of the uncertainty $\log M(v_2; n)$ for each v_2 value averaged over the different v_2 values according to a probability with which v_2 occurs in the measurements of A. The difference is a measure of the reduction in uncertainty, due to the additional information available to C compared to B.

Equation (4.19) provides the relation between the entropy $S(T_2; n)$ above and the entropy $S_{T_2}(v_2)$ defined in equation (4.8) in section 4.2. We can view (4.19) as an expectation value of $S_{T_2}(v_2)$ taken over the probabilities $\frac{M(v_2; n)}{p(n)}$. Thus,

$$S(T_2; n) = \log(p(n)) - \mathbb{E}\left(S_{T_2}(v_2)\right). \tag{4.20}$$

It is interesting to plot the entropy (4.17) as a function of n using data we already have, which is motivated by questions such as: what is this entropy as a function of n? How does it behave at large n? This data is presented in table 3. Next consider $T_2, T_3, \cdots, T_k$. We want to think about the values $\{v_2, v_3, \cdots v_k\}$ of the normalized characters $\widehat{\chi}_2^R, \widehat{\chi}_3^R, \cdots, \widehat{\chi}_k^R$. This vector of normalized characters lives in the space $\mathcal{V}_{\{2,3,\cdots,k\}}$. The size of this set is $N_{(2,3,\cdots,k)}$. The multiplicity of a given value-set is

$$M(v_2, \cdots, v_k; n) = \sum_{R \vdash n} \delta(v_2, \widehat{\chi}_2^R) \delta(v_3, \widehat{\chi}_3^R) \cdots \delta(v_k, \widehat{\chi}_k^R). \tag{4.21}$$

---

[2]This informal discussion of uncertainty can be made more precise by using the Shannon Noiseless Coding Theorem [49, 50].
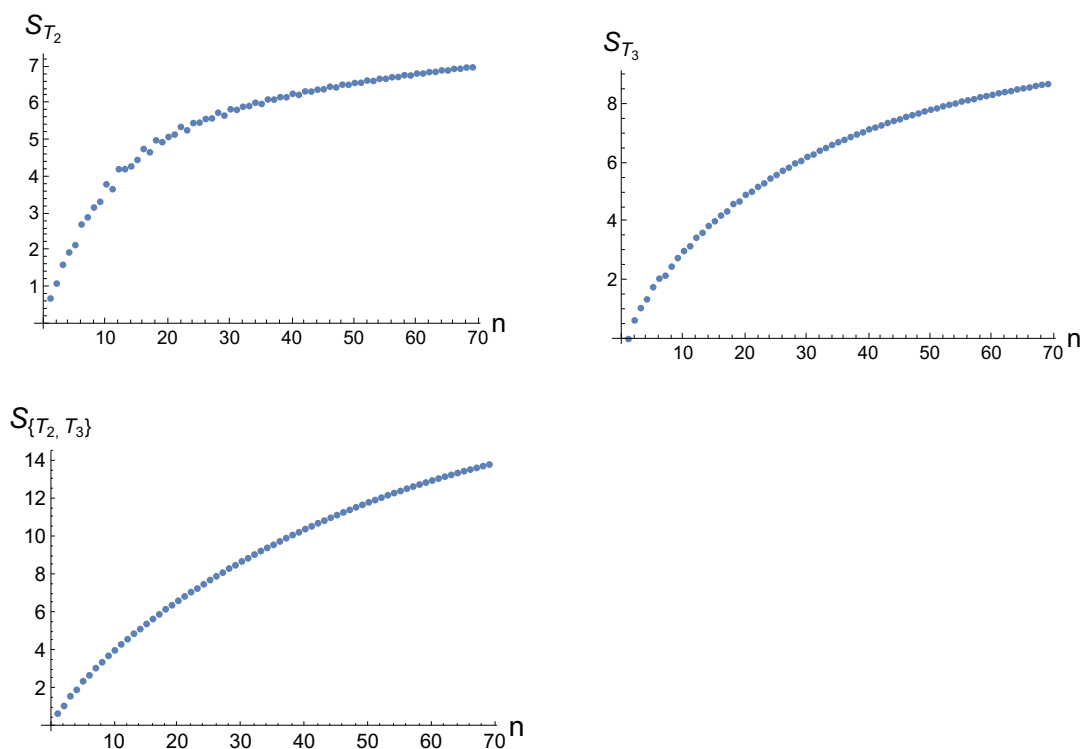
Figure 6. Figures showing entropy S_{T_2}, S_{T_3} and for $S_{\{T_2,T_3\}}$.

Now we can define an entropy for this set of generators

$$
\begin{aligned}
S(k;n) &= S(T_2, \cdots, T_k; n) \\
&= -\frac{1}{p(n)} \sum_{v_2} \sum_{v_3} \cdots \sum_{v_k} \Big[M(v_2, \cdots, v_k; n) \log M(v_2, v_3, \cdots, v_k; n) \\
&\qquad\qquad\qquad\qquad -M(v_2, v_3, \cdots, v_k; n) \log p(n) \Big]
\end{aligned}
\tag{4.22}
$$

Entropy data for $\{T_2, T_3\}$ is also found in table 3. In addition, the entropy data in this table is plotted in figure 6.

These entropies demonstrate the following behaviours. For all values of n listed in the table 3

$$
\begin{aligned}
S_{T_2} &\leq S_{\{T_2,T_3\}} \\
S_{T_3} &< S_{\{T_2,T_3\}}.
\end{aligned}
\tag{4.23}
$$

The only n values for which $S_{T_2} = S_{\{T_2,T_3\}}$ are the ones where T_2 generate the centre: $n = 2, 3, 4, 5$ and 7. This is exactly as expected from our interpretation of these entropies in terms of information gained by knowing in addition to the energy n, the specified charges. We generically gain more information from knowing more charges, unless the more limited set of charges is sufficient to determine the Young diagram entirely.

When analyzing the relative sizes between S_{T_2} and S_{T_3} the data indicates

$$
\begin{aligned}
S_{T_3} &< S_{T_2} \text{ for } n \leq 23 \\
S_{T_2} &< S_{T_3} \text{ for } 24 \leq n \leq 70.
\end{aligned}
\tag{4.24}
$$

Again, it is natural to again conjecture that $S_{T_2} < S_{T_3}$ for all n larger than 24. If this conjecture, along with the corresponding conjectures in sections 4.1 and 4.2, are true, they would support the plausible conclusion that different measures of uncertainty - co-dimensions and variations in the choice of entropy function - give the same ranking of information provided by different conserved charges in the limit of large n.

The above entropies are relevant to AdS/CFT questions when $N > n$. We can also define finite N entropies, where $n > N$, motivated by the discussion of [12]. In this case, we are interested in Young diagrams with no more than N rows, which we express as $l(R) \leq N$, using $l(R)$ to refer to the vertical length of the Young diagram.

$$M(v_2, \cdots, v_k; n, N) = \sum_{\substack{R \vdash n \\ l(R) \leq N}} \delta(v_2, \widehat{\chi}_2^R) \delta(v_3, \widehat{\chi}_3^R) \cdots \delta(v_k, \widehat{\chi}_k^R) \tag{4.25}$$

The associated entropies are

$$S(k; n, N) = -\frac{1}{p(n, N)} \sum_{v_2} \sum_{v_3} \cdots \sum_{v_k} \Big[M(v_2, \cdots, v_k; n, N) \log M(v_2, v_3, \cdots, v_k; n, N) \\ - M(v_2, v_3, \cdots, v_k; n, N) \log p(n, N) \Big]. \tag{4.26}$$

Here $p(n, N)$ is the number of Young diagrams with no more than N rows. Of particular interest, from the discussion in [12], is the large N behaviour of

$$S(k = N^{1/4}, n = N^2, N) \tag{4.27}$$

$k = N^{1/4}$ corresponds to the Planck scale, while N^2 is the dimension of CFT operators which cause a significant backreaction in the geometry. We leave a systematic computation and discussion of these finite N version of the co-dimensions and entropies we have discussed for the future.

5 Content distribution functions and diophantine equations

In section 3 we considered the problem of generating $\mathcal{Z}(\mathbb{C}(S_n))$ using central elements $\{T_2, T_3, \cdots T_k\}$. We used results in the representation theory of $\mathbb{C}(S_n)$ to show that this question can be answered by inspection of the normalized characters of the T_i. This allowed us to generate interesting data on the subspaces generated by T_2, by T_3 and the pair $\{T_2, T_3\}$. In this section, we once again rephrase the problem in terms of the so-called content polynomials, which have been used to produce elegant expressions for normalized characters in [43, 44]. We show that the normalized characters of $\{T_2, T_3, \cdots, T_k\}$ can be expressed in terms of the first k content polynomials. One can then reformulate the problem of generating the centre, and distinguishing all Young diagrams, in terms of these polynomials. This new formulation is used to write simple code in Mathematica to determine the values of n for which the first k polynomials distinguish all Young diagrams (see appendix A). In section 5 we define the content distribution function (CDF) for a Young diagram. We prove that each Young diagram can be uniquely specified by its corresponding CDF. The content polynomials can be expressed as moments of the distribution

functions, analogous to moments of a probability distribution. We show that knowledge of all n moments for a Young diagram uniquely determines its CDF. We show that the event of two diagrams having k degenerate moments ($k < n$) translates to a set of k vanishing moment equations in the difference of the two respective CDFs. Lastly we provide some examples of CDFs and CDF plots for degenerate diagrams at $n = 6, 15, 24$. We explain a visualization of the CDF plots in terms of segmented open strings with Dirichlet boundary conditions.

5.1 Content polynomials and normalized characters

An important result in [44] (lemma 3.1) is

$$\frac{\chi_R(T_k)}{d_R} = \langle c^{k-1}(R)\rangle + \sum_{l=0}^{k-2}\sum_{\lambda \vdash l} \Omega_{k-1}^{\lambda}(n)\langle c^{\lambda_1}(R)\rangle\langle c^{\lambda_2}(R)\rangle \cdots \langle c^{\lambda_r}(R)\rangle \tag{5.1}$$

We are summing over partitions $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_r)$ where $\lambda_1 + \lambda_2 + \cdots + \lambda_r = l \leq (k-2)$. The $\langle c^k(R)\rangle$ are defined as

$$\langle c^k(R)\rangle = \sum_{(i,j)\in R} (j-i)^k \tag{5.2}$$

where we are summing over the coordinates of the boxes of a Young diagram: i is the row number and j is the column number. For example the Young diagram with row lengths $[4,3,1]$ has boxes with $(i,j) \in \{(1,1),(1,2),(1,3),(1,4),(2,1),(2,2),(2,3),(3,1)\}$. The number $j - i$ is called the content of the box labeled by (i,j). $\Omega_{k-1}^{\nu}(n)$ is a polynomial of degree at most $(k-1-l)/2+1-r$. Equation (5.1) implies a triangular structure in the linear transformation between the normalized characters and the content polynomials. Knowing all the content polynomials up to $k-1$ means that we know the normalized character for T_k. For $k = 2$ we have $\widehat{\chi}_2^R = \langle c(R)\rangle$. For $k = 3$, we have $\widehat{\chi}_3^R = \langle c^2(R)\rangle + \Omega_2^1(n)\langle c(R)\rangle$. Lastly, for $k = 4$,

$$\widehat{\chi}_4^R = \langle c^3(R)\rangle + \Omega_3^1(n)\langle c(R)\rangle + \Omega_3^{(2,0)}(n)\langle c^2(R)\rangle + \Omega_3^{(1,1)}(n)\langle c(R)\rangle^2. \tag{5.3}$$

It is immediately apparent that degeneracies in the normalized characters $\widehat{\chi}_2^R$, $\widehat{\chi}_3^R$ and $\widehat{\chi}_4^R$ translate to degeneracies in the content polynomials $\langle c(R)\rangle$, $\langle c^2(R)\rangle$ and $\langle c^3(R)\rangle$.

5.2 Computing with content polynomials

If we are given the $T_2, \cdots, T_k$ what is the smallest n, where these fail to generate the centre of $\mathbb{C}(S_n)$? This question can be reformulated in terms of the content polynomials. Given $\{\langle c(R)\rangle, \langle c^2(R)\rangle, \cdots, \langle c^k(R)\rangle\}$, what is the smallest n such that the sequence of these lists as R runs over Young diagrams with n boxes has degeneracies, i.e. multiple R have the same list. The experimental answer is for k starting from 1, is displayed in table 4.

For up to $n = 5$, the first content polynomial $\langle c(R)\rangle$, where R labels the Young diagram, is able to distinguish all $R \vdash 5$. The first time $\langle c(R)\rangle$ fails to distinguish all Young diagrams is at $n = 6$. However, the set $\{\langle c(R)\rangle, \langle c^2(R)\rangle\}$ is unique to each $R \vdash 6$. These two polynomials together are then able to distinguish all Young diagrams for $n = 6$ up to

k	first n
1	6
2	15
3	24
4	42
5	80

Table 4. Table showing the smallest values of n for which $T_2, \cdots, T_k$ fail to generate the centre $\mathbb{C}(S_n)$. We give the n values for k from 1 to 5. This table was generated by computing the content polynomials for each irrep R at fixed n. Degeneracies in the content polynomials translate to degeneracies in the normalized characters.

$n = 14$. The first n for which this set has a multiplicity, and thus fails to distinguish the Young diagrams is at $n = 15$. This set is no longer unique to each R. We then extend the set to $\{\langle c(R) \rangle, \langle c^2(R) \rangle, \langle c^3(R) \rangle\}$, which then works for up to $n = 24$ where it fails for the first time. It is remarkable that just five numbers, $\{\langle c(R) \rangle, \langle c^2(R) \rangle, \langle c^3(R) \rangle, \langle c^4(R) \rangle, \langle c^5(R) \rangle\}$, are able to distinguish all Young diagrams corresponding to partitions of n up to $n = 79$. The first time these five polynomial values fail is for $n = 80$. It is instructive to look at these multiplicities and at the Young diagrams which have these degenerate content polynomials. The degeneracies (defined as multiplicities minus one, i.e. difference in length of the list of content vectors and the length of the list with repetitions removed) for $k = 1$ at $n = 6, 7, 8, 9$ are

$$2, 0, 3, 5 \tag{5.4}$$

The degeneracies for $k = 2$ at $n = 15, 16, 17, 18$ are

$$3, 4, 4, 8 \tag{5.5}$$

The degeneracies at $k = 3$ for $n = 24, 25, 26, 27, 28$ are

$$1, 0, 0, 2, 2 \tag{5.6}$$

The degeneracies for $k = 4$ at $n = 42, 43, 44, 45, 46$ are

$$2, 2, 6, 8, 8 \tag{5.7}$$

The degeneracies for $k = 5$ at $n = 80, 81, 82, 83, 84, 85$ are

$$3, 0, 2, 2, 11, 12 \tag{5.8}$$

It is interesting that the degeneracies start off very low. It is useful to look at the degenerate Young diagrams which share the same set of Casimirs, when we are at these thresholds of distinguishability.

5.3 Content distribution functions

Every box in a Young diagram has a content c given by $j - i$. For example at $n = 3$, we have 3 Young diagrams which can be described in terms of their row lengths $[3], [2, 1], [1^3]$. The contents are $\{0, 1, 2\}, \{0, 1, -1\}, \{0, -1, -2\}$. These can be plotted on a content axis labelled c, and the content multiplicities $f(c)$. The respective content multiplicities are given by $(c, f(c))$

$$
\begin{aligned}
[3] &: (0, 1), (1, 1), (2, 1) \\
[2, 1] &: (0, 1), (1, 1), (-1, 1) \\
[1^3] &: (0, 1), (-1, 1), (-2, 1)
\end{aligned}
\tag{5.9}
$$

The content distribution functions can be visualized as piecewise linear or segmented open strings defined in the content space from $-n$ to n. Each segment has slope $-1, 0$ or 1. The open string has Dirichlet boundary conditions of zero at its ends. This means that $f(\pm n) = 0$.

Proposition 1. *A Young diagram is uniquely specified by its content distribution function $f(c)$.*

We begin this proof by noting that a Young diagram has a depth d. This is the number of boxes along the diagonal at 45 degrees when to the horizontal. All the boxes along the diagonal have content 0. Alongside d, there is a set of parameters

$$
(d; k_1^+, l_1^+, \cdots, k_{p-1}^+, l_{p-1}^+, k_p^+; k_1^-, l_1^-, \cdots, k_{q-1}^-, l_{q-1}^-, k_q^-)
\tag{5.10}
$$

These parameters are illustrated in figure 7. Going up from corner of the deepest box with content 0, we have k_1^+ boxes before we reach a corner, then we go l_1^+ steps horizontally to get to the next corner, then k_2^+ steps vertically to the next corner, then l_2^+ steps horizontally. This continues until we have k_p^+ steps up for some positive p. Similarly going to the left from corner of the deepest box of content zero, we have l_1^- steps to the next corner, going down from there we have l_1^- steps to the next corner and so forth, until we get to l_q^- steps left for some positive q. These parameters uniquely specify the Young diagram. Note that

$$
\begin{aligned}
d &= k_1^+ + k_2^+ + \cdots + k_p^+ \\
d &= l_1^- + l_2^- + \cdots + l_q^-
\end{aligned}
\tag{5.11}
$$

The content distributions consist of segments of slope $0, -1, 1$. Define three functions of the content c, with parameters k, a, b. a, b are integers with $b \geq a$. k is a positive integer equal to the value of the function at a so that $\Theta(k, a, b; c = a) = k$. The three functions are

$$
\Theta_0(k, a, b; c) = k \sum_{i=a}^{b} \delta(c, i)
$$

$$
\Theta_-(k, a, b; c) = \sum_{i=a}^{b} (k + a - i)\delta(c, i)
$$

$$
\Theta_+(k, a, b; c) = \sum_{i=a}^{b} (k - a + i)\delta(c, i)
\tag{5.12}
$$
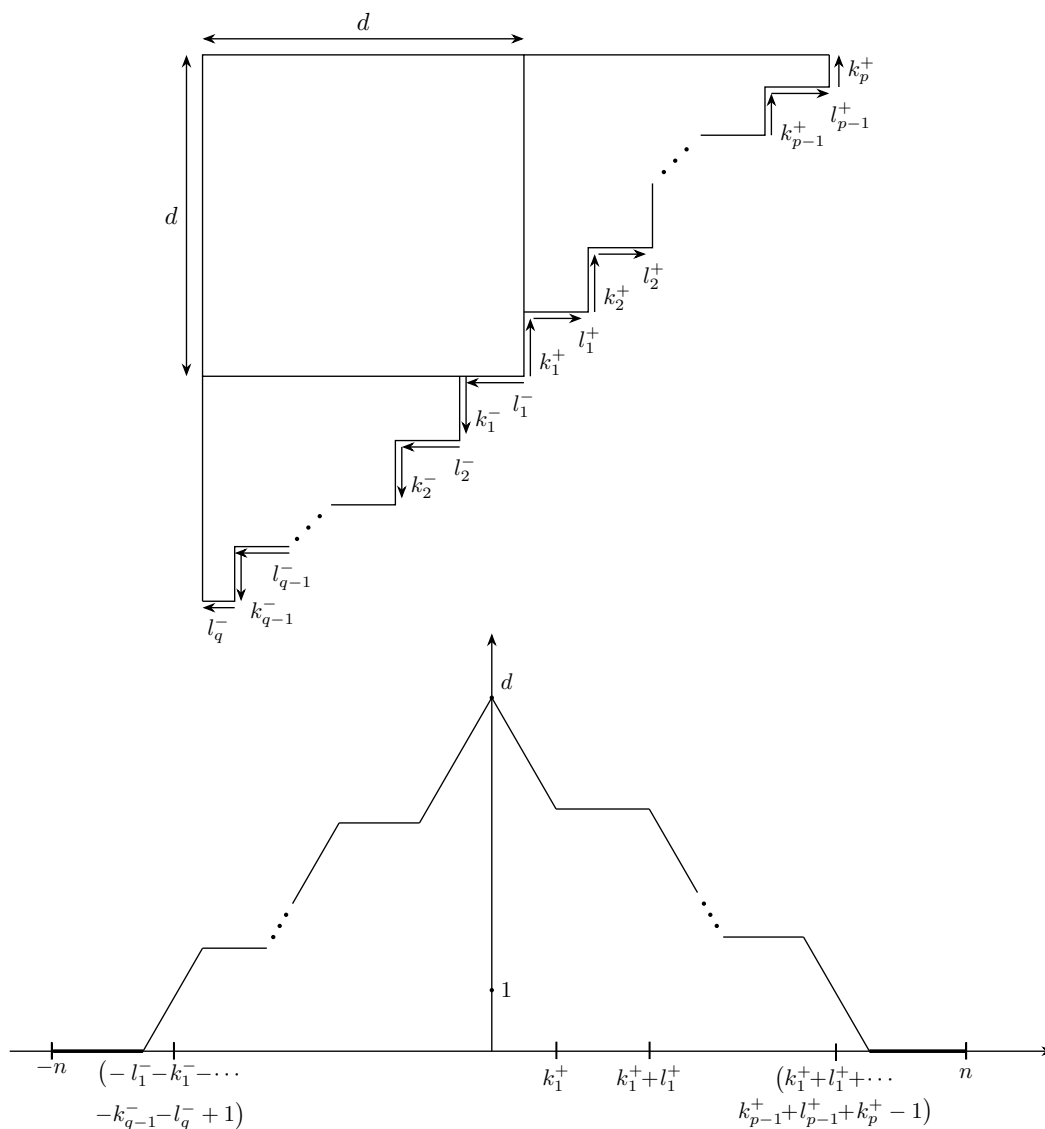
Figure 7. Above: Young diagram being specified by the parameters $d, k_i^+, l_i^+, k_i^-, l_i^-$. Below: a typical content distribution plot. The content of the boxes along the main diagonal is $c = 0$ and its multiplicity is d. The content of the box at the top right most corner of the diagram is $k_1^+ + l_1^+ + \cdots k_p^+ - 1$ and its multiplicity is 1. Similarly, the content of the box at the bottom left most corner is $l_1^- - k_1^- - \cdots l_q^- + 1$ also with multiplicity of 1. At the end points of n and $-n$, the CDF open string is at zero for all Young diagrams.

The content distribution function is easily written in terms of these.

$$f(c) = f_+(c) + f_-(c) \tag{5.13}$$

where f_+, f_- are given below.

$$
\begin{aligned}
f_+(c) = {} & \Theta_-\big(d, 0, k_1^+ - 1; c\big) \\
& + \Theta_0\big(d - k_1^+, k_1^+, k_1^+ + l_1^+ - 1; c\big) + \Theta_-\big(d - k_1^+, k_1^+ + l_1^+, k_1^+ + l_1^+ + k_2^+ - 1; c\big) \\
& + \cdots \\
& + \Theta_0\bigg(d - \sum_{i=1}^{p-1} k_i^+, \sum_{i=1}^{p-2}(k_i^+ + l_i^+) + k_{p-1}^+, \sum_{i=1}^{p-1}(k_i^+ + l_i^+) - 1; c\bigg) \\
& + \Theta_-\bigg(d - \sum_{i=1}^{p-1} k_i^+, \sum_{i=1}^{p-1}(k_i^+ + l_i^+), \sum_{i=1}^{p-1}(k_i^+ + l_i^+) + k_p^+ - 1; c\bigg) \\
& + \Theta_0\bigg(d - \sum_{i=1}^{p} k_i^+, \sum_{i=1}^{p-1}(k_i^+ + l_i^+) + k_p^+, n; c\bigg),
\end{aligned}
\tag{5.14}
$$

and

$$
\begin{aligned}
f_-(c) = {} & \Theta_+\big(d - l_1^-, -l_1^-, -1; c\big) + \Theta_0\big(d - l_1^-, -l_1^- - k_1^-, -l_1^- - 1; c\big) \\
& + \Theta_+\big(d - l_1^- - l_2^-, -l_1^- - k_1^- - l_2^-, -l_1^- - k_1^- - 1; c\big) \\
& + \Theta_0\big(d - l_1^- - l_2^-, -l_1^- - k_1^- - l_2^- - k_2^-, -l_1^- - k_1^- - l_2^- - 1; c\big) \\
& + \cdots \\
& + \Theta_0\bigg(d - \sum_{i=1}^{q-1} l_i^-, -\sum_{i=1}^{q-1}(l_i^- + k_i^-), -\sum_{i=1}^{q-2}(l_i^- + k_i^-) - l_{q-1}^- - 1; c\bigg) \\
& + \Theta_+\bigg(d - \sum_{i=1}^{q} l_i^-, -\sum_{i=1}^{q-1}(l_i^- + k_i^-) - l_q^-, -\sum_{i=1}^{q-1}(l_i^- + k_i^-) - 1; c\bigg) \\
& + \Theta_0\bigg(d - \sum_{i=1}^{q} l_i^-, -n, -\sum_{i=1}^{q-1}(l_i^- + k_i^-) - l_q^-; c\bigg).
\end{aligned}
\tag{5.15}
$$

The Young diagram in equation (5.3) illustrates an example of the counting in $f_+(c)$ and $f_-(c)$. Here, $k_1^+ = 3, k_2^+ = 2, k_3^+ = 3, l_1^+ = 5, l_2^+ = 5$. Then $l_1^- = 2, l_2^- = 3, l_3^- = 2, l_4^- = 1, k_1^- = 2, k_2^- = 2, k_3^- = 3$. The function Θ_- will count the content multiplicity of the shaded regions to the right of, and including, the main diagonal (whose content is zero). It counts content multiplicity from $c = 0$ to $k_1^+ - 1$, (i.e. from $c = 0$ to $c = 2$). Then it counts the multiplicity from $k_1^+ + l_1^+$ to $k_1^+ + l_1^+ + k_2^+ - 1$, (i.e. from $c = 8$ to $c = 9$). Lastly, Θ_- evaluates the multiplicity for contents $k_1^+ + l_1^+ + k_2^+ + l_2^+$ to $k_1^+ + l_1^+ + k_2^+ + l_2^+ + k_3^+ - 1$, (i.e. from $c = 15$ to $c = 17$). The shaded regions below the main diagonal are handled by the function Θ_+. For this example, it will count the contents -1 to $-l_1^-$ (i.e. from $c = -1$ to $c = -2$), then $-l_1^- - k_1^- - 1$ to $-l_1^- - k_1^- - l_2^-$ (i.e. from $c = -5$ to $c = -7$) and then $-l_1^- - k_1^- - l_2^- - k_2^- - 1$ to $-l_1^- - k_1^- - l_2^- - k_2^- - l_3^-$ (i.e. from $c = -10$ to $c = -11$). Lastly,

the content multiplicity of the unshaded regions are evaluated by the function Θ_0.
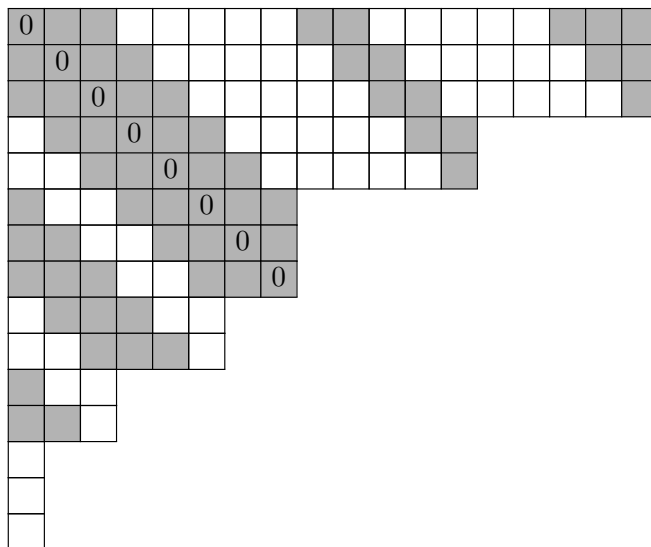
The CDF $f(c)$ is defined from $-n$ to n. For a generic Young diagram $f(c) = 0$ from $\sum_{i=1}^{p-1}(k_i^+ + l_i^+) + k_p^+$ to n. Since $d = \sum_p^{i=1} k_i^+$ (as seen from equation (5.11)), the last Θ_0 function in $f_+(c)$ is zero over this range. The slope of $f(c)$ over this range is also zero. For the special case of the Young diagram being a single row of n boxes, this Θ_0 is defined only at $c = n$, and is still equal to zero. However the slope of $f(c)$ from $c = n-1$ to $c = n$ is -1. This last Θ_0 function, even though it is equal to zero, is included in our definition so that the CDF is indeed defined up to n. Similarly for the last Θ_0 function in $f_-(c)$. Generically, $f(c) = 0$ from $-\sum_{i=1}^{q-1}(l_i^- + k_i^-) - l_q^-$ to $-n$.

Again from (5.11), $d = \sum_q^{i=1} l_i^-$. Thus, the last Θ_0 function in $f_+(c) = 0$ over this range. The CDF's slope is zero here in this region. In the special case of a single column of n boxes the slope is $+1$ from $1 - n$ to $-n$.

The Young diagram is uniquely determined by the parameters $d, k_i^+, l_i^+, k_i^-, l_i^-$ which determine the steps from the box at the end of the leading diagonal to the corners. In turn the content distribution function is uniquely determined by the location of the kinks. The correspondence between the kinks in the CDF and the corners in the Young diagram are given by the equations above.

Proposition 2. *The content distribution functions for Young diagrams with n boxes are uniquely determined by the n moments*

$$\sum_c c^k f(c) = \langle c^k \rangle \text{ for } k \in \{0, 1, \cdots, n-1\} \tag{5.16}$$

$\langle c^0 \rangle = n$.

This follows from the fact we have proved in section 3.2 that $T_2, \cdots, T_n$ generate the centre of $\mathbb{C}(S_n)$, and that these in turn are determined by the above moments.

Experimental data in fact shows that the number of moments needed is in fact lower than n. Indeed, we have already proved (remark 1) that for any n that $\{T_2, \cdots, T_{n-1}\}$ suffice to generate the centre, so the following stronger theorem is true.

Proposition 3. *The content distribution functions for Young diagrams with n boxes are uniquely determined by the $n - 1$ moments*

$$\sum_c c^k f(c) = \langle c^k \rangle \; for \; k \in \{0, 1, \cdots, n-2\} \tag{5.17}$$

Some properties of content distribution functions. They are valued in the positive integers and have the following properties.

1.

$$f(c+1) - f(c) \in \{0, 1\} \text{ for } c \geq 0$$
$$f(c) - f(c-1) \in \{0, 1\} \text{ for } c \leq 0 \tag{5.18}$$

2.

$$f(c) \text{ is a connected curve, extending from} [-c_1, c_2], \text{ where } c_1, c_2 \geq 0 \tag{5.19}$$

3. Boundary values

$$f(-c_1 + 1) = 1$$
$$f(r_1 - 1) = 1 \tag{5.20}$$

   Note that c_1 is the number of non-zero row lengths, or the length of the first column. r_1 is the length of the first row.

4. It follows from the above properties that the maximum value of f occurs at 0, and is equal to the depth of the Young diagram (i.e. the length of the diagonal).

5. It follows from the above that the maximum f_0 is $\lfloor \sqrt{n} \rfloor$.

The problem of generating the centre and distinguishing all Young diagrams at given n may now be recast in the language of the content distribution functions and moments. Given functions obeying the conditions in (5.18) (5.19) (5.20), what is the smallest positive integer $n = \langle c^0 \rangle$, such that the equations

$$\langle c^0 \rangle_f = \langle c^0 \rangle_{f'}$$
$$\vdots$$
$$\langle c^k \rangle_f = \langle c^k \rangle_{f'} \tag{5.21}$$

imply that $f = f'$ for all $m < n$. This is a reformulation of the results of section 5.2 now in terms of CDFs and moments. The data presented in table 4 are the smallest integers where the equality of moments up to k allows a multiplicity of content distribution functions. For $n = 6$, there are two degenerate Young diagram pairs. $[4, 1, 1]$ and $[3, 3]$

both have $<c>=3$. Here we are describing the Young diagrams by their row lengths. The content distributions are

$$[4,1,1] \rightarrow f(c) = 1 \text{ for the range } [-2,3]$$
$$[3,3] \rightarrow f(c) = 2 \text{ for } \{0,1\}$$
$$= 1 \text{ for } \{-1,2\}. \tag{5.22}$$

The CDF plots for these two partitions are displayed in figure 8. Define the difference between the two CDFs for $[4,1,1]$ and $[3,3]$: $\Delta_i = f_i - f'_i$ defined over $i \in [-2,3]$. This six dimensional vector is

$$\{-1,0,1,1,0,-1\}. \tag{5.23}$$

Obviously it satisfies $\sum_i \Delta_i = 0$ and $\sum_i i\Delta_i = 0$. Another degenerate pair of diagrams at $n = 6$ which satisfy the same equations is

$$[3,1^3] \rightarrow \langle c \rangle = -3$$
$$[2^3] \rightarrow \langle c \rangle = -3 \tag{5.24}$$

These are conjugates of the $[4,1,1], [3,3]$ degenerate pair. From this simple example, it seems useful to study the data in terms of these content distribution functions, and their differences. Given any two content distribution functions f_i and f'_i, define

$$\Delta_i = f_i - f'_i \tag{5.25}$$

These are positive or negative integers. These differences have the properties that

$$\Delta_{i+1} - \Delta_i \in \{0, -1, 1\} \tag{5.26}$$

Existence of degenerate moments implies that the following equations have non-tivial solutions

$$\sum_i \Delta_i = 0$$
$$\sum_i i\Delta_i = 0$$
$$\vdots$$
$$\sum_i i^k \Delta_i = 0 \tag{5.27}$$

This looks like a function with a set of vanishing moments up to k. For low values of n, there are no solutions for Δ_i to these equations. Functions with vanishing moments up to a certain maximum are studied in the literature on wavelets (see for example [51–53]). Discrete wavelets are also an active area of research. So this could be a way to approach an answer to our question.

The problem of extending the sequence $n_*(k)$ displayed in table 4 to larger values of k is very interesting. It is indeed plausible that the CDFs could play a major role in this
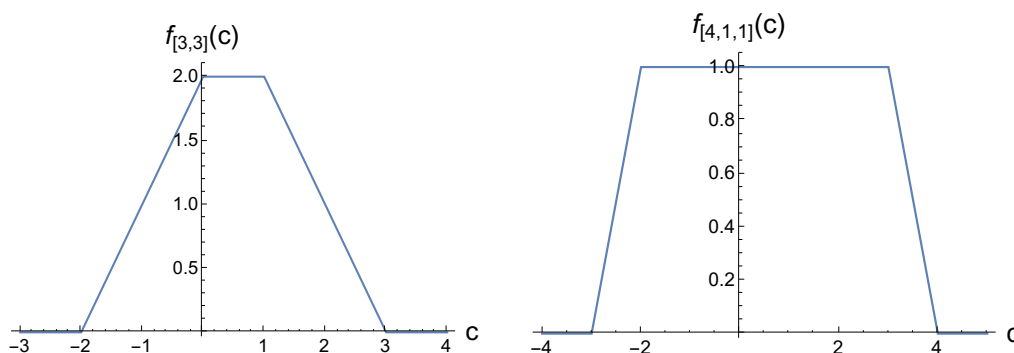
Figure 8. The CDF plots for the partitions degenerate in $\langle c \rangle$ at $n = 6$.

regard. The degeneracies that arise at $n_*(k)$ capture the failure of the first k to distinguish the Young diagrams. These degeneracies are related to the existence of vanishing moment equations for differences of CDFs. Some of these solutions are exhibited using the computer generated data in section 5.4. The CDFs may lead to new approaches for determining $n_*(k)$ drawing on number theory (diophantine equations) and probability theory.

5.4 Degenerate CDF examples at $n = 15$ and $n = 24$

At $n = 15$, there are three degenerate pairs for $\{\langle c \rangle, \langle c^2 \rangle\}$. See table 1 and equation (5.5). Below, we list these partitions, give their values for $\{\langle c \rangle, \langle c^2 \rangle\}$, as well as their CDFs. Lastly, we construct Δ for each degenerate pair.

$$\{[6, 3, 2^3], [5^2, 2, 1^3]\} \qquad \text{have}\{\langle c \rangle, \langle c^2 \rangle\} = \{0, 100\}$$
$$\{[7, 4, 2^2], [6^2, 1^3]\} \qquad \text{have}\{\langle c \rangle, \langle c^2 \rangle\} = \{15, 115\}$$
$$\{[5, 2^5], [4^2, 2^2, 1^3]\} \qquad \text{have}\{\langle c \rangle, \langle c^2 \rangle\} = \{-15, 115\} \tag{5.28}$$

For $\lambda_1 = [6, 3, 2^3]$, $\lambda_2 = [5^2, 2, 1^3]$, we have

$$f_{\lambda_1} = \{1, 2, 2, 2, 2, 2, 1, 1, 1, 1\}, \quad \text{over range} \quad [-4, 5] \tag{5.29}$$
$$f_{\lambda_2} = \{1, 1, 1, 1, 2, 2, 2, 2, 2, 1\}, \quad \text{over range} \quad [-5, 4], \tag{5.30}$$
$$\Delta_{(\lambda_1, \lambda_2)} = \{1, 0, -1, -1, 0, 0, 0, 1, 1, 0, -1\}. \tag{5.31}$$

For $\lambda_3 = [7, 4, 2^2]$, $\lambda_4 = [6^2, 1^3]$, we have

$$f_{\lambda_3} = \{1, 2, 2, 2, 2, 2, 1, 1, 1, 1\}, \quad \text{over range} \quad [-3, 6] \tag{5.32}$$
$$f_{\lambda_4} = \{1, 1, 1, 1, 2, 2, 2, 2, 2, 1\}, \quad \text{over range} \quad [-4, 5], \tag{5.33}$$
$$\Delta_{(\lambda_3, \lambda_4)} = \{1, 0, -1, -1, 0, 0, 0, 1, 1, 0, -1\}. \tag{5.34}$$
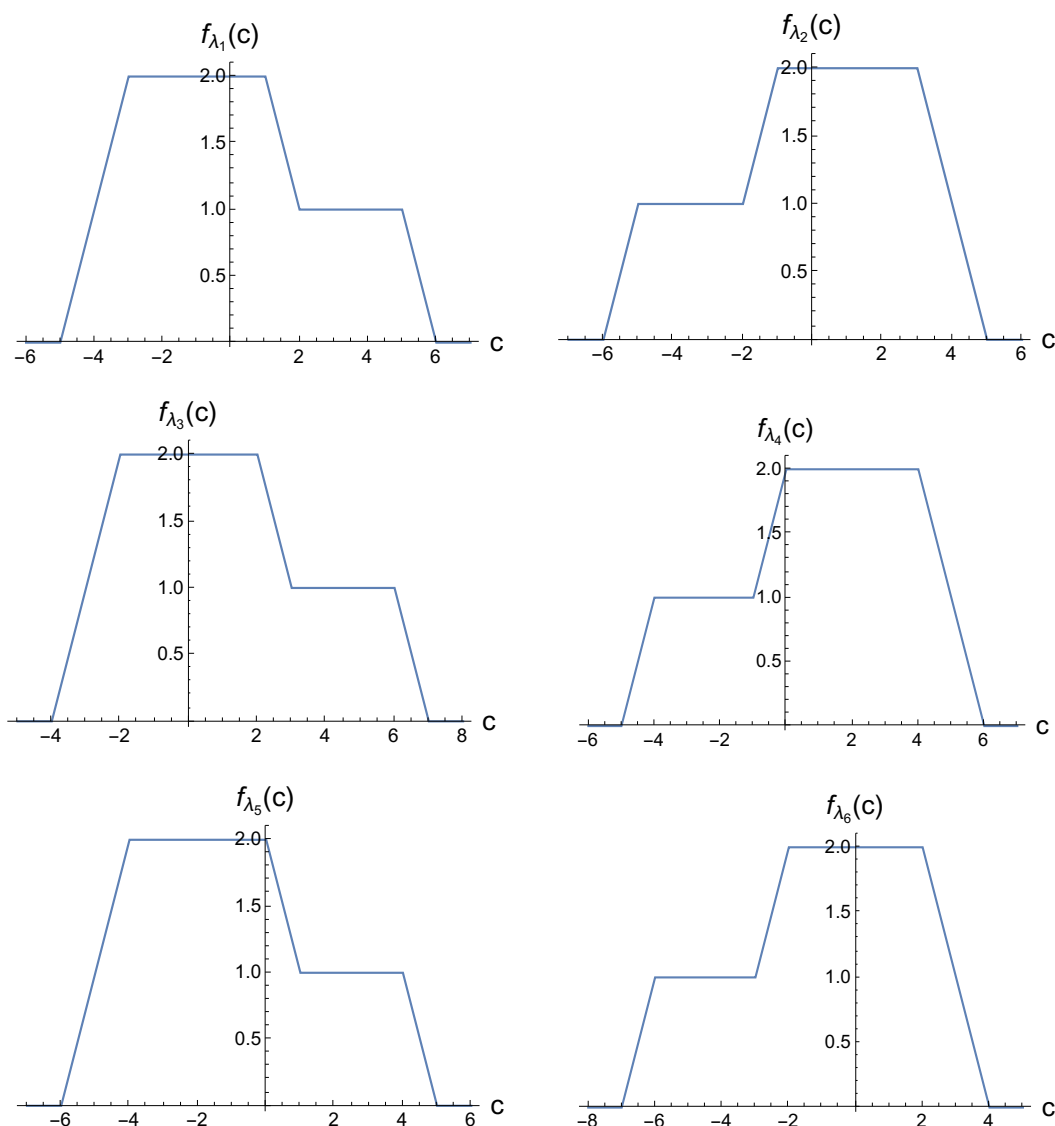
For $\lambda_5 = [5, 2^5]$, $\lambda_6 = [4^2, 2^2, 1^3]$, we have

$$f_{\lambda_5} = \{1, 2, 2, 2, 2, 2, 1, 1, 1, 1\}, \quad \text{over range} \quad [-5, 4] \tag{5.35}$$
$$f_{\lambda_6} = \{1, 1, 1, 1, 2, 2, 2, 2, 2, 1\}, \quad \text{over range} \quad [-6, 3], \tag{5.36}$$
$$\Delta_{(\lambda_5, \lambda_6)} = \{1, 0, -1, -1, 0, 0, 0, 1, 1, 0, -1\}. \tag{5.37}$$

The CDF plots for the three degenerate pairs are displayed in figure 9.

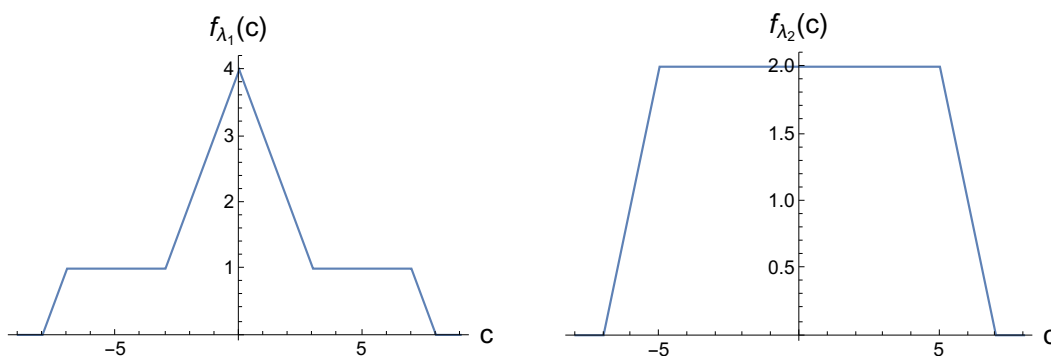**Figure 9**. The CDF plots for the partitions degenerate in $\{\langle c \rangle, \langle c^2 \rangle\}$ at $n = 15$.

As a final example, at $n = 24$, has a single degenerate pair for moments $\{\langle c \rangle, \langle c^2 \rangle, \langle c^3 \rangle\}$. See equation (5.6). The partitions are $\{[8, 4^3, 1^4], [7^2, 2^5]\}$ and the values for $\{\langle c \rangle, \langle c^2 \rangle, \langle c^3 \rangle\} = \{0, 292, 0\}$, and the partitions are $\{[8, 4^3, 1^4], [7^2, 2^5]\}$. The CDFs and Δ are

$$f_{\lambda_1} = \{1, 1, 1, 1, 1, 2, 3, 4, 3, 2, 1, 1, 1, 1, 1\}, \quad \text{over range} \quad [-7, 7] \tag{5.38}$$

$$f_{\lambda_2} = \{1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1\}, \quad \text{over range} \quad [-6, 6], \tag{5.39}$$

$$\Delta_{(\lambda_1, \lambda_2)} = \{-1, 0, 1, 1, 1, 0, -1, -2, -1, 0, 1, 1, 1, 0, -1\}. \tag{5.40}$$

The CDF plots for these two partitions are shown in figure 10.

Figure 10. The CDF plots for the partitions degenerate in $\{\langle c \rangle, \langle c^2 \rangle, \langle c^3 \rangle\}$ at $n = 24$.

6 Summary and outlook

We have proved that central elements $T_2, T_3, \cdots, T_n$ associated with reduced cycle structures $\mathcal{C}_{(1)}, \mathcal{C}_{(2)}, \cdots, \mathcal{C}_{(n-1)}$ generate the centre $\mathcal{Z}(\mathbb{C}(S_n))$. We then showed that restricting to a subset $\{T_2, \cdots, T_k\}$ generates the centre for all n up to $n_*(k)$. We used computational methods to determine $n_*(k)$ for k up to 6. For the classes T_2, T_3 and the collection $\{T_2, T_3\}$, we computed the dimensions of the subspaces generated. We showed that these dimensions are directly encoded in the normalized characters of these conjugacy classes in Young diagrams R with n boxes, which are in turn related to Casimir eigenvalues for the U(N) representation associated with R. The multiplicities of the normalized characters can be used to quantify the amount of information available with specified sets such as $\{T_2\}$, $\{T_3\}$ or $\{T_2, T_3\}$, using Shannon entropy functions. These entropies were calculated and led to the conclusion that the dimensions as well as entropies give sensible measures of the amount of information available with the normalized characters, equivalently specified Casimir charges. We presented some conjectures on the large n behaviours of relative dimensions and entropies for T_2 and T_3, based on the plausible expectation that at large n the different measures should give the same ranking.

We have observed that the power-sums of contents can be viewed as moments of a content distribution function. This is simply a discrete function on a finite set of points in the range $[-n+1, n-1]$, which was shown to uniquely determine a Young diagram. Some initial steps in the direction of using CDFs in order to understand the degeneracies between normalized characters that occur for n just above $n_*(k)$ have been made. We have observed that differences of Young diagram CDFs obey some diophantine vanishing moment conditions above $n_*(k)$. Using computational work, we obtained some instances of these solutions to diophantine equations. It is reasonable to expect that a combination of techniques from combinatorics and number theory will, in the future, allow a general analytic treatment of these diophantine equations for general k and provide further information on $n_*(k)$ as k increases.

The centre $\mathcal{Z}(\mathbb{C}(S_n)$ is one of a class of interesting permutation centralizer algebras which are relevant to multi-matrix and tensor invariants [27, 29]. There is a 2-parameter

algebra $\mathcal{A}(m,n)$ relevant to the 2-matrix system with U(N) gauge symmetry, with structure closely related to Littlewood-Richardson coefficients. There is a 1-parameter algebra $\mathcal{K}(n)$ relevant to 3-index tensor systems, which is closely related to Kronecker coefficients. The algebras $\mathcal{A}(m,n)$ were recently used to derive identities involving contents of Young diagrams, which have applications in quantum information processing tasks [55]: the present paper is another link between permutation algebras and information theoretic perspectives directly motivated, in the present case, by information theoretic questions in AdS/CFT. Another connection between the 2-matrix system and small black holes in AdS/CFT is proposed in [56]. It is evident that we are only beginning to scratch the surface of the story linking AdS/CFT, information and permutation algebras. Analogous algebras play a role in matrix/tensor systems with O(N)/Sp(N) symmetry [57–62]. The results of this paper, developing the connection between Casimirs and the structure of permutation algebras should admit a generalization to these cases. It will be fascinating to explore these systems using the combination of analytic and computational techniques we have used here, to generate sequences analogous to $n_*(k)$ for these cases.

Acknowledgments

A   Mathematica code

We begin by writing code to calculate the content polynomial or the content power sum for a partition specified by P.

```
ContentPowerSum[k_,P_]:=
        Sum[Sum[(a-m)^k,{a,1,P[[m]]}],{m,1,Length[P]}]
ListContentPowerSums[k_,n_]:=
        Table[ContentPowerSum[k,P],{P,IntegerPartitions[n]}]
```

The integer k specifies the power of the terms in the sum, while P specifies the actual partition. For example, taking partitions of 3, and $k = 2$, the above code computes, for the three partitions $(3), (2,1)$ and $(1,1,1)$:

```
In[1]:= ContentPowerSum[2,{3}]
        ContentPowerSum[2,{2, 1}]
        ContentPowerSum[2,{1, 1, 1}]

Out[1]= 5

Out[2]= 2

Out[3]= 5.
```

The definition in the second line of code "ListContentPowerSums" gives a list of these power sums that runs over all the partitions on n. After running the above code for $k = 1$ and $n = 6$, we find:

```
In[3]:= ListContentPowerSums [1,6]

Out[3]= {15, 9, 5, 3, 3, 0, -3, -3, -5, -9, -15}
```

Now we wish to compare the lists of the content power sums for different values of k. Below S will be a set of positive integers specifying the powers of contents to be summed over; this function will produce the list of vectors of content power sums for the partitions of n, with the powers specified by S:

```
ListSetContentPowerSums[S_ ,n_ ]:=
        Table[Table[ContentPowerSum[i,P],{i,S}],
            {P,IntegerPartitions[n]}]
```

This code will allow us to compare the set $\{\langle c(R)\rangle, \langle c(R)^2\rangle, \cdots\}$. For example

```
In[1]:= ListSetContentPowerSums[{1,2},6]

Out[1]= {{15,55}, {9,31}, {5,15}, {3,19}, {3,7}, {0,10},
        {-3,19}, {-3,7}, {-5,15}, {-9,31}, {-15,55}}.
```

We can see that the list $\{\langle c(R)\rangle, \langle c(R)^2\rangle\}$ contain no degeneracies at $n = 6$. The code to compute codimension data for T_2, then T_3 and then for $\{T_2, T_3\}$ respectively is found below.

```
Table[Length[ListSetContentPowerSums[{1},n]] -
        Length[DeleteDuplicates[ListSetContentPowerSums[{1},n]]],
        {n,Range[2,70]}]

Table[Length[ListSetContentPowerSums[{2},n]] -
        Length[DeleteDuplicates[ListSetContentPowerSums[{2},n]]],
        {n,Range[2,70]}]

Table[Length[ListSetContentPowerSums[{1,2},n]] -
        Length[DeleteDuplicates[ListSetContentPowerSums[{1,2},n]]],
        {n,Range[2,70]}].
```

The idea here is simply to generate the list of content polynomials, and count the number of uniques elements by subtracting the length of the list when duplicates have been deleted. The above code generates codimension data for n from $n = 2$ to $n = 70$.

References

[1] J.M. Maldacena, *The large N limit of superconformal field theories and supergravity*, [*Int. J. Theor. Phys.* **38** (1999) 1113](#) [*Adv. Theor. Math. Phys.* **2** (1998) 231] [[hep-th/9711200](#)] [[INSPIRE](#)].

[2] J. McGreevy, L. Susskind and N. Toumbas, *Invasion of the giant gravitons from Anti-de Sitter space*, *JHEP* **06** (2000) 008 [`hep-th/0003075`] [ɪɴSPIRE].

[3] J.M. Maldacena and A. Strominger, *AdS₃ black holes and a stringy exclusion principle*, *JHEP* **12** (1998) 005 [`hep-th/9804085`] [ɪɴSPIRE].

[4] V. Balasubramanian, M. Berkooz, A. Naqvi and M.J. Strassler, *Giant gravitons in conformal field theory*, *JHEP* **04** (2002) 034 [`hep-th/0107119`] [ɪɴSPIRE].

[5] S. Corley, A. Jevicki and S. Ramgoolam, *Exact correlators of giant gravitons from dual $N = 4$ SYM theory*, *Adv. Theor. Math. Phys.* **5** (2002) 809 [`hep-th/0111222`] [ɪɴSPIRE].

[6] D. Berenstein, *A toy model for the AdS/CFT correspondence*, *JHEP* **07** (2004) 018 [`hep-th/0403110`] [ɪɴSPIRE].

[7] A. Bissi, C. Kristjansen, D. Young and K. Zoubos, *Holographic three-point functions of giant gravitons*, *JHEP* **06** (2011) 085 [`arXiv:1103.4079`] [ɪɴSPIRE].

[8] H. Lin, *Giant gravitons and correlators*, *JHEP* **12** (2012) 011 [`arXiv:1209.6624`] [ɪɴSPIRE].

[9] C. Kristjansen, S. Mori and D. Young, *On the regularization of extremal three-point functions involving giant gravitons*, *Phys. Lett.* **B 750** (2015) 379 [`arXiv:1507.03965`] [ɪɴSPIRE].

[10] P. Caputa, R. de Mello Koch and K. Zoubos, *Extremal versus non-extremal correlators with giant gravitons*, *JHEP* **08** (2012) 143 [`arXiv:1204.4172`] [ɪɴSPIRE].

[11] H. Lin, O. Lunin and J.M. Maldacena, *Bubbling AdS space and 1/2 BPS geometries*, *JHEP* **10** (2004) 025 [`hep-th/0409174`] [ɪɴSPIRE].

[12] V. Balasubramanian, B. Czech, K. Larjo and J. Simon, *Integrability versus information loss: a simple example*, *JHEP* **11** (2006) 001 [`hep-th/0602263`] [ɪɴSPIRE].

[13] R.C. Myers and O. Tafjord, *Superstars and giant gravitons*, *JHEP* **11** (2001) 009 [`hep-th/0109127`] [ɪɴSPIRE].

[14] V. Balasubramanian, V. Jejjala and J. Simon, *The library of Babel*, *Int. J. Mod. Phys.* **D 14** (2005) 2181 [`hep-th/0505123`] [ɪɴSPIRE].

[15] V. Balasubramanian, J. de Boer, V. Jejjala and J. Simon, *The library of Babel: on the origin of gravitational thermodynamics*, *JHEP* **12** (2005) 006 [`hep-th/0508023`] [ɪɴSPIRE].

[16] D. Berenstein and A. Miller, *Reconstructing spacetime from the hologram, even in the classical limit, requires physics beyond the Planck scale*, *Int. J. Mod. Phys.* **D 25** (2016) 1644012 [`arXiv:1605.05288`] [ɪɴSPIRE].

[17] V. Balasubramanian and B. Czech, *Quantitative approaches to information recovery from black holes*, *Class. Quant. Grav.* **28** (2011) 163001 [`arXiv:1102.3566`] [ɪɴSPIRE].

[18] J. Simon, *Correlations vs. connectivity in R-charge*, *JHEP* **10** (2018) 048 [`arXiv:1805.11279`] [ɪɴSPIRE].

[19] V. Balasubramanian et al., *Emergent classical spacetime from microstates of an incipient black hole*, *JHEP* **01** (2019) 197 [`arXiv:1810.13440`] [ɪɴSPIRE].

[20] V. Balasubramanian, D. Berenstein, B. Feng and M.-x. Huang, *D-branes in Yang-Mills theory and emergent gauge symmetry*, *JHEP* **03** (2005) 006 [`hep-th/0411205`] [ɪɴSPIRE].

[21] V. Balasubramanian, M.-x. Huang, T.S. Levi and A. Naqvi, *Open strings from $N = 4$ superYang-Mills*, *JHEP* **08** (2002) 037 [`hep-th/0204196`] [ɪɴSPIRE].

[22] Y. Kimura and S. Ramgoolam, *Branes, anti-branes and brauer algebras in gauge-gravity duality*, *JHEP* **11** (2007) 078 [`arXiv:0709.2158`] [ɪɴSPIRE].

[23] T.W. Brown, P.J. Heslop and S. Ramgoolam, *Diagonal multi-matrix correlators and BPS operators in N = 4 SYM*, *JHEP* **02** (2008) 030 [arXiv:0711.0176] [INSPIRE].

[24] T.W. Brown, P.J. Heslop and S. Ramgoolam, *Diagonal free field matrix correlators, global symmetries and giant gravitons*, *JHEP* **04** (2009) 089 [arXiv:0806.1911] [INSPIRE].

[25] R. Bhattacharyya, S. Collins and R. de Mello Koch, *Exact multi-matrix correlators*, *JHEP* **03** (2008) 044 [arXiv:0801.2061] [INSPIRE].

[26] R. Bhattacharyya, R. de Mello Koch and M. Stephanou, *Exact multi-restricted Schur polynomial correlators*, *JHEP* **06** (2008) 101 [arXiv:0805.3025] [INSPIRE].

[27] Y. Kimura and S. Ramgoolam, *Enhanced symmetries of gauge theory and resolving the spectrum of local operators*, *Phys. Rev.* **D 78** (2008) 126003 [arXiv:0807.3696] [INSPIRE].

[28] E. Witten, *Anti-de Sitter space and holography*, *Adv. Theor. Math. Phys.* **2** (1998) 253 [hep-th/9802150] [INSPIRE].

[29] P. Mattioli and S. Ramgoolam, *Permutation centralizer algebras and multi-matrix invariants*, *Phys. Rev.* **D 93** (2016) 065040 [arXiv:1601.06086] [INSPIRE].

[30] Y. Kimura, *Noncommutative Frobenius algebras and open-closed duality*, arXiv:1701.08382 [INSPIRE].

[31] J. Pasukonis and S. Ramgoolam, *Quivers as calculators: counting, correlators and Riemann surfaces*, *JHEP* **04** (2013) 094 [arXiv:1301.1980] [INSPIRE].

[32] Y. Kimura, *Multi-matrix models and noncommutative Frobenius algebras obtained from symmetric groups and Brauer algebras*, *Commun. Math. Phys.* **337** (2015) 1 [arXiv:1403.6572] [INSPIRE].

[33] D.J. Gross and W. Taylor, *Twists and Wilson loops in the string theory of two-dimensional QCD*, *Nucl. Phys.* **B 403** (1993) 395 [hep-th/9303046] [INSPIRE].

[34] R. Goodman and N.R. Wallach, *Representations and invariants of finite groups*, Cambridge University Press, Cambridge U.K. (1998).

[35] S. Doty, A. Lauve and G.H. Seelinger, *Canonical idempotents of multiplicity-free families of algebras*, arXiv:1606.08900.

[36] J. Ben Geloun and S. Ramgoolam, *Counting tensor model observables and branched covers of the 2-sphere*, arXiv:1307.6490 [INSPIRE].

[37] J. Ben Geloun and S. Ramgoolam, *Tensor models, Kronecker coefficients and permutation centralizer algebras*, *JHEP* **11** (2017) 092 [arXiv:1708.03524] [INSPIRE].

[38] P. Diaz and S.-J. Rey, *Orthogonal bases of invariants in tensor models*, *JHEP* **02** (2018) 089 [arXiv:1706.02667] [INSPIRE].

[39] R. de Mello Koch, D. Gossman and L. Tribelhorn, *Gauge invariants, correlators and holography in bosonic and fermionic tensor models*, *JHEP* **09** (2017) 011 [arXiv:1707.01455] [INSPIRE].

[40] R. De Mello Koch, D. Gossman, N. Hasina Tahiridimbisoa and A.L. Mahu, *Holography for tensor models*, arXiv:1910.13982 [INSPIRE].

[41] H. Itoyama, A. Mironov and A. Morozov, *Tensorial generalization of characters*, *JHEP* **12** (2019) 127 [arXiv:1909.06921] [INSPIRE].

[42] R.C. Avohou, J. Ben Geloun and N. Dub, *On the counting of O(N) tensor invariants*, arXiv:1907.04668 [INSPIRE].

[43] M. Lassalle, *An explicit formula for the characters of the symmetric group*, *Math. Ann.* **340** (2008) 383.

[44] S. Corteel, A. Goupil and G. Schaeffer, *Content evaluation and class symmetric functions*, *Adv. Math.* **188** (2004) 315.

[45] S. Cordes, G.W. Moore and S. Ramgoolam, *Lectures on 2D Yang-Mills theory, equivariant cohomology and topological field theories*, *Nucl. Phys. Proc. Suppl.* **41** (1995) 184 [hep-th/9411210] [INSPIRE].

[46] O. Ganor, J. Sonnenschein and S. Yankielowicz, *The string theory approach to generalized 2 − D Yang-Mills theory*, *Nucl. Phys.* **B 434** (1995) 139 [hep-th/9407114] [INSPIRE].

[47] Wikipedia, *Set partitions*, https://en.wikipedia.org/wiki/Partition_of_a_set.

[48] Wikipedia, *Stirling numbers of the second kind*, https://en.wikipedia.org/wiki/Stirling_numbers_of_the_second_kind.

[49] M.A. Nielsen and I.L Chuang, *Quantum computation and quantum information*, Cambridge University Press, Cambridge U.K. (2010).

[50] P. Grunwald and P. Vitanyi, *Shannon information and Kolmogorov complexity*, cs/0410002.

[51] S. Patilkulkarni and H.C. Vijaylakshmi, *Vanishing moments of a wavelet system and feature set in face detection problem for color images*, *Int. J. Comput. Appl.* **66** (2013) 36.

[52] M. Stephane, *A wavelet tour of signal processing*, 3rd edition, Academic Press, U.S.A. (2009).

[53] F. Keinert, *Wavelets and multiwavelets*, Chapman and Hall/CRC, U.S.A. (2003).

[54] I.G. Macdonald, *Symmetric functions and Hall polynomials*, Oxford classic texts in the physical sciences, Clarendon Press, Oxford U.K. (1998).

[55] S. Ramgoolam and M. Sedlák, *Quantum information processing and composite quantum fields*, *JHEP* **01** (2019) 170 [arXiv:1809.05156] [INSPIRE].

[56] D. Berenstein, *Submatrix deconfinement and small black holes in AdS*, *JHEP* **09** (2018) 054 [arXiv:1806.05729] [INSPIRE].

[57] O. Aharony, Y.E. Antebi, M. Berkooz and R. Fishman, *'Holey sheets': Pfaffians and subdeterminants as D-brane operators in large N gauge theories*, *JHEP* **12** (2002) 069 [hep-th/0211152] [INSPIRE].

[58] P. Caputa, R. de Mello Koch and P. Diaz, *Operators, correlators and free fermions for SO(N) and Sp(N)*, *JHEP* **06** (2013) 018 [arXiv:1303.7252] [INSPIRE].

[59] P. Caputa, R. de Mello Koch and P. Diaz, *A basis for large operators in N = 4 SYM with orthogonal gauge group*, *JHEP* **03** (2013) 041 [arXiv:1301.1560] [INSPIRE].

[60] G. Kemp, *Restricted Schurs and correlators for SO(N) and Sp(N)*, *JHEP* **08** (2014) 137 [arXiv:1406.3854] [INSPIRE].

[61] G. Kemp, SO(N) *restricted Schur polynomials*, *J. Math. Phys.* **56** (2015) 022302 [arXiv:1405.7017] [INSPIRE].

[62] C. Lewis-Brown and S. Ramgoolam, *BPS operators in* $\mathcal{N} = 4$ SO(N) *super Yang-Mills theory: plethysms, dominoes and words*, *JHEP* **11** (2018) 035 [arXiv:1804.11090] [INSPIRE].