

Raz and Spieker on the log-rank conjecture, revisited

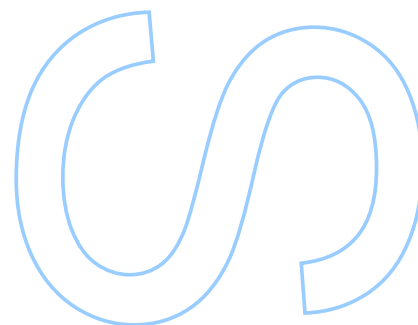
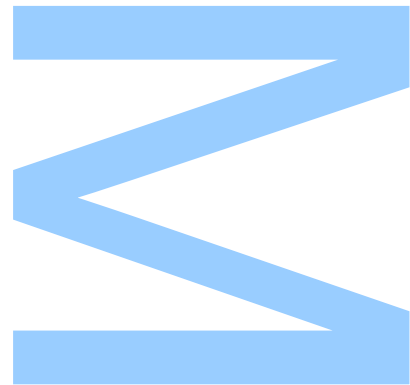
Jorge Pereira

Master in Mathematical Engineering

[Department of Mathematics](#)

Faculty of Sciences of the University of Porto

2024



Raz and Spieker on the log-rank conjecture, revisited

Jorge Pereira

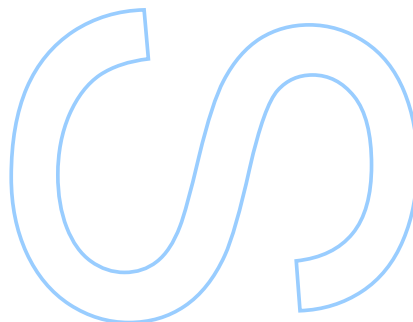
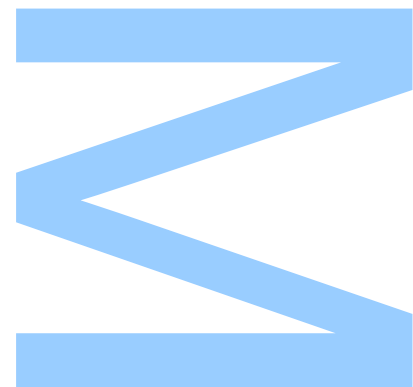
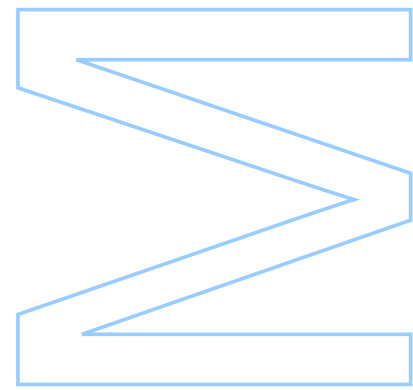
Dissertation carried out as part of
the Masters in Mathematical Engineering

[Department of Mathematics](#)

2024

Supervisor

[Bruno Loff](#), Associate Professor,
Faculty of Sciences of the University of Lisbon



Acknowledgements

I thank Bruno Loff greatly, for all the help and opportunities provided, without him this volume of work would not be possible. His vision and creativity guided me through some of the most wonderful fields in mathematics, with whom I've grown and matured as a mathematician. I extend that gratitude to the LASIGE department at the Faculty of Sciences of the University of Lisbon, for having me for a year.

A special thanks goes also to all the teachers I've had in the past years, they certainly helped me become the mathematician I am today.

Finally I want to thank the Faculty of Sciences of the University of Porto, for all I've learned there and everyone I've met. A place that will forever be in my memory, fondly.

Abstract

We revisit the *R. Raz* and *B. Spieker* paper "On the "log rank"-conjecture in communication complexity" [RS95], explaining the results therein in full detail and paying particular attention to the first part of the paper, and its use of Representation Theory. For that we include a full exposition of Representation Theory of Finite Groups, focused on the modern approach to the representation theory of symmetric groups, by *A.M. Vershik* and *A.Yu. Okounkov* in "A New Approach to the Representation Theory of the Symmetric Groups" [AA05].

Keywords: representation theory, Okounkov-Vershik approach, symmetric groups, complexity theory, communication complexity theory, log-rank conjecture

Table of Contents

List of Figures	vi
Notation	vii
1. Introduction	1
2. Communication Complexity Theory	4
2.1. Model and complexity	4
2.1.1. Yao's model	5
2.1.2. The fundamental theorem of C.C.	7
2.2. Lower bounds	10
2.3. Notes on nondeterminism	14
2.4. The log-rank conjecture	18
3. On the log rank conjecture - Part II	23
4. Representation theory of finite groups	52
4.1. Representations	53
4.1.1. Definitions	53
4.1.2. Examples	59
4.1.3. G -homomorphisms and equivalent representations	63
4.1.4. G -subspaces and direct sums	65
4.1.5. The dual representation	69
4.1.6. Tensor products	71
4.1.7. Matrix coefficients	75
4.2. Complete reducibility	76
4.2.1. Maschke's theorem	76
4.2.2. Schur's lemma	80
4.2.3. Multiplicities and isotypic components	84
4.2.4. The structure of $\text{End}_G(V)$	86
4.3. Character theory	89
4.3.1. Group characters	91
4.3.2. Orthogonality of group characters	99
4.3.3. Consequences	104
4.4. The group algebra	111
4.5. Normal subgroups and lifting	117
4.6. Induced representations	121
4.7. Permutation representations	128
4.7.1. Gelfand pairs	134

5. The Okounkov–Vershik approach	140
5.1. Multiplicity free subgroups	140
5.2. Classical representation theory of S_n	144
5.2.1. Young tableaux	145
5.2.2. Specht modules	148
5.2.3. Young graph	150
5.2.4. Coxeter generators	156
5.3. The O-V approach	159
5.3.1. The Gelfand-Tsetlin basis	160
5.3.2. The Gelfand-Tsetlin algebra	164
5.3.3. Simple branching	169
5.4. Modern representation theory of S_n	170
5.4.1. Simple branching for S_n	171
5.4.2. The YJM elements	173
5.4.3. Coxeter generators II	179
5.4.4. Content vectors	184
5.4.5. The spectrum of the YJM elements	190
5.4.6. Young’s seminormal and orthogonal forms	198
5.4.7. The hook-length formula	204
5.5. Character theory of S_n	213
5.5.1. The case of a cycle	214
5.5.2. The Littlewood-Richardson rule	216
5.5.3. The Murnhagan–Nakayama rule	221
6. On the log rank conjecture - Part I	227
7. Conclusion	233
Bibliography	233
A. Binary trees	238
B. Entropy	240
C. Linear Algebra	244
D. Modules and Algebras	249
D.1. Module theory	249
D.1.1. Modules	249
D.1.2. Direct sums and products	253
D.1.3. Tensor products	256
D.1.4. Semisimple modules	257
D.2. Algebras	258
D.2.1. Involutive algebras	259

D.2.2. Complexification.....	259
------------------------------	-----

List of Figures

2.1. Example of protocol for a binary function $f : X \times Y \rightarrow \{0, 1\}$ given in the upper left corner.....	7
2.2. Example of partitioning of the input space of the function f by the protocol given in figure 2.1.....	8
2.3. Example of a 0-monochromatic rectangle of a function $f : \{0, 1\}^3 \times \{0, 1\}^3 \rightarrow \{0, 1\}$. 9	
2.4. Matrix M_{EQ} of the equality function $EQ : \{0, 1\}^3 \times \{0, 1\}^3 \rightarrow \{0, 1\}$	13
2.5. Example of a covering of the input space of a function $f : X \times Y \rightarrow \{0, 1\}$, into two 0-monochromatic rectangles and two overlapping 1-monochromatic rectangles.	16
2.6. Partition of the input space of a function $f : X \times Y \rightarrow \{0, 1\}$ into five monochromatic rectangles.	16
2.7. Complete bipartite graph $K_{3,3}$	19
2.8. A perfect matching on the complete bipartite graph $K_{3,3}$	20
2.9. The union of two perfect matchings of the graph $K_{3,3}$	21
2.10The union of two perfect matchings of the graph $K_{3,3}$, rearranged.....	21
3.1. Example of union of matching f and g_f	23
3.2. Example of union of matching f and g_f , rearranged.	23
3.3. Plot of $\ln(-\ln(x))$ and marked inflection point.	42
3.4. Plot of $\ln(-\ln(x))$ (dashed) and its inflection point, and plot of $\ln(2 - \ln(x + 0.01))$ and its inflection point.....	43
5.1. Bottom part of the Young graph, starting at level $n = 4$ going down to $n = 1$	152
5.2. Branching graph of the chain $S_1 \leq S_2 \leq S_3 \leq S_4$	161
5.3. Two partitions λ, λ' at level n both covering some μ_1 and μ_2 at level $n - 1$	196
5.4. Two partitions λ, λ' at level n covering a single partition μ at level $n - 1$	196
5.5. Subgraph of the Young graph produces by considering two partitions λ, μ with $\lambda \searrow \mu$.197	
5.6. Two partitions $\nu_1, \nu_2 \vdash n - 1$ such that $\lambda \rightarrow \nu_i \rightarrow \mu$ for $i \in [2]$	198
5.7. High level view of a Ferrers diagram with canceling boxes and boxes of interest.	209
5.8. High level view of a Ferrers diagram with boxes of interest.	209
5.9. Ferrers diagram segmented by the boxes $X_i, i \in [m]$	213
5.10Graph of recursive skew hook elimination starting at the Ferrers diagram of $\lambda = (4, 4, 3)$, for the conjugation class $\zeta = (5, 4, 2)$	225
6.1. The union of two perfect matchings X_1 and X_2 of the graph $K_{5,5}$, edges of X_1 are in black and of X_2 in gray.....	227
6.2. The union of two perfect matchings of the graph $K_{5,5}$, edges of X_1 are in black and of X_2 in gray.	228

Notation

Throughout this dissertation we'll make extensive use of mathematical notation, some widely known and other that needs to be properly stated. In most cases we introduce notation once required and in opportune time, to avoid any confusion. Here we focus on the prerequisites and give some simplifications of well known notation.

We denote the set of natural number by $\mathbb{N}$ (we do not assume 0 is an element of $\mathbb{N}$), the set of integers by $\mathbb{Z}$, the set of real numbers by $\mathbb{R}$ and the set of complex number by $\mathbb{C}$. More generally we denote a field by $\mathbb{F}$.

To simplify notation we denote a set $\{1, 2, 3, \dots, n\} \subseteq \mathbb{N}$ by $[n]$. For the cardinality/size of a finite set S , we denote it by $|S|$.

We denote the symmetric group of order n by S_n [Rot95, chp. 1], thus an element $\sigma \in S_n$ is a permutation of $[n]$.^{*} Regarding permutations there are two useful types of notation, cycle notation and table notation, in table notation we think of $\sigma \in S_n$ as a function $\sigma : [n] \rightarrow [n]$ and denote it by

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & \cdots & n \\ \sigma(1) & \sigma(2) & \sigma(3) & \cdots & \sigma(n) \end{pmatrix}$$

For cycle notation, that we'll mostly use, we denote a permutation as a product of its disjoint cycles, like for instance $\sigma = (3\ 1\ 4)(2\ 5)$, for more on this refer to [Rot95, chp. 1].

When considering logarithms we'll only need two bases here, base 2 and base e , thus we'll only make use of the natural logarithm $\log_e$ whose simplified notation is $\ln$, and the logarithm $\log_2$ that we simplify to $\lg$.

Finally we define asymptotic notation, which is widely used in computer science and in particular in computational complexity theory. For this we may restrict the domain of functions over $\mathbb{R}$ to the set $\mathbb{N}$.

Given a function $f : \mathbb{N} \rightarrow \mathbb{R}$, in complexity theory we are generally not interested in multiplicative constant and lower order term of $f(n)$, or in other words we are interested in its behavior as n grows to infinity. Thus for instance if $f(n) = 7n^2 + 5n + 3$ then we'd really only be interested in the factor n^2 as this will largely determine the behavior of the function f as n grows large. Moreover we are interested in determining functions that either upper

^{*}This may be generalized to permutations on any finite set X , where then we denote the respective symmetric group by S_X .

bound or lower bound f . For this purpose we define three sets of functions:

$$\Omega(g) = \{f : \mathbb{N} \rightarrow \mathbb{R} \mid \exists_{n_0 \in \mathbb{N}, c > 0} \text{ such that for } n \geq n_0, cg(n) \leq f(n)\}$$

$$O(g) = \{f : \mathbb{N} \rightarrow \mathbb{R} \mid \exists_{n_0 \in \mathbb{N}, c > 0} \text{ such that for } n \geq n_0, f(n) \leq cg(n)\}$$

$$\Theta(g) = \{f : \mathbb{N} \rightarrow \mathbb{R} \mid \exists_{n_0 \in \mathbb{N}; c_1, c_2 > 0} \text{ such that for } n \geq n_0, c_1g(n) \leq f(n) \leq c_2g(n)\}$$

Using the example above one may easily verify that $f \in \Theta(n^2)$, which tells us that $f(n)$ is upper and lower bounded by n^2 (multiplied by appropriate constants), in particular it tells us that the behavior of $f(n)$ as n grows large may be approximated by n^2 . Extending this we have the sets $\Omega(g)$ and $O(g)$ that pertain to lower bounds and upper bounds, respectively. This is of course useful when we are talking about some function f whose behavior as n grows large we have yet to determine, and it's in this context that asymptotic notation is used in complexity theory, thus we seek to prove statements like $f \in O(g)$, for some function $g : \mathbb{N} \rightarrow \mathbb{R}$, thus determining that, asymptotically, $f(n)$ may be upper bounded by $cg(n)$ for some positive constant $c \in \mathbb{R}$.

In complexity theory it's customary to write $f = O(g)$ instead of $f \in O(g)$ (or with any of the other sets above), thus we adopt this custom. For more on asymptotic notation in computer science refer to [Cor+09, sec. 3.1].

Regarding citations/references to the bibliography we write, for instance, [Gre75, 2.24] to refer to a result (theorem, proposition, corollary, *etc.*) or an equation numbered 2.24 in reference [Gre75] in the bibliography, and write [Gre75, chp. 3] or [Gre75, sec. 3.2] to refer to a chapter or a section there in, repetitively. If we simply write [Gre75] then we are referring to the whole reference and nothing in it specifically.

1. Introduction

Communication complexity is a relatively new field in complexity theory, focusing on a particular computation model where one assumes a function is to be computed by two or more parties, by cooperation. The problem emerges from the assumption that each party in the model has only part of the input to the function and thus they have to communicate with each other in order to evaluate the function at that input. The complexity of the problem here is then how much information has to be shared amongst the parties to produce the function's output.

Computing the communication complexity of a function is generally no trivial matter, requiring extensive effort, yet several techniques for lower bounding communication complexity have been found and are extensively used in one way or another. One such technique uses the log of the rank of a matrix associated with the function in order to lower bound its communication complexity, a quite powerful technique, yet questions emerged of how good it really was, that is how tight was this lower bound really. This problem became known as the "*log-rank conjecture*", the conjecture being that the log of the rank of the aforementioned matrix was a tight lower bound for the communication complexity of the function.

Initial result only managed to prove that a constant gap between the communication complexity and the log of the rank exists [Raz92], but in 1995 *R. Raz* and *B. Spieker* proved that in fact a non-constant gap exists, in their paper "On the "log rank"-conjecture in communication complexity" [RS95], this is the result we revisit here in full detail. Although largely forgotten due to another paper by *N. Nisan* and *A. Wigderson* [NW94] being published the following year, that proved a stronger result than the one found by Raz and Spieker, renewed interest in it emerged due to the technique there used to find the rank of the matrix of the function.

Unfortunately the paper is notoriously difficult to follow, proofs lack context and intuition, and a strong background in representation theory of finite groups is required.

The paper starts by defining the function to be used, it may then be divided into two parts. In the first part the rank of the matrix of the function is computed, using representation theory of finite groups. In the second part a lower bound for the communication complexity of the function is determined, using an information theoretic argument.

Here we'll cover everything needed to understand the paper, we start with a short exposition of communication complexity and proceed to the second part of the Raz and Spieker paper, any complementary information needed is diverted to the appendices. We choose to start with the second part as it requires less background to be understood, moreover much of this background is known to computer scientists. The second part is also not the part we're most interested in, that will be the first part as its use of representation theory opens a door for future applications of this theory to communication complexity and more generally to complexity theory. Yet to understand the application given there we need to build a strong background in representation theory of finite groups, as such we give a proper introductory exposition of this theory. We strive to do this in the simplest of ways, making the theory intuitive, but also building it from an algebraic perspective so that we may extend it to more modern results.

As we'll see the function used in the paper may be described using permutations, and for this symmetric groups end up playing an important role. In particular the authors of the paper are able to relate the matrix of the function to the representation theory of symmetric groups.

It was a fortunate moment when the supervisor of this dissertation, Bruno Loff, noticed that the representation theory of symmetric groups, as was known at the time by Raz and Spieker, suffered from significant deficiencies, and luckily for us a new approach to this theory was recently developed by *A.M. Vershik* and *A.Yu. Okounkov* in "A New Approach to the Representation Theory of the Symmetric Groups" [AA05], that managed to resolve the issues of the past. Unfortunately, again, we saw ourselves looking at yet another paper that was very hard to digest, requiring a strong background in module theory. Moreover not a lot of resources exist that attempt to explain this new approach in an accessible way, some exist but not without their problems, thus a new requirement emerged, to produce a document explaining the results in the Vershik and Okounkov paper, in full detail.

This document would fulfill two objectives, first it would give a complete treatment of the representation theory needed to understand the first part of the paper by Raz and Spieker, and second it would give a detailed exposition of the Vershik and Okounkov paper.

This way newcomers to the Raz and Spieker paper would be provided with everything needed, from a modern perspective, to understand the paper and work on ways to extend its application of representation theory to other problems in complexity theory.

With the extensive work done in our exposition of representation theory the first part of the paper by Raz and Spieker follows straightforwardly.

2. Communication Complexity Theory

The concept of communication is as old as civilization itself, yet the field of Communication Complexity focuses on something less general but still as fundamental, the focal idea is cooperation.

In the general idea of communications we include the case where two or more parties (be it computers, components, systems or humans) are interested in purely sharing information, taking it to a more computational context we have as examples the case where we want to download a file off the internet, or send an email. These cases are heavily studied in the field of information theory, yet in Communication Complexity the communication arises from the necessity of two or more parties to jointly solve a problem, be it from a lack of computation resources by any single party or a lack of information, here we'll assume the latter is the case. Thus here, unlike in information theory, we don't start with a type of message or channel for communications to occur but with a problem to be solved, like for instance having two parts holding a bit string and jointly determine if they're equal. So the big problem here is that each part holds only part of the information and doesn't know what the others hold, thus have to communicate in order to solve the problem.

The exposition of Communication Complexity theory given here is based on the one given in [KN96], where we refer the reader to for a more complete treatment of the subject.

2.1. Model and complexity

The question is, for a given problem, how much communication has to occur for it to be solved. We'll focus on the case where there are only two parties, in contrast to multiparty communication.

The simplest message both parties can communicate is a single bit, a single yes/no, and so binary digits are the fundamental information we'll be concerned with and allow both parties to extensively use. Any other type of communication considered can be brought down to the binary level [AB09, sec. 0.1], thus we are in no way limiting ourselves. That being said our measure of how complex/complicated a problem is can be expressed as a function of how many bits have to be communicated to solve it. Yet we'll be concerned with the minimum amount of bits required, after all we can always complicate things by

having useless bits communicated. This complexity measure is our communication complexity and is related only to a problem and not any particular solution of it.

We'll care only about the communication and won't account for any computational efforts made by any single part of the system, thus we'll look only at what happens between the two parties and not how costly it is for a single party to figure out what it's going to communicate to the other. In other words, we give each party infinite computational power.

2.1.1 Yao's model

In 1979 Andrew Yao proposed the simplest mathematical model to capture the scenario of two party communication mentioned above [Yao79], it follows from a simple set of assumptions.

- There are only two parties in the system.
- Each party of the system gets a fixed part of the input information for the problem.
- The only resource we care about is communication.
- The task is the computation of some predetermined function of the input (the problem).

These assumption help us concentrate on the core issue of communication, yet it produces a very rich and flexible model.

We proceed now to formalize our model from this set of assumption. Let X, Y, Z be arbitrary sets and $f : X \times Y \rightarrow Z$ an arbitrary function, we say $X \times Y$ is the *domain* and Z the *range* of f . By our previous remarks we may always assume that the set Z is of the form $\{0, 1\}^n$ for some $n \in \mathbb{N}$, and as we'll see this will allow us to focus on simpler cases. We'll name each party of the system after the two most well know persons in computer science, Alice and Bob. Alice gets an element $x \in X$ and Bob an element $y \in Y$, their objective is to jointly compute $f(x, y)$.

The communication between them are carried out by a *protocol* $\mathcal{P}$ which consists of the parties/players (Alice and Bob) sending bits to one another until the value of $f(x, y)$ is determined. At each stage the protocol $\mathcal{P}$ defines who has to communicate, what the

player communicates, who communicates next and when the protocol execution terminates. This information depends solely on (x, y) , the input, and what has been communicated so far during the run of the protocol, as this is the only knowledge common to both players. As a player only communicates bits at each stage we have two possibilities, which hints to a binary tree structure.

We now begin to properly define all ideas gathered so far.

Definition 2.1.1. *A protocol $\mathcal{P}$ with domain $X \times Y$ and range Z is a finite binary tree (directed acyclical graph) whose internal nodes are labeled with functions from a set $A \cup B$ where A is a set of functions of the form $a_i : \{\alpha\} \times X \rightarrow \{0, 1\}$ where α is a tuple containing a record of the communication that occurred previously, and B contains function of the form $b_i : \{\alpha\} \times Y \rightarrow \{0, 1\}$ where α is as before. We generally omit $\{\alpha\}$ from the input of the functions in $A \cup B$ for convenience, thus $A = \{a_i : X \rightarrow \{0, 1\}\}$, and $B = \{b_i : Y \rightarrow \{0, 1\}\}$. The leaves of the tree are labeled with the elements of Z .*

Applying $\mathcal{P}$ to an input $(x, y) \in X \times Y$ produces a path in the binary tree, from root to a leaf, determined by going left if the result of the function on a node is 0 and right if it's 1, i.e. if $a_v(x) = 0$ or $b_v(y) = 0$ (depending on what function we have) we go left, and we go right if the result of that function is 1. The result of applying $\mathcal{P}$ to input (x, y) is the label of the leaf the path terminates in. We say $\mathcal{P}(x, y) = z$ to indicate that the application of $\mathcal{P}$ to (x, y) results in $z \in Z$.

Intuitively the functions in the set A are attributed to communications done by Alice, and the ones in B by Bob. Bellow in figure 2.1 we give an example of a protocol, we remark that although in this example communications by Alice and Bob are done in turns this does not always have to be the case, it may well be that in some cases (branches in the graph of a protocol) Alice or Bob communicate twice or more times repeatedly.

Note that we defined a protocol without making an explicit connection to any given function, the following definition established the required connection between a protocol $\mathcal{P}$ and a problem f .

Definition 2.1.2. *Given a protocol $\mathcal{P}$ with domain $X \times Y$ and range Z , and a function $f : X \times Y \rightarrow Z$, we say that $\mathcal{P}$ computes f if for all $(x, y) \in X \times Y$, $\mathcal{P}(x, y) = f(x, y)$.*

So by the above definition, given a function f , a protocol $\mathcal{P}$ that computes f is essentially a solution to the problem of having two parties computing f by cooperating with each other. As we're interested in a solution that communicates the least amount of bits we need a metric to compare protocols, in other words we need to determine the "cost" of a protocol.

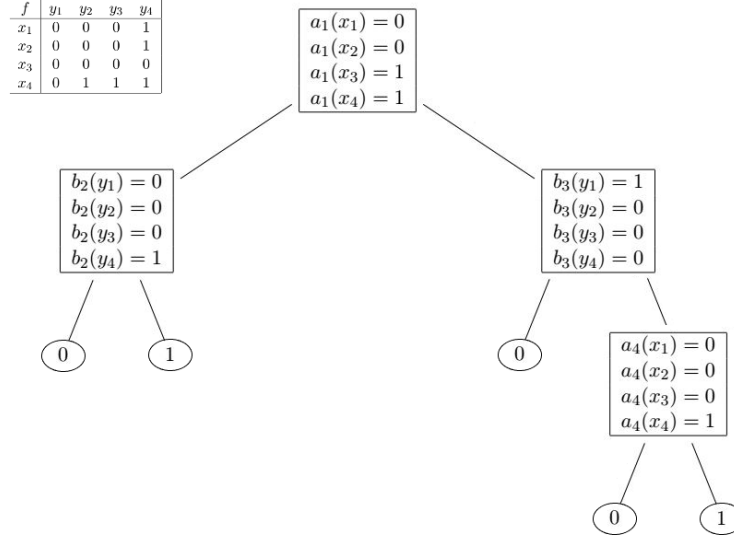


FIGURE 2.1: Example of protocol for a binary function $f : X \times Y \rightarrow \{0, 1\}$ given in the upper left corner.

Definition 2.1.3. Given a protocol $\mathcal{P}$ we define the cost of $\mathcal{P}$ as the height of its binary tree, i.e. the maximum length of a path in the tree from root to leaf (A.0.3).

This definition makes sense as it gives us the maximum number of bits communicated by the protocol in any given execution. With this we can now define the communication complexity of a function/problem.

Definition 2.1.4. Given a function $f : X \times Y \rightarrow Z$ we define its communication complexity as the minimum cost over all protocols $\mathcal{P}$ that compute f . We denote it by $D(f)$.

We remark that the above definition indeed makes sense as there must exist a minimum element in the set of costs of all protocols that compute f , as this set is in particular a subset of $\mathbb{N}$ due to the cost of a protocol being the height of a binary tree, which is a whole number. Thus by this it is the case that $D(f) \in \mathbb{N}$.

The question now is how to make this tractable. Although we have the definition of communication complexity of a function we can't possibly go over all protocol that compute it to determine what the optimal one for it is, at least not generally, thus we need to further explore the properties of a protocol that computes the function in order to gain further insight into the problem at hand.

2.1.2 The fundamental theorem of C.C.

As we'll see the fact that each player knows only their respective input can be of use to us, moreover because each node in the tree is labeled by a function, as seen in definition

2.1.1, and the path traversed by applying $\mathcal{P}$ to (x, y) is a result of the output of those functions, an input results in a single path in the tree, and this allows us to partition the input space.

Definition 2.1.5. Let $\mathcal{P}$ be a protocol and v a node in it. We define R_v to be the set of inputs $(x, y) \in X \times Y$ whose application of $\mathcal{P}$ results in a path that includes node v .

Proposition 2.1.6. Let L be the set of leaves in a protocol $\mathcal{P}$, then $\{R_l\}_{l \in L}$ is a partition of $X \times Y$.

Proof. As stated above an input (x, y) results in a single path in $\mathcal{P}$, as the path is determined by a sequence of functions, thus the existence of $(x, y) \in R_{l_1} \cap R_{l_2}$ for some $l_1, l_2 \in L$ with $l_1 \neq l_2$ contradicts the definition of a function. $\square$

In figure 2.2 we give an example of the partition alluded above, using the protocol in figure 2.1, there we observe that the protocols partitions the input space into 5 parts.

f	y_1	y_2	y_3	y_4
x_1	0	0	0	1
x_2	0	0	0	1
x_3	0	0	0	0
x_4	0	1	1	1

FIGURE 2.2: Example of partitioning of the input space of the function f by the protocol given in figure 2.1.

So we now have that a protocol induces a partition in the input space of the respective function it computes, and that partition has as many parts as the protocol has leaves, which is a quite powerful result for if we know how many leaves the protocol has we can lower bound its cost using theorem A.0.7 in Appendix A, yet more can be said about the nature of the elements of the partition. In that regard we begin with a definition.

Definition 2.1.7. A combinatorial rectangle (or just rectangle) in $X \times Y$ is a set $R \subseteq X \times Y$ such that $R = A \times B$ with $A \subseteq X$ and $B \subseteq Y$.

Proposition 2.1.8. $R \subseteq X \times Y$ is a (combinatorial) rectangle if and only if

$$(x_1, y_1), (x_2, y_2) \in R \Rightarrow (x_1, y_2) \in R$$

Proof. This follows from the fact that $R = A \times B$ with $A \subseteq X$ and $B \subseteq Y$. $\square$

Given $(x_1, y_1), (x_2, y_2) \in R_l$, with $l \in L$, as both this inputs produce paths that terminate in the same leaf l of $\mathcal{P}$, by proposition A.0.6 in Appendix A it must be the case that they

produce the same path in $\mathcal{P}$. Thus Alice and Bob perform the same communications for (x_1, y_1) as for (x_2, y_2) , yet the functions in the set A in definition 2.1.1 depend only on X and the record of communication done so far, and the function in B depend only on Y and the record of communication done so far, which by what was remarked above are equal both for (x_1, y_1) and (x_2, y_2) , thus (x_1, y_2) and (x_2, y_1) result in the same path as (x_1, y_1) (and (x_2, y_2)), hence $(x_1, y_2), (x_2, y_1) \in R_l$. This gives us the following proposition.

Proposition 2.1.9. *For any protocol $\mathcal{P}$ and leaf l in it, R_l is a rectangle.*

We can additionally note that for any $l \in L$, f will be constant when restricted to R_l , for those inputs result in the same output of f (assuming of course that we're talking about a protocol that computes f). We can formalize this with the following definition.

Definition 2.1.10. *A set $R \subseteq X \times Y$ is said to be f -monochromatic for a function $f : X \times Y \rightarrow Z$, or just monochromatic when f is clear from context, if f is constant when restricted to R .*

Below in figure 2.3 we give an example of a monochromatic rectangle of a function $f : \{0, 1\}^3 \times \{0, 1\}^3 \rightarrow \{0, 1\}$ where f is constant and equal to 0, such a rectangle is said to be a 0-monochromatic rectangle, a similar idea goes for the definition of a 1-monochromatic rectangle. Note that a monochromatic rectangle need not be of a contiguous shape.

f	000	001	010	011	100	101	110	111
000	0	1	1	0	1	0	0	0
001	1	0	0	0	0	0	0	1
010	1	0	0	0	1	0	0	0
011	0	0	1	0	0	0	0	1
100	1	0	0	0	1	0	0	1
101	1	1	1	0	0	0	1	1
110	0	0	0	0	1	0	0	0
111	0	1	1	0	1	1	0	1

FIGURE 2.3: Example of a 0-monochromatic rectangle of a function $f : \{0, 1\}^3 \times \{0, 1\}^3 \rightarrow \{0, 1\}$.

We are now ready to state the theorem alluded to in the title of this section, that we fondly call *the fundamental theorem of Communication Complexity*.

Theorem 2.1.11. *Any protocol $\mathcal{P}$ that computes a given function $f : X \times Y \rightarrow Z$ induces a partition of $X \times Y$ into f -monochromatic (combinatorial) rectangles. The number of rectangles in the partition is equal to the number of leaves in the protocol.*

The above theorem produces the following remarkable result.

Theorem 2.1.12. *If any partition of the input space of a given function f into monochromatic rectangles requires at least k such rectangles then $D(f) \geq \lg(k)$.*

Proof. Given a protocol to compute $f : X \times Y \rightarrow Z$, by the previous corollary it induces a partition of $X \times Y$ into f -monochromatic rectangles, thus if we find a lower bound for the number of parts in such partition we find a lower bound for the number of leaves a protocol that computes f can have. Indeed by contradiction if no partition with less than k parts exists then no protocol with less than k leaves can exist.

Now let k be that lower bound, by theorem A.0.7 in Appendix A a binary tree with k leaves has height at least $\lg(k)$, thus by definition 2.1.3 a protocol that computes f must have cost at least $\lg(k)$. The conclusion follows immediately from definition 2.1.4 of $D(f)$. $\square$

In the following section we'll present several techniques for lower bounding $D(f)$, using the above theorem.

One may easily upper bound $D(f)$ by noting that for any $f : X \times Y \rightarrow Z$, Alice may simply send her entire input to Bob and let him compute $f(x, y)$ and communicate that answer back, or vice versa as both are assumed to have infinite computing power, and this constitutes a perfectly valid protocol for f , thus we have the following.

Proposition 2.1.13. *For any $f : X \times Y \rightarrow Z$*

$$D(f) \leq \lg(\min(|X|, |Y|)) + \lg(|Z|)$$

From now on we'll focus on the case where $Z = \{0, 1\}$, functions with this range are usually called *predicates*. We note that the problem of computing a function f with a range Z' other than $\{0, 1\}$ may be reduced to the problem of computing predicates by considering the functions $f_i(x, y)$, for $i \in \lg(|Z'|)$, that compute the i -th bit of $f(x, y)$.

2.2. Lower bounds

The fundamental property we exploit here is the partition into monochromatic rectangles a protocol induces on the input space of a function, seen in theorem 2.1.11, via theorem 2.1.12.

Our first lower bounding technique seeks to find a subset of the input space of a function f for which no two elements can be in the same monochromatic rectangle, hence there must

be at least as many monochromatic rectangles partitioning the input space as elements in this subset, or in other words we get a lower bound for the number of such rectangles in the partition induced by a protocol. We call this subset a *fooling set* for f . As can be expected we'll make use of definition 2.1.10 of an f -monochromatic rectangle.

Definition 2.2.1. Consider $f : X \times Y \rightarrow \{0, 1\}$. A set $S \subseteq X \times Y$ is said to be a *fooling set* for f if there exists $z \in \{0, 1\}$ such that

- a) For all $(x, y) \in S$, $f(x, y) = z$. That is f is constant and equal to z on S .
- b) Given $(x_1, y_1), (x_2, y_2) \in S$ either $f(x_1, y_2) \neq z$ or $f(x_2, y_1) \neq z$, that is (x_1, y_1) and (x_2, y_2) can't be in the same f -monochromatic rectangle.

Using the notation above, we see that property b) of a fooling set for f is built to precisely contradict the definition of a monochromatic rectangle by contradicting the alternative definition of a rectangle in proposition 2.1.8 and the fact that f must be constant on a monochromatic rectangle, so that no two elements of S can be in the same monochromatic rectangle, as intended. Thus if we find such set for f we can directly use theorem 2.1.12 to get a lower bound for $D(f)$, as seen below.

Theorem 2.2.2. If a function $f : X \times Y \rightarrow \{0, 1\}$ has a fooling set S then

$$D(f) \geq \lg(|S|)$$

Proof. Indeed by the observations above a protocol for f must induce a partition on $X \times Y$ with at least $|S|$ monochromatic rectangles, thus the conclusion follows from theorem 2.1.12. □

Note that as the function f is constant on the fooling set S , say equal to $z \in \{0, 1\}$, each element in S is in a monochromatic rectangle where f is equal to z , thus the lower bound we get is on the number of monochromatic rectangles where f is equal to z . If for instance $z = 1$ then we get a lower bound for the number of monochromatic rectangles where f is equal to 1, also called the 1-monochromatic rectangles. To improve the lower bound we can seek a fooling set where $z = 0$ and get a lower bound for the 0-monochromatic rectangles. The sum of these two lower bounds, that is $|S_0| + |S_1|$ where S_0 is the fooling set when $z = 0$ and S_1 when $z = 1$, result in a tighter lower bound for $D(f)$ as $\lg(|S_0| + |S_1|)$.

The question now is how does one in practice obtain a fooling set. The question of course very much depends on the function f , as such we next present an example.

Example 2.2.3. Let $x, y \in \{0, 1\}^n$, Alice holds x and Bob holds y . The objective for both players is to determine if these bit strings are equal, we call this function $\text{EQ} : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}$, where $\text{EQ}(x, y) = 1$ if $x = y$, otherwise $\text{EQ}(x, y) = 0$.

If we consider the set

$$S = \{(\alpha, \alpha) | \alpha \in \{0, 1\}^n\}$$

then it is clear that EQ is constant and equal to 1 on S , moreover given $\alpha_1, \alpha_2 \in S$ if $\alpha_1 \neq \alpha_2$ then $\text{EQ}(\alpha_1, \alpha_2) = 0$, thus S is a fooling set for EQ . As $|S| = 2^n$ we have that

$$D(\text{EQ}) \geq \lg(2^n) = n$$

As any protocol for EQ partitions the set $\{0, 1\}^n \times \{0, 1\}^n$ into monochromatic rectangles and there are pairs (x, y) such that $\text{EQ}(x, y) = 0$ there must exist 0-monochromatic rectangles, thus it must be the case that

$$D(\text{EQ}) > n$$

Now using proposition 2.1.13 we have that

$$D(f) \leq \lg(2^n) + \lg(2) = n + 1$$

Therefore, recalling that it is always the case that $D(f) \in \mathbb{N}$ for any f , we can conclude that $D(f) = n + 1$.

Another very useful technique seeks to upper bound the size of the monochromatic rectangles for a function $f : X \times Y \rightarrow Z$. If we manage to determine that the monochromatic rectangles for f have size at most k then a partition of $X \times Y$ into monochromatic rectangles must have at least $|X||Y|/k$ such parts, and so by theorem 2.1.12

$$D(f) \geq \lg\left(\frac{|X||Y|}{k}\right)$$

Hence we have the following theorem.

Theorem 2.2.4. *Given a function $f : X \times Y \rightarrow Z$, if any monochromatic rectangle for f has size at most k then*

$$D(f) \geq \lg\left(\frac{|X||Y|}{k}\right)$$

The last lower bounding technique we'll present is also perhaps the most interesting, as it uses concepts of linear algebra, thus allowing for the use of more advanced tools in

determining the communication complexity of a problem.

Definition 2.2.5. Given a function $f : X \times Y \rightarrow Z$ we define M_f as the matrix

$$M_f = (f(x, y))_{x, y}$$

Below in figure 2.4 we give the matrix M_{EQ} , of the equality function $\text{EQ} : \{0, 1\}^3 \times \{0, 1\}^3 \rightarrow \{0, 1\}$ seen in example 2.2.3.

	000	001	010	011	100	101	110	111
000	1	0	0	0	0	0	0	0
001	0	1	0	0	0	0	0	0
010	0	0	1	0	0	0	0	0
011	0	0	0	1	0	0	0	0
100	0	0	0	0	1	0	0	0
101	0	0	0	0	0	1	0	0
110	0	0	0	0	0	0	1	0
111	0	0	0	0	0	0	0	1

FIGURE 2.4: Matrix M_{EQ} of the equality function $\text{EQ} : \{0, 1\}^3 \times \{0, 1\}^3 \rightarrow \{0, 1\}$.

Recall the definition of rank of a matrix [Gre75, 2.34; 3.4], we denote the rank of a matrix M by $\text{rank}(M)$.

Theorem 2.2.6. Given a function $f : X \times Y \rightarrow \{0, 1\}$ with corresponding matrix M_f as in the previous definition

$$D(f) \geq \lg(\text{rank}(M_f))$$

Proof. Let $\mathcal{P}$ be a protocol for f and let L_1 be the set of leaves of $\mathcal{P}$ labeled with 1. For each $l \in L_1$ let R_l be the set of inputs in $X \times Y$ whose path in $\mathcal{P}$ terminate in l and define the matrix M_l , with rows indexed by X and columns indexed by Y , where an entry (x, y) is equal to 1 if $(x, y) \in R_l$ and 0 otherwise.

We note that for (x, y) such that $f(x, y) = 0$ we have that in every matrix M_l , $l \in L_1$, the entry (x, y) is 0, whereas if $f(x, y) = 1$ with $(x, y) \in R_l$ only the matrix M_l will have a 1 in position (x, y) , as $\{R_l\}_{l \in L_1}$ is a partition of $X \times Y$, therefore

$$M_f = \sum_{l \in L_1} M_l$$

Now it is a simple property of the rank of a matrix, and one may easily verify, that

$$\text{rank} \left(\sum_{l \in L_1} M_l \right) \leq \sum_{l \in L_1} \text{rank}(M_l)$$

Thus

$$\text{rank}(M_f) \leq \sum_{l \in L_1} \text{rank}(M_l)$$

Now note that as R_l is a rectangle, as in definition 2.1.7, we have that for each $l \in L_1$, $\text{rank}(M_l) = 1$. Therefore, letting L be the set of all leaves of $\mathcal{P}$, we can conclude that

$$\text{rank}(M_f) \leq |L_1| \leq |L|$$

Therefore we have that $\lg(\text{rank}(M_f)) \leq \lg(|L|)$ which as we know is less than the height of $\mathcal{P}$. As $\mathcal{P}$ is any protocol for f we have that

$$D(f) \geq \lg(\text{rank}(M_f))$$

□

Similar to the fooling set technique we note that the above $\lg(\text{rank}(M_f))$ is in fact a lower bound for the number of 1-monochromatic rectangles for f . The result may then be improved by lower bounding the number of 0-monochromatic rectangles, and this can be easily done by applying the above theorem to the function $\text{not}(f) = 1 - f$. As $M_{\text{not}(f)} = J - M_f$, where J is the matrix with 1 everywhere, whose rank is 1, we have that

$$\text{rank}(M_{\text{not}(f)}) \leq 1 + \text{rank}(M_f) \Leftrightarrow \text{rank}(M_{\text{not}(f)}) - 1 \leq \text{rank}(M_f)$$

Moreover as then $M_f = J - M_{\text{not}(f)}$ we have also that $\text{rank}(M_f) \leq \text{rank}(M_{\text{not}(f)}) + 1$, hence the rank of M_f and $M_{\text{not}(f)}$ differ by at most 1, therefore

$$D(f) \geq \lg(2 \text{rank}(M_f) - 1)$$

2.3. Notes on nondeterminism

As with most computation models, there exists a concept of nondeterminism* for the communication protocols presented in the previous sections. In this regard the theory developed so far may be defined as *deterministic* communication complexity.

The protocols as we defined them specify a unique behavior for Alice and Bob, in which at each stage of the protocol the next step is uniquely determined, that is who communicates next and what is communicated is fully and uniquely specified by the protocol and the

*We advise the interested reader to refer to [Sip12, sec. 1.2, 3.2] for more on nondeterminism.

input under consideration. Yet in nondeterminism we extend this to allow for different possibilities to be realized, simultaneously. So when in a node in the protocol it is possible to proceed in more than one direction, and what this effectively means is that the functions in the set $A \cup B$ in definition 2.1.1 are of the form

$$a_i : X \rightarrow \{0, 1, \{0, 1\}\} ; \quad b_i : Y \rightarrow \{0, 1, \{0, 1\}\}$$

Thus we may proceed through the left outgoing edge of a node or through the right one as defined in 2.1.1, or we may proceed through both edges simultaneously, we call this a *nondeterministic step*.

Definition 2.3.1. A nondeterministic protocol $\mathcal{P}$ with domain $X \times Y$ and range Z is defined as in definition 2.1.1 with the distinction that functions in the set $A \cup B$ are of the form

$$a_i : X \rightarrow \{0, 1, \{0, 1\}\} ; \quad b_i : Y \rightarrow \{0, 1, \{0, 1\}\}$$

Applying $\mathcal{P}$ to an input $(x, y) \in X \times Y$ may produce several paths in the binary tree, from root to several leaves, determined by going left if the result of a function on a node is 0, right if it's 1 or through both edges if the result is $\{0, 1\}$. The result of applying $\mathcal{P}$ to an input $(x, y) \in X \times Y$ is a set containing the labels of the leaves the paths terminated on.

The cost of a nondeterministic protocol $\mathcal{P}$ is the height of its binary tree.

Of course such protocols bear consequences on what computing a function means.

Definition 2.3.2. We say a nondeterministic protocol $\mathcal{P}$ with domain $X \times Y$ and range $Z = \{0, 1\}$ computes a function $f : X \times Y \rightarrow Z$ if on input (x, y) the following conditions hold

- If $f(x, y) = 0$ then every path resulting from the application of the protocol $\mathcal{P}$ to (x, y) terminate in a leaf with label 0, thus $\mathcal{P}(x, y) = \{0\}$.
- If $f(x, y) = 1$ then at least one path resulting from the application of the protocol $\mathcal{P}$ to (x, y) terminate in a leaf with label 1, thus $1 \in \mathcal{P}(x, y)$.

One may easily verify that the sets R_l of inputs that terminate in a leaf $l \in L$, where L is the set of leaves of $\mathcal{P}$, are still rectangles as in definition 2.1.7, thus proposition 2.1.9 is still valid. Moreover if l is a leaf labeled with a 1 then R_l is a 1-monochromatic rectangle, however the same can't be said if l is labeled with a 0, yet as we'll see this won't be a problem.

Note also that as an input may produce several different paths, thus terminate in different leafs, an input may belong to more than one rectangle, hence we no longer get a partition of the input space, we get a *cover*.

Definition 2.3.3. A cover of a set A is a family of sets $\{B_i\}_{i \in [k]}$, with $B_i \subseteq A$ and $k \in \mathbb{N}$, such that

$$\bigcup_{i \in [k]} B_i = A$$

with the possibility that for $i, j \in [k]$, $B_i \cap B_j \neq \emptyset$.

Below in figure 2.5 we have an example of a cover, there we see two 0-monochromatic rectangles and two overlapping 1-monochromatic rectangles.

f	y_1	y_2	y_3
x_1	1	1	0
x_2	1	1	1
x_3	0	1	1

FIGURE 2.5: Example of a covering of the input space of a function $f : X \times Y \rightarrow \{0, 1\}$, into two 0-monochromatic rectangles and two overlapping 1-monochromatic rectangles.

Thus a cover is similar to a partition but with the distinction that parts may now intersect with each other. By our remarks above we have the following result.

Proposition 2.3.4. A nondeterministic protocol $\mathcal{P}$ that computes $f : X \times Y \rightarrow Z$ induces a cover of $\{(x, y) \in X \times Y \mid f(x, y) = 1\}$ into 1-monochromatic rectangles. Moreover the number of rectangles in the cover is equal to the number of leaves of $\mathcal{P}$ with label 1.

We call the set $\{(x, y) \in X \times Y \mid f(x, y) = 1\}$ the 1-inputs of f , one may similarly define the 0-inputs of f .

While a deterministic protocol induces a partition of the input space into monochromatic rectangles, the inverse is not the case, to see this consider the following example of a partition of the input space of a function $f : X \times Y \rightarrow \{0, 1\}$, in figure 2.6.

f	y_1	y_2	y_3
x_1	1	0	0
x_2	1	1	1
x_3	0	0	1

FIGURE 2.6: Partition of the input space of a function $f : X \times Y \rightarrow \{0, 1\}$ into five monochromatic rectangles.

Assuming that there exists a protocol $\mathcal{P}$ that induces the partition, as the partition has 5 parts the protocol has 5 leaves, thus there exists a first node, corresponding to the communications of a player, say Alice, for which the function a_i at that node is not constant, that is it has two outgoing edges. As by this the answer that Alice sends on x_1 , x_2 and x_3 can all be the same there are two possibilities:

- Her answer on x_1 and x_2 are different, that is $a_i(x_1) \neq a_i(x_2)$, thus $\{x_1, x_2\} \times \{y_1\}$ can't be a monochromatic rectangle as (x_1, y_1) and (x_2, y_1) don't produce paths on $\mathcal{P}$ that terminate in the same leaf.
- Her answer on x_2 and x_3 are different, that is $a_i(x_2) \neq a_i(x_3)$, thus $\{x_2, x_3\} \times \{y_1\}$ can't be a monochromatic rectangle by the same argument as above.

Hence the partition can't correspond to the protocol. Yet for nondeterministic protocols a cover of the 1-inputs into 1-monochromatic rectangles induces a protocol.

Proposition 2.3.5. *Given $f : X \times Y \rightarrow Z$ and a cover of the 1-inputs of f into 1-monochromatic rectangles, there exists a nondeterministic protocol computing f that induces that cover.*

Proof. Indeed given the cover and an input (x, y) , Alice may simply communicate the name of the several 1-monochromatic rectangles that contain inputs with x as their first component*, taking nondeterministic steps to do this, to which Bob in each path replies with a 0 if y is not in the mentioned rectangle, or if it is he replies 1.

It is clear this protocol computes f and induces the given cover. □

One notes that the protocol defined above has cost equal to the logarithm (base 2) of the number of 1-monochromatic rectangles in the given cover of the 1-inputs of f , and as any such cover induces a valid nondeterministic protocol computing f , and vice versa, we can obtain the protocol with the minimum cost over all nondeterministic protocol that compute f by considering the smallest possible cover of the 1-inputs of f .

Definition 2.3.6. *Given a function $f : X \times Y \rightarrow Z$ we denote by $C^1(f)$ the smallest possible cover of the 1-inputs of f into monochromatic rectangles.*

Definition 2.3.7. *The nondeterministic communication complexity of a function f , denoted $N^1(f)$, is equal to $\lg(|C^1(f)|)$.*

*We note that said names are simply binary numbers, that is if the cover has k rectangles then each name may be represented with $\lg(k)$ bits.

Now we may ask what is the advantage of considering nondeterministic protocols. As any partition of $X \times Y$ is in particular a cover it must be the case that a cover of the input space has at most as many elements as a partition. Coupling this with the fact that we need only to consider covers of the 1-inputs and using theorem 2.1.12 it becomes clear that the nondeterministic communication complexity of a function can be much smaller than the deterministic communication complexity.

Regarding lower bounding techniques for nondeterministic communication complexity we remark that both theorem 2.2.2 and 2.2.4 are still applicable here. Indeed as the fooling set guarantees that no two elements in it can be in the same rectangle the technique applies verbatim, for theorem 2.2.4 as we target the size of the rectangles a cover of the 1-inputs has at least as many rectangles as a partition of the 1-inputs, due to the possibility of intersections between them.

2.4. The log-rank conjecture

As we've seen in the previous section our techniques for lower bounding communication complexity are both applicable for deterministic and nondeterministic complexity, which raises the question of how good are the lower bounds obtained through these methods, the exception being the log (base 2) of the rank of the matrix M_f as seen in theorem 2.2.6, this was originally proven by *K. Mehlhorn* and *M. Schmidt* in [MS82] and became one of the most important methods with which to obtain lower bounds for deterministic communication complexity. Yet again one must question how good can this lower bound be, and thus it became an open problem whether the lower bound obtained was tight, in the sense that it would be asymptotically equal to the exact complexity, or more precisely, given a function f

$$D(f) = \Theta(\lg(\text{rank}(M_f)))$$

If so one would need only to compute the rank of the matrix M_f to get the deterministic communication complexity of a function f . This open problem came to be known as *the log-rank conjecture*.

The paper by Raz and Spieker revisited here managed to prove that in fact the log-rank conjecture has a false answer, by proving that there exists a non-constant gap between the communication complexity of a particular predicate f and the log of the rank of M_f .

To define the predicate f they used we first need to look at some definitions. Recall the definition of graph, we say a graph Γ is a tuple (V, E) where V is a set of vertices and E is a set of edges, members of this set are of the form $\{v, v'\}$ for $v, v' \in V$ with $v \neq v'$. We use the set notation for edges instead of tuple notation as edges in a graph have no direction (contrary to a directed graph), thus notionally $(v, v') = (v', v) = \{v', v\}$. Note that under this notation an edges of the form (v, v) would be simply denoted as $\{v\}$, these types of edges are known as *loops*, yet we will not consider graphs that contain them here.

Definition 2.4.1. A complete bipartite graph $K_{n,n}$ is a graph with $2n$ vertices, where we consider them as divided into two disjoint sets of n vertices each, V_1 and V_2 . With this in mind the graph has every vertex in the set V_1 connected by an edge to every vertex in the set V_2 , thus a vertex in V_1 is connected by an edge to every vertex in V_2 .*, therefore we have that $\Gamma = (V, E)$ with $V = V_1 \cup V_2$ and

$$E = \{\{v, v'\} \mid v \in V_1, v' \in V_2\}$$

Below in figure 2.7 we give the graph $K_{3,3}$.

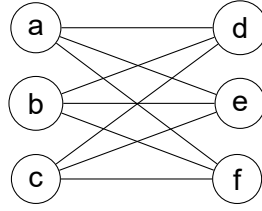


FIGURE 2.7: Complete bipartite graph $K_{3,3}$.

Definition 2.4.2. Given a graph $\Gamma = (V, E)$ a path in it is a sequence of vertices and edges

$$v_1 \rightarrow e_1 \rightarrow v_2 \rightarrow e_2 \rightarrow v_3 \rightarrow \cdots \rightarrow e_{n-1} \rightarrow v_n$$

where $e_i \in E$ is an edge and $v_i \in V$ a vertex, such that the edge e_i has as endpoints the vertices v_i, v_{i+1} , that is $e_i = \{v_i, v_{i+1}\}$, and such that it has no repeating vertices in the sequence. A cycle on the other hand is a closed path, that is the final vertex is equal to the first, so that only the first and last vertices in the sequence repeat.

A Hamiltonian cycle is a cycle that contains each vertex of Γ exactly once (with the exception of the last in the sequence which is then equal to the first).

As examples we may consider in the graph $K_{3,3}$ (figure 2.7 above) the path

$$a \rightarrow \{a, d\} \rightarrow d \rightarrow \{d, b\} \rightarrow b \rightarrow \{b, f\} \rightarrow f$$

*Or the other way around as the graph is not a directed graph.

or the cycle

$$a \rightarrow \{a, d\} \rightarrow d \rightarrow \{d, b\} \rightarrow b \rightarrow \{b, f\} \rightarrow f \rightarrow \{f, a\} \rightarrow a$$

We may also consider the Hamiltonian cycle

$$a \rightarrow \{a, d\} \rightarrow d \rightarrow \{d, b\} \rightarrow b \rightarrow \{b, e\} \rightarrow e \rightarrow \{e, c\} \rightarrow c \rightarrow \{c, f\} \rightarrow f \rightarrow \{f, a\} \rightarrow a$$

For what follows we say an edge e is *incident* on a vertex v if $e = \{v, v'\}$ for some vertex v' .

Definition 2.4.3. A matching on a graph $\Gamma = (V, E)$ is a subset of its edges $X \subseteq E$ such that no two are incident on a same vertex, that is if $e_1 = \{v_1, v'_1\}, e_2 = \{v_2, v'_2\} \in X$ then $v_1 \neq v_2, v_1 \neq v'_2, v'_1 \neq v_2$ and $v'_1 \neq v'_2$.

A perfect matching X is a matching such that given any vertex v in V there exists an edge $e \in X$ incident on v .

When looking at a complete bipartite graph $K_{n,n} = (V_1 \cup V_2, E)$, as in definition 2.4.1, one notes that a perfect matching X on it induces a bijection $\theta : V_1 \rightarrow V_2$, by simply letting for $v \in V_1$, $\theta(v) = v'$ such that $\{v, v'\} \in X$. This is well defined as it can't be the case that $\{v'', v'\} \in X$ for some vertex $v'' \neq v$, by definition of matching. For illustrative purposes an example of a perfect matching on $K_{3,3}$ is given below in figure 2.8, if we let $V_1 = \{a, b, c\}$ and $V_2 = \{d, e, f\}$ then

$$\theta(a) = f; \quad \theta(b) = d; \quad \theta(c) = e$$

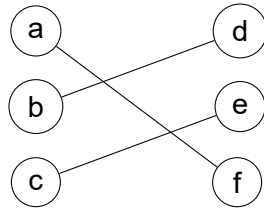


FIGURE 2.8: A perfect matching on the complete bipartite graph $K_{3,3}$.

Consider now a complete bipartite graph $K_{n,n}$, for some $n \in \mathbb{N}$, and let $\mathcal{M}$ be the set of all perfect matchings on $K_{n,n}$. Given two perfect matchings $X_1, X_2 \in \mathcal{M}$ we may consider their union, indeed if $K_{n,n} = (V, E)$ then as $X_i \subseteq E$ for $i \in [2]$ then $\Gamma = (V, X_1 \cup X_2)$ is a graph (yet be mindful that it may occur that we get multiple edges between two vertices in V). Below in figure 2.9 we give an example of such a graph, we mark the edges from X_1 with an x and the ones from X_2 with a y .

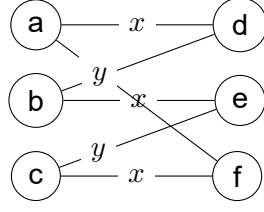


FIGURE 2.9: The union of two perfect matchings of the graph $K_{3,3}$.

The predicate considered by Raz and Spieker in their paper asks if this union of perfect matchings produces a Hamiltonian cycle in $K_{n,n}$. In the example above we may rearrange the vertices as seen in figure 2.10, thus making it clear that in this case we do get a Hamiltonian cycle.

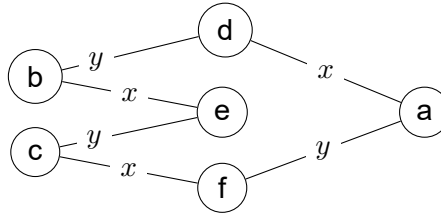


FIGURE 2.10: The union of two perfect matchings of the graph $K_{3,3}$, rearranged.

Hence the predicate considered may be defined as

$$\begin{aligned} \mathfrak{F} : \mathcal{M} \times \mathcal{M} &\rightarrow \{0, 1\} \\ (X_1, X_2) &\mapsto \begin{cases} 1 & \text{if } X_1 \cup X_2 \text{ gives a Hamiltonian cycle in } K_{n,n} \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (2.1)$$

Going back to communication complexity and the rank of the matrix $M_{\mathfrak{F}}$, Raz and Spieker proved in their paper that

1. $\lg(\text{rank}(M_{\mathfrak{F}})) = O(n)$
2. $D(\mathfrak{F}) = \Omega(n \lg(\lg(n)))$

Therefore a gap of $\Omega(\lg(\lg(n)))$ exists between deterministic communication complexity and the log of the rank of a function's matrix. This in turn produced a new log-rank conjecture as

$$D(f) = O(\lg(\text{rank}(M_f)) \lg(\lg(\lg(\text{rank}(M_f))))$$

That is we now wonder if the deterministic communication complexity is upper bounded by the expression of the rank above.

The first point above is proven in the first part of the paper, the rank of matrix is obtained resorting to representation theory, that we'll cover extensively here. The second point is proven in the second part of the paper, it is proven in a convoluted way but overall using theorem 2.2.4, thus by what we've seen in the previous section it actually gives a lower bound for the nondeterministic communication complexity of $\mathfrak{F}$, and indeed this is what is stated in the paper, yet we need not to state it as such, for fundamentally we are interested in the log-rank conjecture.

As the first part of the paper requires an extensive knowledge of representation theory, that will be provided, we will present the second part of the paper first, and then after a great venture we'll come back to the first part.

It's worth mentioning that the paper by Raz and Spieker first appeared as a preprint in the 1993 edition of the Symposium on Foundations of Computer Science, and between 1993 and 1995, the date of publishing of their paper, a larger gap for the log-rank conjecture was found by *N. Nisan* and *A. Wigderson* in their 1994 paper "*On Rank vs. Communication Complexity*" [NW94], their work resulted in the new conjecture that

$$D(f) = O(\lg(\text{rank}(M_f))^\alpha)$$

with $\alpha = \log_3(2)$, or in other words the deterministic communication complexity of a predicate is polynomially upper bounded by the log of the rank. Since then greater gaps have been found and today the log-rank conjecture has the form

$$D(f) = O\left(\sqrt{\text{rank}(M_f)}\right)$$

This due to a recent paper by *B. Sudakov* and *I. Tomon* [ST23].

3. On the log rank conjecture - Part II

As stated in the previous section we consider the predicate $\mathcal{F}$ as defined in 2.1, and prove a lower bound for its communication complexity by an argument following theorem 2.2.4. The proof will make heavy use of *entropy*, thus we advise the reader to refresh, or get acquainted with, this concept in Appendix B.

Theorem 3.0.1. *The deterministic communication complexity of $\mathcal{F}$ satisfies*

$$D(\mathcal{F}) = \Omega(n \lg(\lg(n)))$$

Proof. As we've seen in the previous section we can think of a perfect matching on $K_{n,n}$ as a bijection $f : [n] \rightarrow [n]$, or equivalently a permutation of $[n]$ where $i \in [n]$ is permuted with $f(i)$. Letting then $\mathcal{M}$ be the set of all such matchings we have that $\mathcal{M}$ is essentially the set of all bijections on $[n]$, indeed given such a bijection we can define a perfect matching on $K_{n,n}$ and as seen the same goes the other way around, hence $|\mathcal{M}| = n!$.

For each $f \in \mathcal{M}$ define $g_f \in \mathcal{M}$ by, for $i \in [n]$, $g_f(i + 1) = f(i)$, where $i + 1$ is taken mod(n) so that $g_f(1) = f(n)$ (it is straightforward to verify that indeed g_f is a bijection on $[n]$ and so it's in the set $\mathcal{M}$). Let's look at a pair (f, g_f) , an example is given below in figure 3.1 when $n = 3$, edges that correspond to f or g_f are respectively labeled.

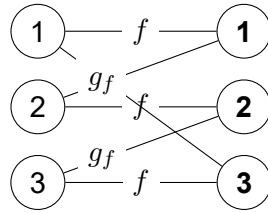


FIGURE 3.1: Example of union of matching f and g_f .

One can easily see that this is the same as seen in figure 3.2 below. Note also that this

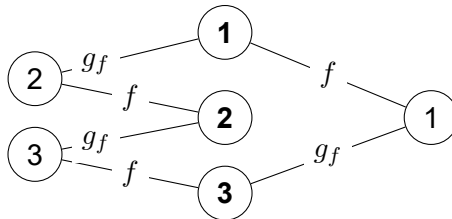


FIGURE 3.2: Example of union of matching f and g_f , rearranged.

is the same example as given in figure 2.9, which is a Hamiltonian cycle on $K_{3,3}$.

Considering our predicate $\mathfrak{F}$, it has matrix $M_{\mathfrak{F}}$ defined as

$$(M_{\mathfrak{F}})_{f,g} = \mathfrak{F}(f,g)$$

for $f, g \in \mathcal{M}$. It's straightforward to verify that $(M_{\mathfrak{F}})_{f,g_f} = 1$, indeed we'll have a cycle in the graph of the form $1 \rightarrow f(1) \rightarrow 2 \rightarrow f(2) \rightarrow \dots \rightarrow n \rightarrow f(n) \rightarrow 1$. Now because f is a bijection we clearly pass through all the nodes of $K_{n,n}$, and if this cycle is not Hamiltonian then we have within it a cycle $f(i) \rightarrow i+1 \rightarrow \dots \rightarrow f(j)$, with $j > i$ and $f(i) = f(j)$, which is not possible because f is injective and thus we must have that $j = i$.

Define now the set $\mathcal{E} = \{(f, g_f) \in \mathcal{M} \times \mathcal{M} \mid f \in \mathcal{M}\}$, with g_f as before. We think of this set as being the diagonal of the matrix $M_{\mathfrak{F}}$ (by permuting the columns of $M_{\mathfrak{F}}$ if necessary), note that each element of $\mathcal{E}$ satisfies the predicate $\mathfrak{F}$ so it's a set of 1-inputs of $\mathfrak{F}$. Let $M_1 \times M_2$ be a 1-monochromatic rectangle for $\mathfrak{F}$ and let Y be the projection of $\mathcal{E} \cap M_1 \times M_2$ on the first coordinate, that is

$$Y = \{f \in M_1 \mid (f, g_f) \in \mathcal{E} \cap M_1 \times M_2\}$$

Our goal is to prove an upper bound on the size of Y , with which we'll be able to compute a lower bound on the number of 1-monochromatic rectangles required to partition $\mathcal{E}$, thus lower bounding the number of 1-monochromatic rectangles required to partition the 1-inputs of $\mathfrak{F}$, as per theorem 2.2.4.

We start by noting two properties of the set Y , in what follows we always consider elements $i, j \in \mathbb{N}$ modulo n , so that $n+1 \equiv 1$.

1. $\forall f, f' \in Y \quad \forall i \in [n] \quad f(i) \neq f'(i+1)$

Proof. Let $f, f' \in Y$ and let's assume, for a contradiction, that there exists an $i \in [n]$ such that $f(i) = f'(i+1)$. By the definition of Y we must have that $(f, g_f), (f', g_{f'}) \in M_1 \times M_2$ and because $M_1 \times M_2$ is a (combinatorial) rectangle we must have that $(f', g_f) \in M_1 \times M_2$. But $f'(i+1) = f(i) = g_f(i+1)$ thus (f', g_f) can't possibly form a Hamiltonian cycle as there is no path connecting any $i \in [n] \setminus \{i+1\}$ to $i+1$. Thus $(M_{\mathfrak{F}})_{f',g_f} = 0$, which is a contradiction of the fact that $M_1 \times M_2$ is 1-monochromatic. $\square$

2. Let $i < j$ with $j+1 \not\equiv i-1 \pmod{n}$, then $\forall f, f' \in Y \quad \forall i, j \in [n] \quad (f(i) = f'(i), f(i+1) = f'(i+1), \dots, f(j-1) = f'(j-1), f(j) = f'(j)) \Rightarrow f(i-1) \neq f'(j+1)$

Proof. First note that if $j + 1 = i - 1 \bmod(n)$ then $\{i - 1, i, i + 1, \dots, j, j + 1\} = [n]$ so we'd be talking about the whole graph produced by the union of the matchings, thus the equality of the matchings from i to j would force $f(i - 1) = f'(j + 1)$ considering there would be no other choice for f and f' , so let $j + 1 \neq i - 1 \bmod(n)$. Let's assume for a contradiction that $f(i - 1) = f'(j + 1)$. As before we must have that $(f', g_f) \in M_1 \times M_2$, but now we have that $f'(\{i, i + 1, \dots, j, j + 1\}) = g_f(\{i, i + 1, \dots, j, j + 1\})$, that is, the image of the set $\{i, i + 1, \dots, j, j + 1\}$ by f' coincides with the image by g_f , indeed

$$\begin{aligned} f'(i) &= f(i) = g_f(i + 1) \\ f'(i + 1) &= f(i + 1) = g_f(i + 2) \\ &\vdots \\ f'(j) &= f(j) = g_f(j + 1) \\ f'(j + 1) &= f(i - 1) = g_f(i) \end{aligned}$$

Therefore, as f' is a bijection, it must be the case that $f(i - 1) \in [n] \setminus \{i, \dots, j + 1\}$, and the same goes for g_f . Thus (f', g_f) can't be a Hamiltonian cycle as there is no path connecting any $i \in [n] \setminus \{i, \dots, j + 1\}$ to $\{i, i + 1, \dots, j, j + 1\}$, so we get that $(M_{\mathfrak{F}})_{f', g_f} = 0$, which by the same reasons as before is not possible. $\square$

Note that the second property is essentially a generalization of the first.

The proof of the upper bound on Y is diverted to a lemma entitled by Raz and Spieker as *the anti-symmetry lemma*, which is stated and proven below in 3.0.2 so as to avoid any confusions now, since we need to introduce it properly and give it our undivided attention. Yet we can use it now in order to lower bound the deterministic communication complexity of the predicate $\mathfrak{F}$ and then dedicate ourselves fully to it. The lemma essentially states that

$$|Y| \leq \left(\frac{n}{\lg^\epsilon(n)} \right)^n$$

Thus by a simple rectangle size argument, considering the size of $\mathcal{E}$ is the same as the size of $\mathcal{M}$, which is $n!$, we get that the number of 1-monochromatic rectangles required to partition $\mathcal{E}$ must be at least

$$|\mathcal{E}| / \left(\frac{n}{\lg^\epsilon(n)} \right)^n = 2^{\Omega(n \lg(\lg(n)))}$$

Where the equality follows by using Sterling's approximation, $n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$, and asymptotic notation. $\square$

We may now divert our attention to *the anti-symmetry lemma*, but to avoid simply stating it and proceed with a proof without any intuitive notion of what's happening we're going to build our intuition first.

More generally we may consider the set Y as being a subset of $[n]^n$, where then a function $f \in Y$ is represented by an n -tuple $(x_1, x_2, \dots, x_n)$ such that $f(i) = x_i$. Now we can have an x_i be any element of $[n]$, but of course the size of Y must be smaller than n^n due to the two constraining properties it satisfies. So a fair question to ask is what's the (upper bound on the) average number of possibilities for an x_i . In essence we'd get an upper bound on Y of the form $(n/\alpha)^n$. With this in mind we note that by property 2. of Y if two tuples coincide on a contiguous interval, say from index i to index j , with $i < j$, then they must differ in the indices $i-1$ and $j+1$ (unless as stated $i-1 = j+1 \pmod{n}$). Taking this a step further, let A_{ij}^k be the set of tuples $(a_{i-j}, a_{i-j+1}, \dots, a_{i-1})$ for which there exists $(x_1, x_2, \dots, x_n) \in Y$ with $x_l = a_l$ for $l \in \{i-j, \dots, i-1\}$ and $x_i = k$, and B_{ij}^k the set of tuples $(a_{i+1}, a_{i+2}, \dots, a_{i+j})$ for which there exists $(x_1, x_2, \dots, x_n) \in Y$ with $x_l = a_l$ for $l \in \{i+1, \dots, i+j\}$ and $x_i = k$. Then by property 2. we must have that $A_{ij}^k \cap B_{i-(j+1),j}^k = \emptyset$. In fact it's straightforward to verify that if

$$(i' - i + 1) \pmod{n} \leq j, j'$$

then $A_{i'j'}^k \cap B_{ij}^k = \emptyset$. Moreover these properties would remain true even if we were to consider the subsets of Y whose members are consistent with the tuples in A_{ij}^k and B_{ij}^k respectively, so we can redefine the sets A_{ij}^k and B_{ij}^k as being those subsets of Y , this makes more sense probability wise when one considers the probability $\Pr(A_{ij}^k)$ when given a distribution over Y , so we want the sets A_{ij}^k and B_{ij}^k to be subsets of Y . This gives us the additional property that, if $j > j'$ the $A_{ij}^k \subseteq A_{ij'}^k$ and the same goes for the sets B_{ij}^k . All of this will be properly formalized in the beginning of the proof of the lemma.

The proof of the *anti-symmetry lemma* follows by induction, we first assume that a set $S \subset [n]^m$ that satisfies properties 1. and 2. above (recasted in a probabilistic setting) has a lower bound, of the form $(n/\alpha)^m$ with α dependent on n , that we set to be high enough to result in a contradiction. We then start with a set Y in those conditions and proceed to produce a new set $\tilde{Y}$ by fixing, on Y , a contiguous interval of indices (that is the variables x_i associated with those indices) so that $\tilde{Y}$ also satisfies properties 1. and 2. and so by

assumption satisfies the mentioned lower bound on its size (we will additionally forget the fixed indices on the set $\tilde{Y}$ thus $\tilde{Y} \subset [n]^{\tilde{m}}$ where $\tilde{m} < m$).

To achieve the contradiction we associate to Y , and in the same way to $\tilde{Y}$, a *potential function* that in a way measures the average size of the sets B_{ij}^k , such that the smaller the sets are the larger the function will be, and such that it has an upper bound. We then prove that this potential function grows substantially after the fixing, that is the function associated to $\tilde{Y}$ gets substantially larger than that of Y for a particular fixing. Thus if we repeat the fixing process a number of times, resulting in a chain of sets $\tilde{Y}$, we'll arrive at the contradiction that the potential function gets larger than its upper bound and so the assumed lower bound on the sets $\tilde{Y}$ can't hold for all sets in the chain, hence, eventually, we arrive at one set $\tilde{Y}$ for which the lower bound can't hold. From that we'll be able to prove that Y itself can't be large.

That being said the trick here is to find a fixing for which the *potential function* grows substantially. Let A_{ij}^k, B_{ij}^k the aforementioned sets associated with Y and $\tilde{A}_{ij}^k, \tilde{B}_{ij}^k$ be the analogously defined sets associated with $\tilde{Y}$, we will fix $j = n/100 = l$ and focus on the resulting sets. To our end we first prove that there exists a large contiguous interval of indices $i \in [m], [i_0 - 1, \dots, i_0 - s]$ with $s < l$, for which the sets A_{il}^k and B_{il}^k are large for a considerable number of indices i in that interval and $k \in [n]$. Then the fixing will be over a small (contiguous) portion of that interval, $[i_0 - 1, \dots, i_0 - s_0]$, with $s_0 < s$ so as to assure that a significant amount of sets B_{il}^k that are large are in the interval $[i_0 - s_0 - 1, \dots, i_0 - s]$. Then we are left to prove that of those large sets B_{il}^k in the interval $[i_0 - s_0 - 1, \dots, i_0 - s]$, after the fixing the corresponding sets $\tilde{B}_{il}^k$ associated with $\tilde{Y}$ are significantly smaller, thus inflating the potential function. This will be largely done by requiring the interval $[i_0 - 1, \dots, i_0 - s]$ to satisfy a particular property.

We now state *the anti-symmetry lemma*.

Lemma 3.0.2. *Let $Y \subset [n]^m$, and let $x = (x_1, \dots, x_n)$ be a random variable uniformly distributed on Y . If x has the following two anti-symmetry properties:*

1. $\forall_{i,k} \Pr(x_i = k) \Pr(x_{i+1} = k) = 0$
2. $\forall_{i,j,k,a_i,a_{i+1},\dots,a_{j-1},a_j} \text{ if } i-1 \neq j+1 \text{ then}$

$$\Pr(x_{i-1} = k | x_i = a_i, \dots, x_j = a_j) \cdot \Pr(x_{j+1} = k | x_i = a_i, \dots, x_j = a_j) = 0$$

where the sums on indices is taken mod(m)

Then, for n big enough

$$|Y| \leq \left(\frac{n}{\sqrt{\alpha}} \right)^m$$

with $\alpha = \lg(n)^{1/1000}$ and $m = n$.

The two properties stated here are the probability analogues of the two properties stated earlier in the proof of theorem 3.0.1. We use two different parameters, namely n and m , precisely for the fixing process described earlier, that is we can initially start with a set $Y \subset [n]^n$ but after a fixing we get a set $Y \subset [n]^m$ with $m < n$, as such the result must be more general to follow by induction.

Proof. Throughout the proof we think of the indices $i \in [m]$ as placed on a circle, which is to say we consider operations on the indices as taken $\text{mod}(m)$. Also, as we're pursuing an asymptotic result, namely Theorem 3.0.1, we are free to assume when needed that n is big enough for our purposes.

Consider the mentioned random variable $x = (x_1, \dots, x_n)$ on Y and for $i, j \in [m]$ and $k \in [n]$ define the sets A_{ij}^k and B_{ij}^k as before, their definition may be made more precise by noting that, for $a \in Y$

$$a \in A_{ij}^k \Leftrightarrow \Pr(x_i = k \mid x_{i-1} = a_{i-1}, \dots, x_{i-j} = a_{i-j}) > 0$$

$$a \in B_{ij}^k \Leftrightarrow \Pr(x_i = k \mid x_{i+1} = a_{i+1}, \dots, x_{i+j} = a_{i+j}) > 0$$

So, taking A_{ij}^k for example, if for a tuple $a \in Y$ we have that $a \in A_{ij}^k$ then there must exist a tuple $b \in Y$ such that $b_{i-1} = a_{i-1}, \dots, x_{i-j} = b_{i-j}$ and $b_i = k$ (and of course that being said b is also in A_{ij}^k).

As a consequence of these definitions it's straightforward to verify that if $j > j'$ then

$$A_{ij}^k \subseteq A_{ij'}^k \text{ and } B_{ij}^k \subseteq B_{ij'}^k$$

Moreover by the second property in the lemma

$$A_{ij}^k \cap B_{i-j-1,j}^k = \emptyset$$

And more generally for $i < i'$

$$(i - (i' + 1)) \bmod(m) \leq j, j' \Rightarrow A_{ij}^k \cap B_{i'j'}^k = \emptyset \quad (3.1)$$

This basically states that if the interval $[i-j, i-j+1, \dots, i]$ intersect with $[i'+1, \dots, i'+j']$, with $i'+1$ in the intersection, then if there exists a tuple a in the intersection $A_{ij}^k \cap B_{i'j'}^k$, we get a contradiction of property 1. or property 2. in the lemma.

As stated before we'll be looking in particular to the sets A_{il}^k and B_{il}^k with $l = n/100$. Note that l is quite large, as we'll see this is no coincidence.

If $|Y|$ is small then the entropy of x , $H(x)$, must be small for $H(x) = \lg(|Y|)$ considering we have a uniform distribution on Y (B.0.3). So we assume for a contradiction that

$$H(x) > m \left(\lg \left(\frac{n}{\alpha} \right) \right) = m(\lg(n) - \lg(\alpha)) \quad (3.2)$$

With $\alpha = \lg(n)^{1/1000}$.

The proof of the lemma is divided into three parts. In the first part we show that for a considerable number of pairs (i, k) , A_{il}^k and B_{il}^k are large. This is achieved in Claim 4. More explicitly we prove, for some $C_1, C_2 \in \mathbb{R}$ both polynomial on α^{-1} , the existence of a set I of at least $C_1 nm$ pairs (i, k) such that $\Pr(A_{il}^k), \Pr(B_{il}^k) > C_2$.

In the second part we work with the set I and show some properties of it resulting in the definition of the contiguous interval of indices we are looking to fix. The goal of this part, achieved in Claim 6, is to prove, for some $C_3, C_4, C_5 \in \mathbb{R}$ all polynomial on α^{-1} , the existence of indices $i_0, s_0 \in [m]$ and $s = s_0/C_5 < l$ such that for at least a fraction C_3 of the $k \in [n]$ the following conditions hold:

- $(i_0, k) \in I$
- For at least a fraction C_4 of the indices in the interval $[i_0 - 1, \dots, i_0 - s]$, $(i, k) \in I$
- $\Pr(A_{i_0s}^k \mid A_{i_0s_0}^k) \geq 1 - C_5$

This last property will be fundamental for us to be able to prove that the aforementioned sets $\tilde{B}_{il}^k$ are significantly smaller, after the fixing, than B_{il}^k .

As the authors of the original paper state, the heart of the proof is the third part. Here we select two indices $i_0, s_0 \in [m]$, dictated by the first two parts of the proof, and fix the variables $x_{i-1}, \dots, x_{i-s_0}$, the new set $\tilde{Y}$ resulting from this fixing will then be a subset of $[n]^{\tilde{m}}$ where $\tilde{m} = m - s_0$, this considering that we forget the fixed indices. The fixing will be chosen to satisfy two conditions; (1) the number of elements of Y that are consistent with the fixing is large (at least $|Y|/n^{s_0+3}$); (2) the value of the potential function goes up

substantially from Y to $\tilde{Y}$. We then repeat the fixing process until 3.2 can't hold anymore. The main claims of this part of the proof are Claim 9 and 10.

We'll consider the potential function $L = \sum_{i,k} \ln(2 - \ln(\Pr(B_{il}^k) + 1/n))$. Using the first two parts of the proof we show that after each fixing step the potential function is much larger, at least by $ns_0\alpha^{10}/2$, than before. Since there is an upper bound on L , namely that $L < nm(\ln(\ln(n)) + 1)$, it follows that we cannot repeat the fixing process more than a small number of times, and that the total number of fixed variables is also small, less than n/α . Therefore after that small number of possible fixings we must end up in a situation for which assumption 3.2 can't be true, and so $H(\tilde{x}) \leq \tilde{m}(\lg(n) - \lg(\alpha))$, where $\tilde{x}$ is the random variable associated with the particular set $\tilde{Y}$ we stop at. It will then follow that the original set Y is small (we'll later make this clearer).

Although originally we want the result for a set $Y \subset [n]^n$, as mentioned after a fixing we are left with a set $\tilde{Y} \subset [n]^m$, with $m < n$, and because we consider this generality of m in the proof of the lemma we must be aware of how small we let m be compared to n . As will be the case we make use of an additional assumption, that $m \geq \frac{3}{4}n$.

We start then with the first part of the proof where we prove that for a considerable number of pairs (i, k) , A_{il}^k and B_{il}^k are both large. The first two claims are for technical reasons to assist in Claim 4 and the third will help our intuition.

Claim 1.

$$\sum_{i \in [m]} H(x_i \mid x_{i-1}, \dots, x_{i-l}) \geq m \left(\lg(n) - \frac{m}{m-l} \lg(\alpha) \right) \geq m(\lg(n) - 2 \lg(\alpha))$$

$$\sum_{i \in [m]} H(x_i \mid x_{i+1}, \dots, x_{i+l}) \geq m \left(\lg(n) - \frac{m}{m-l} \lg(\alpha) \right) \geq m(\lg(n) - 2 \lg(\alpha))$$

Proof. We prove the first inequality, the second is analogous. By the chain rule for entropy, as in B.3, we have that

$$H(x) = H(x_1, \dots, x_m) = H(x_1) + H(x_2 \mid x_1) + \dots + H(x_m \mid x_1, x_2, \dots, x_{m-1})$$

Now note that, by proposition B.0.3 and B.0.7, for $i \leq l$ we have that

$$H(x_i \mid x_1, x_2, \dots, x_{i-1}) \leq H(x_i) \leq \lg(n)$$

and for $i > l$ by equation B.4 we have that

$$H(x_i \mid x_1, x_2, \dots, x_{i-1}) \leq H(x_i \mid x_{i-1}, \dots, x_{i-l})$$

Thus by 3.2

$$\begin{aligned} m(\lg(n) - \lg(\alpha)) &\leq H(x) \leq l \lg(n) + \sum_{i=l+1}^m H(x_i \mid x_{i-1}, \dots, x_{i-l}) \Leftrightarrow \\ &\Leftrightarrow \sum_{i=l+1}^m H(x_i \mid x_{i-1}, \dots, x_{i-l}) \geq (m-l) \left(\lg(n) - \frac{m}{m-l} \lg(\alpha) \right) \end{aligned}$$

Considering that $H(x_1, \dots, x_m) = H(x_i, \dots, x_m, x_1, \dots, x_{i-1})$ we obtain the same conclusion above for any rotation of the indices in $H(x_1, \dots, x_m)$, therefore we have that for all $j \in [m]$

$$\sum_{i=l+j+1}^{m+j} H(x_i \mid x_{i-1}, \dots, x_{i-l}) \geq (m-l) \left(\lg(n) - \frac{m}{m-l} \lg(\alpha) \right)$$

Thus summing over all $j \in [m]$ we get

$$\sum_j \sum_{i=l+j+1}^{m+j} H(x_i \mid x_{i-1}, \dots, x_{i-l}) \geq m(m-l) \left(\lg(n) - \frac{m}{m-l} \lg(\alpha) \right)$$

Now, it is easy to see that

$$(m-l) \sum_{i \in [m]} H(x_i \mid x_{i-1}, \dots, x_{i-l}) = \sum_j \sum_{i=l+j+1}^{m+j} H(x_i \mid x_{i-1}, \dots, x_{i-l})$$

Indeed as we sum over all $j \in [m]$ each summand $H(x_i \mid x_{i-1}, \dots, x_{i-l})$ is repeated $m-l$ times. Thus we can conclude that

$$\sum_{i \in [m]} H(x_i \mid x_{i+1}, \dots, x_{i+l}) \geq m \left(\lg(n) - \frac{m}{m-l} \lg(\alpha) \right)$$

Finally, because $m-l > m/2$ ($m > 3n/4$ and $l = n/100$) the final inequality follows immediately.* □

Define now constants $\alpha_i \in \mathbb{R}$ as

$$H(x_i \mid x_{i-1}, \dots, x_{i-l}) = \lg(n) - \lg(\alpha_i) \tag{3.3}$$

We could have a similar equality for $H(x_i \mid x_{i+1}, \dots, x_{i+l})$ yet we omit this to avoid excessive notation and unwanted redundancy. Note that the above definition pertains to the sets A_{il}^k (note the indices in the entropy). In the following claims we focus on the sets

*In fact $m-l > 296m/300$.

A_{il}^k but the proofs are general enough for the statements to apply to the sets B_{il}^k , for this reason we refrain from having two separate definitions for the constants α_i .

By Claim 1, through a simple substitution of the summands with the equality above, we get

$$\frac{1}{m} \sum_{i \in [m]} \lg(\alpha_i) \leq \frac{m}{m-l} \lg(\alpha) \leq 2 \lg(\alpha) \quad (3.4)$$

The following claim will be useful for the proof of Claim 4, the main claim in this first part of the proof of the lemma.

Claim 2.

$$\sum_{k \in [n]} \Pr(A_{il}^k) \geq n\alpha_i^{-1}$$

Proof. Denote by a an event $(x_{i-1} = a_{i-1}, \dots, x_{i-l} = a_{i-l})$ and let $\mathcal{A}$ be the set of all such events. Also denote by K_a the number of integers $k \in [n]$ for which $\Pr(x_i = k \mid a) > 0$. Then

$$\begin{aligned} \sum_{k \in [n]} \Pr(A_{il}^k) &= \sum_{k \in [n]} \sum_{a \in A_{il}^k} \Pr(a) = \sum_{k \in [n]} \sum_{\{a \in \mathcal{A} \mid \Pr(x_i = k \mid a) > 0\}} \Pr(a) \\ &= \sum_{a \in \mathcal{A}} \sum_{\{k \mid \Pr(x_i = k \mid a) > 0\}} \Pr(a) = \sum_{a \in \mathcal{A}} K_a \Pr(a) = E_{\mathcal{A}}(K_a) \end{aligned}$$

where by $E_{\mathcal{A}}(K_a)$ we mean the expected value of K_a over all events $a \in \mathcal{A}$. Now note that $H(x_i \mid a) \leq \lg(K_a)$ by the entropy bound (proposition B.0.3) and the definition of K_a , therefore, by 3.3 and the definition of conditional entropy, we have that

$$\lg(n) - \lg(\alpha_i) = H(x_i \mid x_{i-1}, \dots, x_{i-l}) = E_{\mathcal{A}}(H(x_i \mid a)) \leq E_{\mathcal{A}}(\lg(K_a))$$

By the concavity of the $\lg$ function we can conclude that

$$E_{\mathcal{A}}(\lg(K_a)) \leq \lg(E_{\mathcal{A}}(K_a)) = \lg\left(\sum_{k \in [n]} \Pr(A_{il}^k)\right)$$

Thus we have that

$$\lg(n) - \lg(\alpha_i) \leq \lg\left(\sum_{k \in [n]} \Pr(A_{il}^k)\right) \Leftrightarrow \sum_{k \in [n]} \Pr(A_{il}^k) \geq n\alpha_i^{-1}$$

□

The following claim is not used in the proof of *the anti-symmetry lemma*, yet it provides valuable information for Claim 4. In this part of the proof we wish to find a set I (of considerable

size) of pairs (i, k) for which the sets A_{il}^k and B_{il}^k are large, how large we need those sets to be will affect how many we can find, so we can't be too greedy here because we do need I to be quite large.

Claim 3.

$$\frac{1}{mn} \sum_{\substack{i \in [m], \\ k \in [n]}} \Pr(A_{il}^k) \geq \alpha^{-2} \quad \frac{1}{mn} \sum_{\substack{i \in [m], \\ k \in [n]}} \Pr(B_{il}^k) \geq \alpha^{-2}$$

We prove the first inequality, the second has a completely analogous proof (via a result analogous to Claim 2 but for the sets B_{il}^k).

Proof. By the concavity of the lg function, Claim 2 and 3.4

$$\begin{aligned} \lg \left(\frac{1}{mn} \sum_{i,k} \Pr(A_{il}^k) \right) &\geq \frac{1}{m} \sum_i \lg \left(\frac{1}{n} \sum_k \Pr(A_{il}^k) \right) \geq \\ &\geq \frac{1}{m} \sum_i -\lg(\alpha_i) \geq -2 \lg(\alpha) \end{aligned}$$

□

So on average (over $i \in [m]$ and $k \in [n]$) the sets A_{il}^k and B_{il}^k have a probability mass of at least α^{-2} .

Claim 4. *There exists a set I of pairs (i, k) such that $|I| \geq \frac{1}{2}nm\alpha^{-4}$ and for $(i, k) \in I$*

$$\Pr(A_{il}^k), \Pr(B_{il}^k) \geq \alpha^{-10}$$

Proof. The proof is divided in two parts. Our main objective is to find for each i how many k are there such that $\Pr(A_{il}^k), \Pr(B_{il}^k) \geq \alpha^{-10}$. To do so we first divide the problem and then conquer it. In the first part we work with the conditions $\Pr(A_{il}^k) \geq \alpha^{-10}$ and $\Pr(B_{il}^k) \geq \alpha^{-10}$ independently and only then we'll have enough information to attack the two conditions simultaneously. In the same spirit of previous proofs, for the first part, we work only with the sets A_{il}^k but the arguments are general enough to be directly applicable to the sets B_{il}^k .

Fix $i \in [m]$ uniformly at random and define the sets

$$K' = \{k \mid \Pr(A_{il}^k) < \alpha^{-10}\}$$

$$K'' = \{k \mid \Pr(A_{il}^k) \geq \alpha^{-10}\}$$

$$Y' = \{x \in Y \mid x_i \in K'\}$$

$$Y'' = \{x \in Y \mid x_i \in K''\}$$

In this first part of the proof our objective is to find an upper bound on the average, over $i \in [m]$, size of the sets Y' , this is the information we need for the second part of the proof.

Let x' be a random variable distributed uniformly on Y' and x'' analogously distributed on Y'' . Let $\gamma = |Y'|/|Y|$ and $\bar{\gamma}$ be the average of γ over $i \in [m]$ (we'll use the bar over a character to denote the average over $i \in [m]$ of that measure).

Define the additional binary random variable $\hat{x}$ by $\hat{x} = 0$ if $x_i \in K'$ and $\hat{x} = 1$ if $x_i \in K''$. Note that $\hat{x}$ depends solely on x_i , moreover we have a partition of Y , namely $\{Y', Y''\}$, according to the i 'th previously fixed index. With this in mind we have that

$$H(x_i \mid x_{i-1}, \dots, x_{i-l}) = H(x_i, \hat{x} \mid x_{i-1}, \dots, x_{i-l})$$

This can be easily seen by what we said above regarding $\hat{x}$ and x_i and the definition of conditional entropy. Now by the chain rule of entropy we have that

$$H(x_i, \hat{x} \mid x_{i-1}, \dots, x_{i-l}) = H(x_i \mid x_{i-1}, \dots, x_{i-l}, \hat{x}) + H(\hat{x} \mid x_{i-1}, \dots, x_{i-l})$$

We can expand the first entropy on the right side of the above equality (again using just the definition of conditional entropy) to get

$$\begin{aligned} H(x_i \mid x_{i-1}, \dots, x_{i-l}) &= \gamma H(x'_i \mid x'_{i-1}, \dots, x'_{i-l}) + \\ &+ (1 - \gamma) H(x''_i \mid x''_{i-1}, \dots, x''_{i-l}) + H(\hat{x} \mid x_{i-1}, \dots, x_{i-l}) \end{aligned}$$

Of course we can always say that $H(\hat{x} \mid x_{i-1}, \dots, x_{i-l}) \leq 1$ and

$H(x''_i \mid x''_{i-1}, \dots, x''_{i-l}) \leq \lg(n)$ by the entropy bound, thus

$$H(x_i \mid x_{i-1}, \dots, x_{i-l}) \leq \gamma H(x'_i \mid x'_{i-1}, \dots, x'_{i-l}) + (1 - \gamma) \lg(n) + 1$$

So now all we need is to upper bound $H(x'_i \mid x'_{i-1}, \dots, x'_{i-l})$ in order to proceed to upper bound $\bar{\gamma}$, as we'll see. One might ask why not upper bound $H(x''_i \mid x''_{i-1}, \dots, x''_{i-l})$ instead and apply the trivial entropy bound to the other summand, the answer to that should be made clear by the next argument.

If we define A'_{ij} similarly to A_{ij}^k but on the set Y' then by the same argument as in Claim 2 (do verify that Claim 2 is perfectly applicable here) we can conclude that

$$H(x'_i \mid x'_{i-1}, \dots, x'_{i-l}) \leq \lg \left(\sum_k \Pr(A_{il}^k) \right)$$

Now if $k \in K'$ then

$$\Pr(A_{ij}^k) = \frac{|A_{ij}^k|}{|Y'|} = \frac{|A_{ij}^k|}{\gamma|Y|} \leq \frac{|A_{ij}^k|}{\gamma|Y|} = \Pr(A_{ij}^k) \gamma^{-1}$$

and we know that $\Pr(A_{ij}^k) < \alpha^{-10}$ (this is why we choose to upper bound the entropy regarding x' , if we went with the other we wouldn't be getting an upper bound). On the other hand if k is not in K' then $\Pr(A_{il}^k) = 0$, thus we have that

$$\lg \left(\sum_{k \in [n]} \Pr(A_{il}^k) \right) < \lg \left(\sum_{k \in [n]} \gamma^{-1} \alpha^{-10} \right) = \lg(n \gamma^{-1} \alpha^{-10})$$

Therefore

$$H(x'_i \mid x'_{i-1}, \dots, x'_{i-l}) \leq \lg(n) - \lg(\gamma \alpha^{10})$$

Getting all together now we get

$$H(x_i \mid x_{i-1}, \dots, x_{i-l}) \leq 1 + \lg(n) - \gamma \lg(\gamma \alpha^{10}) \leq \lg(n) - 10\gamma \lg(\alpha) + 2$$

By 3.3 and 3.4, when summing over $i \in [m]$, recalling that i is fixed uniformly at random, we have that

$$\begin{aligned} m(\lg(n) - 2\lg(\alpha)) &\leq \sum_i H(x_i \mid x_{i-1}, \dots, x_{i-l}) \leq \\ &\leq \sum_i (\lg(n) - 10\gamma \lg(\alpha) + 2) \\ &\Leftrightarrow \lg(n) - 2\lg(\alpha) \leq \lg(n) - 10\bar{\gamma} \lg(\alpha) + 2 \\ &\Leftrightarrow \bar{\gamma} \leq \frac{2\lg(\alpha) + 2}{10\lg(\alpha)} \end{aligned}$$

Thus, for α large enough, we can conclude that $\bar{\gamma} \leq 1/4$.

As stated before all is equally valid for the same arguments regarding the sets B_{il}^k , as one may verify.

We now proceed to the second part of the proof. We'll use a similar information theoretical argument to the one in the first part, that alone should hint as to why we spent this time finding an upper bound for $\bar{\gamma}$, none the less it will become evident.

As before, fix $i \in [m]$ uniformly at random and define the sets

$$K^0 = \{k \mid \Pr(A_{il}^k), \Pr(B_{il}^k) \geq \alpha^{-10}\}$$

$$K^{00} = [n] \setminus K^0$$

$$Y^0 = \{x \in Y \mid x_i \in K^0\}$$

$$Y^{00} = \{x \in Y \mid x_i \in K^{00}\}$$

As stated before our goal now it to lower bound the sum over all $i \in [m]$ of each set K^0 , or analogously lower bound the average size of K^0 , considering this simply adds a factor of $1/m$ to the computation. So in the end we get a lower bound for the number of pairs (i, k) that satisfy the desired conditions $\Pr(A_{il}^k), \Pr(B_{il}^k) \geq \alpha^{-10}$.

Let x^0 be a random variable uniformly distributed on Y^0 , x^{00} similarly distributed on Y^{00} and let $\delta = |Y^0|/|Y|$. Since $\bar{\gamma} \leq 1/4$ (for the sets Y' , regarding both A_{il}^k and B_{il}^k) then $\bar{\delta} \geq 1/2$ and as before we have

$$\lg(n) - \lg(\alpha_i) = H(x_i | x_{i-1}, \dots, x_{i-l}) \leq$$

$$\delta H(x_i^0 | x_{i-1}^0, \dots, x_{i-l}^0) + (1 - \delta) H(x_i^{00} | x_{i-1}^{00}, \dots, x_{i-l}^{00}) + 1$$

Noting that $H(x_i^0 | x_{i-1}^0, \dots, x_{i-l}^0) \leq \lg(|K^0|)$ and $H(x_i^{00} | x_{i-1}^{00}, \dots, x_{i-l}^{00}) \leq \lg(n)$ we get

$$\lg(n) - \lg(\alpha_i) \leq \delta \lg(|K^0|) + (1 - \delta) \lg(n) + 1$$

Therefore we have that

$$\delta(\lg(n) - \lg(|K^0|)) \leq \lg(\alpha_i) + 1$$

Now for the final blow, recall that in these expressions δ and K^0 are dependent on $i \in [m]$, fixed uniformly at random, so lets give them subscripts to make that evident and consider $\sum_i \delta_i(\lg(n) - \lg(|K_i^0|))$. Let Z be the $m/2$ indices i for which $(\lg(n) - \lg(|K_i^0|))$ is smallest. With this we note the following:

- i) At least for half of the indices i we must have $\delta_i > 0$, otherwise as for all $i \in [m]$, $\delta_i \leq 1$ and the distribution is uniform we couldn't possibly have $\bar{\delta} \geq 1/2$.
- ii) Worst case scenario for the value of the sum $\sum_i \delta_i(\lg(n) - \lg(|K_i^0|))$, $\delta_i = 1$ for the indices $i \in Z$, or in other words this case minimizes the sum $\sum_i \delta_i(\lg(n) - \lg(|K_i^0|))$.

As such we have that

$$\sum_{i \in Z} (\lg(n) - \lg(|K_i^0|)) \leq \sum_{i \in [m]} \delta_i (\lg(n) - \lg(|K_i^0|)) \leq \sum_{i \in [m]} (\lg(\alpha_i) + 1)$$

Now we note that in 3.4 it's actually also true that $m/(m-l) \leq 1.5$, and so by 3.4 with that modification, $\sum_i (\lg(\alpha_i) + 1) \leq 1.5m \lg(\alpha) - m$, and for α big enough so that $m < 0.5m \lg(\alpha)$ we get that $\sum_i (\lg(\alpha_i) + 1) < 2m \lg(\alpha)$, and as a result

$$\sum_{i \in Z} (\lg(n) - \lg(|K_i^0|)) \leq 2m \lg(\alpha)$$

From this we can straightforwardly conclude that

$$\sum_{i \in Z} \lg(|K_i^0|) \geq \frac{1}{2}m \lg(n\alpha^{-4})$$

And by the concavity of the log function

$$\frac{1}{2}m \lg\left(\frac{1}{\frac{1}{2}m} \sum_{i \in Z} |K_i^0|\right) \geq \sum_{i \in Z} \lg(|K_i^0|)$$

Thus

$$\sum_i |K_i^0| \geq \sum_{i \in Z} |K_i^0| \geq \frac{1}{2}mn\alpha^{-4}$$

Therefore, taking $I = \{(i, k) \mid i \in [m] \text{ and } k \in K_i^0\}$ we have that

$$|I| = \sum_i |K_i^0| \geq \frac{1}{2}mn\alpha^{-4}$$

This concludes the proof. □

We now proceed to work with the set I from the last claim, starting by giving it a bit more structure. For every $k \in [n]$ define

$$I_k = \{i \mid (i, k) \in I\}$$

Additionally define

$$Z = \{k \mid |I_k| \geq \frac{1}{4}m\alpha^{-4}\}$$

We want to find a lower bound for $|Z|$ in order to redefine the set I to be the union of the I_k for $k \in Z$, thus maintaining knowledge of a lower bound for the size of the resulting set. To that end we may simply distribute the mass of I over the sets I_k so as to minimize the number of $k \in [n]$ for which $|I_k| \geq \frac{1}{4}m\alpha^{-4}$. That is achieved by first seeing how many I_k we can fill to the maximum of $\frac{1}{4}m\alpha^{-4}$, which is $|I|/((1/4)m\alpha^{-4})$, that considering the

lower bound we have on I from the last claim gives us $2n$, but because we have only n sets I_k we have an excess mass of $n\frac{1}{4}m\alpha^{-4}$ still to distribute, so now all we have to do is see how many I_k can we fill to the maximum of m on top of the $\frac{1}{4}m\alpha^{-4}$ already there, that is

$$\frac{n\frac{1}{4}m\alpha^{-4}}{m - \frac{1}{4}m\alpha^{-4}} > \frac{1}{4}n\alpha^{-4}$$

Thus we can conclude that $|Z| > \frac{1}{4}n\alpha^{-4}$. Now redefine I to be

$$I = \{(i, k) \mid k \in Z \text{ and } i \in I_k\}$$

So we now have that $|I| > \frac{1}{16}nm\alpha^{-4}$.

This concludes the first part of the proof of the Lemma, and as such perhaps it is now a good time to recall the objectives for the second part of the proof.

Fix $k \in Z$ and look at the set I_k . For each $i \in I_k$ and distance $s \in [m]$ look at the set of all indices in $I_k \cap [i - 1, \dots, i - s]$. With this define

$$J_{is}^k = \{j \mid j \in I_k \cap [i - 1, \dots, i - s]\}$$

In this part we'll be looking for an index i_0 and a distance $s < l$ such that for a fraction C_3 of the $k \in Z$, $(i_0, k) \in I$ and $\Pr(A_{i_0s}^k) \geq (1 - C_5) \Pr(A_{i_0s_0}^k)$ where s_0 is such that $s = s_0/C_5$. Moreover for those k at least a fraction C_4 of the $i \in [i_0 - 1, \dots, i_0 - s]$, $(i, k) \in I$. That is, $|J_{i_0s}^k| \geq sC_4$ and $\Pr(A_{i_0s}^k) \geq (1 - C_5) \Pr(A_{i_0s_0}^k)$ for C_3 of the $k \in Z$.

To that end we start with the objective $|J_{is}^k| \geq sC_4$, for some set of triples (i, k, s) , and after this we fix some i_0 and s and lower bound the number of k that also satisfy the other condition $\Pr(A_{i_0s}^k) \geq (1 - C_5) \Pr(A_{i_0s_0}^k)$.

For $k \in Z$, consider I_k and a distance $s \in [m]$. To minimize J_{is}^k over all $i \in I_k$, in order to find a lower bound for their size, we could distribute the $|I_k|$ indices i equidistantly from each other over the $[m]$ possible positions, then for all $i \in I_k$, $|J_{is}^k|$ would be about $\frac{s}{m}|I_k| - 1$, ($\frac{s}{m|I_k|}$ elements in the interval $[i, \dots, i - s]$ minus one because we don't count the position i in the set J_{is}^k). Of course I_k is not necessarily distributed this way over the $[m]$ possible positions and so some of the J_{is}^k , for $i \in I_k$, can be larger than others, so a fair question to ask is how many of them are larger than $\frac{s}{2m}|I_k| - 1$. As it turns out for at least half of the $i \in I_k$, $|J_{is}^k|$ satisfies this bound, as the next claim asserts.

Claim 5. For $k \in Z$ and any distance $s \leq m$ there are at least

$$\frac{1}{2}|I_k| \geq \frac{1}{8}m\alpha^{-4}$$

indices $i \in I_k$ for which

$$|J_{is}^k| \geq \frac{s}{2m}|I_k| - 1 \geq \frac{1}{8}s\alpha^{-4} - 1$$

Proof. The proof follows from a straightforward combinatorial lemma:

Lemma 3.0.3. Let X be a subset of $\{0, \dots, m-1\}$ and let $s \leq m$ be a positive integer. For $i \in X$ let $J_i = X \cap [i-1, \dots, i-s]$. Let X' be the subset of X consisting of the indices i for which $|J_i| \geq |X|s/2m - 1$, then $|X'| \geq |X|/2$

Proof. Let $X'' = X \setminus X'$ (that is the complement of X' in X). By averaging, over all $i \in [m]$ uniformly distributed, and letting s_i be the number of elements of X'' in the interval $[i-1, \dots, i-s]$, we get

$$\frac{1}{m} \sum_i s_i = \frac{1}{m} \sum_i |X'' \cap [i-1, \dots, i-s]| = \frac{s}{m}|X''|$$

This because as we go over all $i \in [m]$ we count the whole set X'' , and any elements of X'' in the interval $[i-1, \dots, i-s]$ will be counted s times in the sum.

Hence there must exist one such interval that contains at least $|X''|s/m$ elements of X'' , let i' be the last element of X'' in this interval. Now as $X'' = X \setminus X'$ and $i' \in X''$, so $|J_{i'}| < |X|s/2m - 1$, we have that $|J_{i'}| \geq |X''|s/m - 1$, therefore by these two inequalities we get that

$$|X''|s/m - 1 < |X|s/2m - 1$$

and so we can conclude that $|X''| < |X|/2$. □

Now to finish the proof of the claim all we have to do is set $X = I_k$ and recall that $|I_k| \geq \frac{1}{4}m\alpha^{-4}$. □

With this done we proceed to the final claim of this part of the proof of the anti-symmetry lemma where we prove the existence of an i_0 and a distance s , per the previous claim, for which we can find a nice lower bound for the number of $k \in Z$ such that our second and final objective, namely that $\Pr(A_{i_0s}^k \mid A_{i_0s_0}^k) \geq (1 - C_5)$ where s_0 is such that $s = s_0/C_5$, also holds. This last objective will be of use later in the proof of the lemma (in Claim 8) and s_0 is in fact the distance we'll use for the fixing process, it basically states that for an

event $a \in A_{i_0 s_0}^k$ there are on average more than $(1 - C_5)$ elements in $A_{i_0 s}^k$ consistent with that event.

Fix $\alpha_d = 10^6 \alpha^{100}$ (a big monomial in α) and let $s_r = \alpha_d^r \lg(n)$ with $r \in [d]$ where $d = \alpha_d^4$. These s_r 's are to be the distances we'll consider. We note that these grow geometrically with r , this will be important, as we are going to find a particular interval in $[m]$ for which $J_{i_0 s_r}^k$ is large for a lot of k , then by fixing the variables in the interval $[i_0 - 1, \dots, i_0 - s_{r-1}]$ we will still be left with a lot of large sets A_{il}^k and B_{il}^k in the interval $[i_0 - s_{r-1}, \dots, i_0 - s_r]$ and thus be left with a greater chance to observe a significant increase in our *potential function*. Also note that $s_d = \alpha_d^4 \lg(n) < l = n/100$,* this again will be important later in the proof (in Claim 8).

In order to make the path ahead more manageable we'll need a couple new definitions that wrap up our objectives here concisely.

Call (i, k, r) a *good triple* if $k \in Z$, $i \in I_k$ and

$$|J_{i s_r}^k| \geq \frac{1}{8} s_r \alpha^{-4} - 1$$

So a triple as in the last claim. Considering that $|Z| \geq \frac{1}{4} n \alpha^{-4}$ and the statements of the previous claim, the number of *good triples* is at least

$$d|Z| \frac{1}{2} |I_k| \geq \frac{d}{32} n m \alpha^{-8}$$

For each *good triple* (i, k, r) such that $r > 1$ define

$$\gamma_{ir}^k = \Pr(A_{i s_r}^k \mid A_{i s_{r-1}}^k)$$

Call (i, k, r) a *nice triple* if $r > 1$ and $\gamma_{ir}^k > 1 - \alpha_d^{-1}$.

This last definition of a *nice triple* pertains to our objective $\Pr(A_{i_0 s}^k \mid A_{i_0 s_0}^k) \geq (1 - C_5)$, after we prove the next claim we'll make this evident. The advantages of having a *good triple* should be clear by now, for the advantages of having a *nice triple* that will be made clear in Claim 8 and the reason we need such property are for technical reasons. As we'll see it will allow us to have a link between the current set Y and the set Y' obtained after the fixing of variables in an interval.

We are now ready to proceed to the next claim.

*Just recall the definition of each variable as a function of n and the simple rule that $x = 2^{\lg(x)}$, and the inequality follows with some straightforward expression manipulations and bound observations

Claim 6. *There exist two indices i_0, r_0 and a set $W \subseteq Z$ such that $|W| \geq 10^{-2}n\alpha^{-8}$ and for all $k \in W$, (i_0, k, r_0) is good and nice.*

Proof. Fix $k \in Z$ and $i \in I_k$. By the definition of I and the fact that if $j > j'$ then $A_{ij}^k \subset A_{ij'}^k$, we have

$$\begin{aligned} \alpha^{-10} &\leq \Pr(A_{il}^k) \leq \Pr(A_{is_d}^k) = \\ &= \Pr(A_{is_1}^k) \Pr(A_{is_2}^k \mid A_{is_1}^k) \dots \Pr(A_{is_d}^k \mid A_{is_{d-1}}^k) \leq \prod_{r=2}^d \gamma_{ir}^k \end{aligned}$$

We note that $\alpha^{-10} \gg (1 - \alpha_d^{-1})^{\alpha_d^2}$, that is α^{-10} is much larger than $(1 - \alpha_d^{-1})^{\alpha_d^2}$,* hence

$$\prod_{r=2}^d \gamma_{ir}^k > (1 - \alpha_d^{-1})^{\alpha_d^2} \quad (3.5)$$

Now because $\gamma_{ir}^k \leq 1$ if more than α_d^2 of the γ_{ir}^k , for $r = 2, \dots, d$, satisfy $\gamma_{ir}^k < 1 - \alpha_d^{-1}$ then even if all the remaining r have $\gamma_{ir}^k = 1$ we couldn't possibly have 3.5 be true. As such we have that, given $k \in Z$ and $i \in I_k$, for at most α_d^2 of the $r \in \{2, \dots, d\}$, $\gamma_{ir}^k < 1 - \alpha_d^{-1}$.

Summing over all possible $k \in Z$ and $i \in I_k$ (at most nm such pairs (i, k) can exist) we have at most $nm\alpha_d^2$ triples (i, k, r) that are **not** nice. But we have at least $dnm\alpha^{-8}/32$ good triples and

$$\frac{\alpha_d^2}{d} = \alpha_d^{-2} = 10^{-12} \lg(n)^{-0.2} \ll \frac{1}{32} \lg(n)^{-0.08} = \frac{1}{32} \alpha^{-8}$$

Therefore $\alpha_d^2 nm \ll \frac{1}{32} dnm\alpha^{-8}$, that is, the number of **not** nice triples is negligible compared to the number of good triples, hence most of the good triples are also nice, so there are more than, say, $10^{-2} dnm\alpha^{-8}$ good and nice triples.

For $i \in [m]$, $r \in \{2, \dots, d\}$ let W be the set of $k \in Z$ for which the triple (i, k, r) is good and nice. If for all $i \in [m]$, $r \in \{2, \dots, d\}$ we have $|W| < 10^{-2}n\alpha^{-8}$ (where W is the set mentioned in the claim) then we couldn't possibly have more than $dm10^{-2}n\alpha^{-8}$ good and nice triples, and so for some i_0 and r_0 we must have that $|W| \geq 10^{-2}n\alpha^{-8}$, completing the proof. $\square$

Note that for (i_0, k, r_0) per the last claim we have a nice triple and so

$$\Pr(A_{i_0 s_{r_0}}^k \mid A_{i_0 s_{r_0-1}}^k) > (1 - \alpha_d^{-1})$$

Also $s_{r_0} = \alpha_d s_{r_0-1} = \frac{s_{r_0-1}}{C_5}$, with then $C_5 = \alpha_d^{-1}$ (additionally $C_3 = 10^{-2}n\alpha^{-8}$ and $C_4 > \frac{1}{8}r_0\alpha^{-4} - 1$) thus the last promise for this part of the proof is left fulfilled.

*Use the definition of each variable and note that for n large $(1 - \alpha_d^{-1})^{\alpha_d^2} \approx e^{-\alpha_d}$.

We are now ready for the third and final part of the proof of the lemma. We'll start by defining the *potential function*, setting some notation and then go over some of the main ideas for this final part of the proof.

Let's then define the *potential function*. For $x \in [0, 1]$ let $f(x) = \ln(2 - \ln(x + 1/n))$, this function behaves like $\ln(-\ln(x))$ seen in figure 3.3 below. The factor $1/n$ is added to give

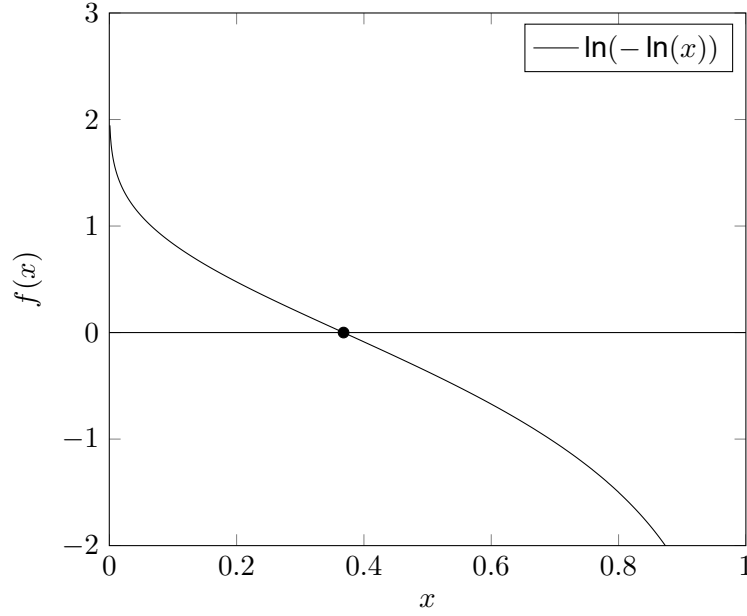


FIGURE 3.3: Plot of $\ln(-\ln(x))$ and marked inflection point.

an upper bound to the function $\ln(-\ln(x))$ in the interval $[0, 1]$, the constant 2 is added to assure the function is convex in the interval $[0, 1]$, the result of this, $f(x)$, is seen in figure 3.4 below, with $n = 100$, note it will also be monotone decreasing in this interval.

Now for the *potential function*, define

$$L = \sum_{i,k} f(\Pr(B_{il}^k))$$

In a way L measures the average size of the sets B_{il}^k , although in an inverted way, if they are mostly small then L will be large and vice versa.

Fix i_0, r_0 and W as per the last claim and set $s_0 = s_{r_0-1}$. As previously said we are going to fix the values of $x_{i_0-1}, \dots, x_{i_0-s_0}$, this is done randomly according to the distribution we have of x on Y . When we fix those s_0 variables we'll be left with a new scheme with $\tilde{m} = m - s_0$ random variables and we disregard the fixed variables, leaving them out of the new scheme. Note that both conditions of the lemma still hold (notionally it's still a subset of Y), except of course now we can no longer have $\tilde{m} = n$ for we want the initial

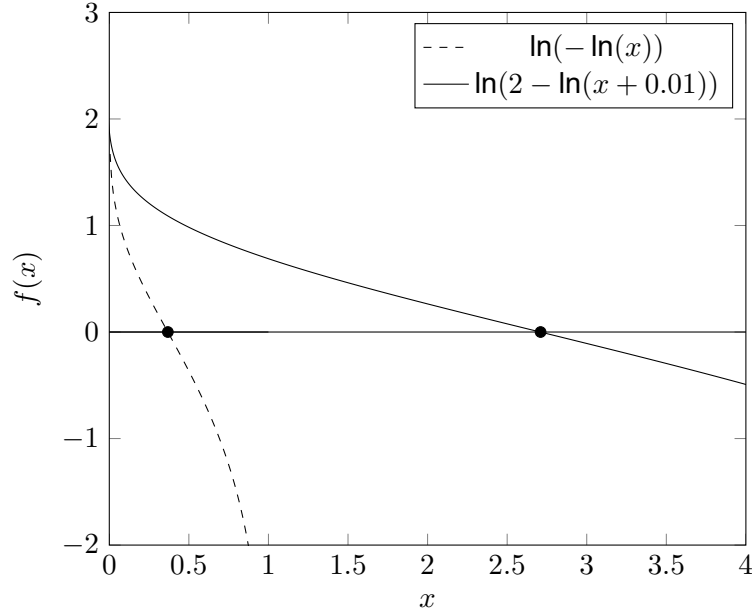


FIGURE 3.4: Plot of $\ln(-\ln(x))$ (dashed) and its inflection point, and plot of $\ln(2 - \ln(x + 0.01))$ and its inflection point.

set Y to enjoy that property, as such by assumption it satisfies 3.2. We'll call this new set $\tilde{Y}$.

For $\tilde{Y}$ define $\tilde{A}_{ij}^k$ and $\tilde{B}_{ij}^k$ similarly as done before for Y and define $\tilde{L}$, the *potential function* associated to this new scheme. Note that $\tilde{L}$ is a random variable on $(x_{i_0-1}, \dots, x_{i_0-s_0})$, it depends solely on the particular fixing we end up choosing over that interval.

We'll denote by a an event $(x_{i_0-1} = a_{i_0-1}, \dots, x_{i_0-s_0} = a_{i_0-s_0})$, i.e. a particular fixing of the variables $(x_{i_0-1}, \dots, x_{i_0-s_0})$, and consider $\tilde{L}$ as a function of the event a .

In an effort to simplify things going forward we note the following, if we use the same indices for the variables in the new scheme as in the previous one (there taken from the set $[m]$) then some indices will be missing (the fixed ones), for example if $i_0 = 8$, $s_0 = 4$ and $m = 9$ we'd get the scheme (x_1, x_2, x_8, x_9) , and as a result the sets $\tilde{A}_{il}^k$ and $\tilde{B}_{il}^k$ can depend on less than l variables, yet correcting for this by increasing l ad-hoc can only reduce the size of these sets (by the properties of these sets) and so increase the *potential function* $\tilde{L}$, as a result using this simplification of the notation for the indices won't bring any harm, as we'll see, and so we adopt it for the new scheme.

The main idea for this part of the proof of the lemma is as follows; because $\tilde{Y}$ satisfies the condition of the lemma, by assumption 3.2 is true for $\tilde{Y}$ and so everything proved so far for Y remains true for $\tilde{Y}$ just the same. The goal of the next four claims is to show the existence of an event a (a fixing), with a probability of occurring that we can lower bound,

for which $\tilde{L}$ is significantly larger than L (achieved finally in Claim 10). With that we'll be able to use induction to prove that for some $\tilde{Y}$ achieved in a bounded number of fixing steps condition 3.2 can't hold anymore.

We assume that Y (a set in the conditions of the lemma) satisfies condition 3.2 and as a result if Y does then $\tilde{Y}$ also does because it satisfies the condition of the lemma, this gives our inductive step, but because after a bounded number of fixing steps $\tilde{L}$ ends up being greater than its upper bound due to its large growth we get a contradiction. Thus either Y doesn't satisfy 3.2, and the upper bound on Y follows immediately, or for some $\tilde{Y}$ achieved after a bounded number of steps 3.2 doesn't hold anymore. Now note that the size of $\tilde{Y}$ is given by $\Pr(a)|Y|$, i.e. $|\tilde{Y}| = \Pr(a)|Y|$, and because we do this fixing a bounded number of times until we get a contradiction, say t times, we have that

$$|Y| = \prod_i^t \Pr(a_i)^{-1} |\tilde{Y}_t|$$

Where a_i is the particular fixing event at step i and $\tilde{Y}_t$ the set obtained after the t fixing steps. Thus because we have an upper bound on the set $\tilde{Y}_t$ due to it not satisfying condition 3.2 all we need is then a lower bound on the probability of each fixing a_i to upper bound $|Y|$.

To prove the existence of the aforementioned event a (for which $\tilde{L}$ is significantly larger than L) we analyze $E_a(\tilde{L})$, that is the expected value of $\tilde{L}$ over all possible events a , and attempt to prove that $E_a(\tilde{L}) > L + C_6$ where C_6 is a polynomial on α . To that end we start with a more local argument (Claim 7 and 8) where we first prove that $E_a(f(\Pr(\tilde{B}_{il}^k))) \geq f(\Pr(B_{il}^k))$ for all possible tuples (i, k) and then that for some of those tuples

$$E_a(f(\Pr(\tilde{B}_{il}^k))) \geq f(\Pr(B_{il}^k)) + C$$

with C a polynomial on α^{-1} .

Claim 7. *If x_i was not fixed * then*

$$E_a(\Pr(\tilde{B}_{il}^k)) = \Pr(B_{il}^k)$$

Proof. Let then index i be not among the fixed indices, we that that

$$\Pr(B_{il}^k) = \sum_a \Pr(B_{il}^k \cap a) = \sum_a \Pr(B_{il}^k | a) \Pr(a) =$$

*If the index i was fixed then $\tilde{B}_{il}^k$ doesn't exist for there would be no index i in $\tilde{Y}$.

$$= \sum_a \Pr(\tilde{B}_{il}^k) \Pr(a) = E_a(\Pr(\tilde{B}_{il}^k))$$

□

With this claim we can conclude, by the convexity of $f(x)$, that

$$E_a(f(\Pr(\tilde{B}_{il}^k))) \geq f(E_a(\Pr(\tilde{B}_{il}^k))) = f(\Pr(B_{il}^k)) \quad (3.6)$$

Claim 7 will be of use in the next claim.

Claim 8. *If $k \in W$, $i \in J_{i_0 s_{r_0}}^k$ and x_i was not fixed (i.e. $i \in J_{i_0 s_{r_0}}^k \cap [i_0 - s_{r_0-1} - 1, i_0 - s_{r_0-1} - 2, \dots, i_0 - s_{r_0}]$) then*

$$E_a(f(\Pr(\tilde{B}_{il}^k))) \geq f(\Pr(B_{il}^k)) + \frac{1}{10}\alpha^{-10}$$

Proof. Consider $f(\Pr(\tilde{B}_{il}^k)) - f(\Pr(B_{il}^k))$ as a function of the event a . Because (i_0, k, r_0) is a nice triple we have that

$$\Pr(A_{i_0 s_{r_0}}^k \mid A_{i_0 s_{r_0-1}}^k) > 1 - \alpha^{-1}$$

So on average, over $a \in A_{i_0 s_{r_0-1}}^k$, we have that

$$E_{a \in A_{i_0 s_{r_0-1}}^k}(\Pr(\tilde{A}_{i_0 s_{r_0}}^k)) > 1 - \alpha^{-1}$$

Now note that, because $i \in J_{i_0 s_{r_0}}^k$, then $i \in [i_0 - 1, \dots, i_0 - s_{r_0}]$ and so

$$(i_0 - (i + 1)) \bmod(m) \leq s_{r_0} < l$$

thus by property 3.1 of the sets $\tilde{A}_{ij}^k$ and $\tilde{B}_{ij}^k$ we must have that $\tilde{A}_{i_0 s_{r_0}}^k \cap \tilde{B}_{il}^k = \emptyset$. So by the previous inequality

$$E_{a \in A_{i_0 s_{r_0-1}}^k}(\Pr(\tilde{B}_{il}^k)) \leq \alpha_1^{-1} < \alpha^{-50}$$

Moreover $i \in J_{i_0 s_{r_0}}^k \subseteq I$ so $\Pr(B_{il}^k) \geq \alpha^{-10}$. With this in mind first note that

$$E_{a \in A_{i_0 s_{r_0-1}}^k}(f(\Pr(\tilde{B}_{ij}^k)) - f(\Pr(B_{ij}^k))) = E_{a \in A_{i_0 s_{r_0-1}}^k}(f(\Pr(\tilde{B}_{ij}^k))) - f(\Pr(B_{ij}^k))$$

Now by the convexity of f

$$E_{a \in A_{i_0 s_{r_0-1}}^k}(f(\Pr(\tilde{B}_{ij}^k))) \geq f(E_{a \in A_{i_0 s_{r_0-1}}^k}(\Pr(\tilde{B}_{ij}^k)))$$

therefore, recalling also that f is a decreasing function, we have that

$$E_{a \in A_{i_0 s_{r_0-1}}^k}(f(\Pr(\tilde{B}_{ij}^k)) - f(\Pr(B_{ij}^k))) > f(\alpha^{-50}) - f(\alpha^{-10}) > \ln(4)$$

for α large enough.*

Recall that the triple (i_0, k, r_0) is good so $(i_0, k) \in I$ and so $\Pr(A_{i_0 l}^k) > \alpha^{-10}$. Considering that $A_{i_0 l}^k \subseteq A_{i_0 s_{r_0-1}}^k$ then

$$\Pr(A_{i_0 s_{r_0-1}}^k) \geq \Pr(A_{i_0 l}^k) > \alpha^{-10}$$

As such the contribution to the average $E_a(f(\Pr(\tilde{B}_{ij}^k)) - f(\Pr(B_{ij}^k)))$, of the events a in the set $A_{i_0 s_{r_0-1}}^k$ must be greater than $\ln(4)\alpha^{-10}$.

When the fixing events a are not members of $A_{i_0 s_{r_0-1}}^k$ then $f(\Pr(\tilde{B}_{ij}^k)) - f(\Pr(B_{ij}^k))$ may of course be negative but by virtue of Claim 7 and the fact that in f we take the log twice we'll prove that these negative numbers are negligible on the average.

Let $\delta = \Pr(B_{ij}^k)$ (so $\delta \geq \alpha^{-10}$) and consider the events a with small $\Pr(\tilde{B}_{ij}^k)$. Construct the set $G = \{a \mid \Pr(B_{ij}^k \mid a) \text{ is small}\}$ by successively adding events a for which $\Pr(\tilde{B}_{ij}^k)$ is the smallest (recall that $\Pr(B_{ij}^k \mid a) = \Pr(\tilde{B}_{ij}^k)$), until $\Pr(G) = \alpha^{-10}$ and let G' be the complement of G , so then $\Pr(G') = 1 - \alpha^{-10}$.

As we've seen $\Pr(A_{i_0 s_{r_0-1}}^k) > \alpha^{-10}$ and

$$E_{a \in A_{i_0 s_{r_0-1}}^k} (f(\Pr(\tilde{B}_{ij}^k)) > f(\Pr(B_{ij}^k))) + \ln(4)$$

therefore

$$E_{a \in G} (f(\Pr(\tilde{B}_{ij}^k)) > f(\Pr(B_{ij}^k))) + \ln(4)$$

as, by the definition of G , and in particular that $\Pr(G) = \alpha^{-10}$, it must be the case that $E_{a \in G}(\Pr(\tilde{B}_{ij}^k)) \leq E_{a \in A_{i_0 s_{r_0-1}}^k}(\Pr(\tilde{B}_{ij}^k))$.

We proceed to further simplify the problem by noting the following, by the convexity of f

$$\begin{aligned} E_a(f(\Pr(\tilde{B}_{ij}^k))) &= E_{a \in G}(f(\Pr(\tilde{B}_{ij}^k))) \Pr(G) + E_{a \in G'}(f(\Pr(\tilde{B}_{ij}^k))) \Pr(G') > \\ &> f(E_{a \in G}(\Pr(\tilde{B}_{ij}^k))) \Pr(G) + f(E_{a \in G'}(\Pr(\tilde{B}_{ij}^k))) \Pr(G') \end{aligned}$$

Letting $g = E_{a \in G}(\Pr(\tilde{B}_{ij}^k))$ and $g' = E_{a \in G'}(\Pr(\tilde{B}_{ij}^k))$ we get

$$E_a(f(\Pr(\tilde{B}_{ij}^k))) > f(g) \Pr(G) + f(g') \Pr(G')$$

Therefore, as far as proving a lower bound for $E_a(f(\Pr(\tilde{B}_{ij}^k)))$ we might as well assume that

*To see the last inequality note that $f(\alpha^{-50}) - f(\alpha^{-10}) = \ln\left(\frac{2 - \ln(\alpha^{-50} + 1/n)}{2 - \ln(\alpha^{-10} + 1/n)}\right)$ and if α (and n) is large enough then $1/n$ is negligible and $-\ln(\alpha^{-10})$ is a large number making the constant 2 also negligible and so $\ln\left(\frac{2 - \ln(\alpha^{-50} + 1/n)}{2 - \ln(\alpha^{-10} + 1/n)}\right) \approx \ln\left(\frac{50 \ln(\alpha)}{10 \ln(\alpha)}\right)$

the $P(\tilde{B}_{il}^k)$ are constant on G , and on G' .^{*} As such we assume from now on that $P(\tilde{B}_{il}^k) = g$ for all $a \in G$ and $P(\tilde{B}_{il}^k) = g'$ for all $a \in G'$. We this we have that $f(g) > f(\delta) + \ln(4)$. Also, by Claim 7

$$\delta = \Pr(B_{ij}^k) = E_a(\Pr(\tilde{B}_{il}^k)) = g \Pr(G) + g' \Pr(G') \geq g' \Pr(G') = g'(1 - \alpha^{-10})$$

Therefore

$$g' \leq \frac{\delta}{1 - \alpha^{-10}}$$

We are now ready to finalize the proof, we have that

$$\begin{aligned} E_a(f(P(\tilde{B}_{il}^k))) &= f(g)\alpha^{-10} + f(g')(1 - \alpha^{-10}) > \\ &> (f(\delta) + \ln(4))\alpha^{-10} + f\left(\frac{\delta}{1 - \alpha^{-10}}\right)(1 - \alpha^{-10}) >^\dagger \\ &> (f(\delta) + \ln(4))\alpha^{-10} + f(\delta(1 + 2\alpha^{-10}))(1 - \alpha^{-10}) \end{aligned}$$

By the Taylor series we can approximate $f(\delta(1 + 2\alpha^{-10}))$ centering the series at δ and considering only the series up to the first degree, i.e. $f(\delta(1 + 2\alpha^{-10})) \approx f(\delta) + 2\delta\alpha^{-10}f'(\delta)$, but because the function is convex and monotone decreasing we actually get that $f(\delta(1 + 2\alpha^{-10})) \geq f(\delta) + 2\delta\alpha^{-10}f'(\delta)$. Hence the above expression is greater than

$$\begin{aligned} (f(\delta) + \ln(4))\alpha^{-10} + (f(\delta) + 2\delta\alpha^{-10}f'(\delta))(1 - \alpha^{-10}) &> \\ &> f(\delta) + \alpha^{-10}(\ln(4) + 2\delta f'(\delta)) \end{aligned}$$

Note that $f'(\delta)$ must be negative, moreover

$$f'(\delta) = -((2 - \ln(\delta + 1/n))(\delta + 1/n))^{-1} > -((2 - \ln(\delta + 1/n))\delta)^{-1}$$

Hence

$$2\delta f'(\delta) > -2(2 - \ln(\delta + 1/n))^{-1} > -2(2 - \ln(1 + 1/n))^{-1} > -1.1$$

for n sufficiently large, therefore

$$E_a(f(P(\tilde{B}_{il}^k))) > f(\delta) + \alpha^{-10}(\ln(4) - 1.1) \geq f(\delta) + \frac{1}{10}\alpha^{-10}$$

With this the proof is complete. □

^{*}One can see this intuitively by looking at the graph of f , in figure 3.4, and noting that if we fix the point $f(E_a(P(\tilde{B}_{il}^k)))$ and consider a value $c = P(\tilde{B}_{il}^k)$ to the left of $E_a(P(\tilde{B}_{il}^k))$ and another value d equidistant but to the right of $E_a(P(\tilde{B}_{il}^k))$ (on the x -axis) then $|f(c) - f(E_a(P(\tilde{B}_{il}^k)))| > |f(d) - f(E_a(P(\tilde{B}_{il}^k)))|$.

[†]Note that $\frac{\delta}{1 - \alpha^{-10}} = \delta \frac{\alpha^{10}}{\alpha^{10} - 1} < \delta \frac{\alpha^{10} + 2}{\alpha^{10}} = \delta(1 + 2\alpha^{-10})$

We can now proceed to our original objective of proving that $E_a(\tilde{L}) > L + C_6$ where C_6 is polynomial on α . Recall that $\tilde{L} = \sum_{i,k} f(\Pr(\tilde{B}_{il}^k))$ where i is not fixed (otherwise $\tilde{B}_{il}^k$ doesn't exist), thus by linearity of expectation

$$E_a(\tilde{L}) = \sum_{i,k} E_a(f(\Pr(\tilde{B}_{il}^k)))$$

Claim 9.

$$E_a(\tilde{L}) > L + ns_0\alpha^{10}$$

Proof. Consider $E_a(f(\Pr(\tilde{B}_{il}^k))) - f(\Pr(B_{il}^k))$ as a function of (i, k) . If (i, k) satisfies Claim 8 then by using it we can lower bound this quantity, if not then if i is not fixed Claim 7 states that this quantity is not negative, and if i is fixed $\tilde{B}_{il}^k$ doesn't exist so we are just left with $-f(\Pr(B_{il}^k))$, yet we can lower bound this by using the known upper bound on f .*

We start by computing how many tuples (i, k) satisfy Claim 8. By Claim 6, $|W| \geq 10^{-2}n\alpha^{-8}$, and for all $k \in W$, $|J_{i_0s_{r_0}}| > (1/8)s_{r_0}\alpha^{-4} - 1$. Yet because $i \in J_{i_0s_{r_0}}^k \cap [i_0 - s_{r_0-1} - 1, i_0 - s_{r_0-1} - 2, \dots, i_0 - s_{r_0}]$ we must have that at least

$$\frac{1}{8}s_{r_0}\alpha^{-4} - 1 - s_{r_0-1} = \frac{1}{8}s_{r_0}\alpha^{-4} - s_0 - 1$$

indices i satisfy Claim 8, for each $k \in W$. Hence Claim 8 holds for at least

$$10^{-2}n\alpha^{-8} \left(\frac{1}{8}s_{r_0}\alpha^{-4} - s_0 - 1 \right)$$

pairs (i, k) . Recall that $s_{r_0} = \alpha_1 s_{r_0-1} = \alpha_1 s_0$, it is then straightforward to conclude that Claim 8 holds for at least $10ns_0\alpha^{25}$ tuples (i, k) , and by Claim 8 for those tuples we have that

$$E_a(f(\Pr(\tilde{B}_{il}^k))) \geq f(\Pr(B_{il}^k)) + \frac{1}{10}\alpha^{-10}$$

For the other tuples with indices not fixed, by 3.6 we have that

$$E_a(f(\Pr(\tilde{B}_{il}^k))) \geq f(\Pr(B_{il}^k))$$

So $E_a(f(\Pr(\tilde{B}_{il}^k))) - f(\Pr(B_{il}^k))$ is not negative. We now take the sum over all (i, k) , let

$$A = \{(i, k) \mid (i, k) \text{ satisfies Claim 8}\}$$

$$B = \{(i, k) \mid (i, k) \text{ doesn't satisfy Claim 8 and } i \text{ is not fixed}\}$$

*Note that we could attempt to additionally use the fact that for i fixed then $i \in [i_0 - 1, \dots, i - s_0]$, which is a big subset of the interval $[i_0 - 1, \dots, i - s_{r_0}]$ where we know that $J_{i_0s_{r_0}}^k$ is large for a significant number of $k \in [n]$, and thus for those tuples (i, k) , $-f(\Pr(B_{il}^k)) > -f(\alpha^{-10})$, and $-f(\alpha^{-10}) > -(\ln(\ln(n)) + 1)$ from the upper bound on f , yet this won't be necessary.

$$C = \{(i, k) \mid (i, k) \text{ doesn't satisfy Claim 8 and } i \text{ is fixed}\}$$

Then

$$\begin{aligned} & \sum_{i,k} (E_a(f(\Pr(\tilde{B}_{il}^k))) - f(\Pr(B_{il}^k))) = \\ &= \sum_{(i,k) \in A} (E_a(f(\Pr(\tilde{B}_{il}^k))) - f(\Pr(B_{il}^k))) + \sum_{(i,k) \in B} (E_a(f(\Pr(\tilde{B}_{il}^k))) - f(\Pr(B_{il}^k))) + \\ &+ \sum_{(i,k) \in C} -f(\Pr(B_{il}^k)) > ns_0\alpha^{15} - s_0n(\ln(\ln(n)) + 1) > ns_0\alpha^{10} \end{aligned}$$

The conclusion follows immediately. $\square$

Now all it's left to do is to lower bound the probability of a particular event a for which $\tilde{L} > L + ns_0\alpha^{10}$, that we know it exists by the last claim. Denote by $\tilde{L}_a$ the value of $\tilde{L}$ resulting from the fixing event a .

Claim 10. *There exists a fixing event a for which*

$$\Pr(a) > n^{-(3+s_0)} \quad L_a > L + \frac{1}{2}ns_0\alpha^{10}$$

Proof. By the bound on f , for all fixings a , $\tilde{L}_a < n\tilde{m}(\ln(\ln(n)) + 1) < n^3$. Denote by γ the probability, over the fixing events a , that $L_a > L + \frac{1}{2}ns_0\alpha^{10}$, then

$$\begin{aligned} L + ns_0\alpha^{10} &< E_a(\tilde{L}) = \sum_a \Pr(a)\tilde{L}_a \leq \\ &\leq \gamma n^3 + (1 - \gamma)(L + \frac{1}{2}ns_0\alpha^{10}) \leq \gamma n^3 + L + \frac{1}{2}ns_0\alpha^{10} \end{aligned}$$

Therefore $\gamma n^3 > \frac{1}{2}ns_0\alpha^{10} > 1$, for n sufficiently large, hence $\gamma > n^{-3}$. But note that

$$\gamma = \sum_{\{a \mid L_a > L + \frac{1}{2}ns_0\alpha^{10}\}} \Pr(a)$$

And trivially $|\{a \mid L_a > L + \frac{1}{2}ns_0\alpha^{10}\}| < n^{s_0}$ (considering a fixing a is a member of $[n]^{s_0}$), so all event can't possibly have $\Pr(a) \leq n^{-(3+s_0)}$ or else we'd get a contradiction, for $\gamma > n^{-3}$. Thus there must exists an event a within $\{a \mid L_a > L + \frac{1}{2}ns_0\alpha^{10}\}$ for which $\Pr(a) > n^{-(3+s_0)}$. $\square$

We are now ready to finalize the proof of *the anti-symmetry lemma*. Remember that correcting l in $\tilde{B}_{il}^k$ when it depends on less than l variables (due to the fixing) can only decrease the size of $\tilde{B}_{il}^k$ and so increase the value of $f(\tilde{B}_{il}^k)$, it is straightforward to verify that this corrections maintains the validity of the last claims.

Fixing a per Claim 10, we get a new scheme with $\tilde{m} = m - s_0$ variables for which the conditions of the lemma still hold and so by assumption so does 3.2. We now let $m = n$, of course we can't have $\tilde{m} = n$ after the fixing process but one can verify that all the claims remain valid for say $\tilde{m} > 3n/4$ and so we can repeat the fixing process a number of times.

To recap we have that, given set Y satisfying the condition of the lemma, and by assumption 3.2, there exists a fixing a producing a new set $\tilde{Y}$ such that $\Pr(a) > n^{-(3+s_0)}$ and $L_a > L + \frac{1}{2}ns_0\alpha^{10}$. So we are now looking for the contradiction resulting from the inductive repetition of this fixing process, namely that after a finite number of fixing steps the potential functions grows larger than its upper bound.

Let $\mathbf{a}$ denote the union of all fixing events a_i for $i \in [t]$, with t the total number of fixing steps, let $\hat{s}$ be the total number of fixed indices and $\tilde{Y}_t$ the set produced after the t steps, so that $\hat{m} = m - \hat{s}$ with $\hat{m}$ the remaining variables after t fixing steps, then by Claim 10

$$\begin{aligned} \tilde{L}_{\mathbf{a}} &> \tilde{L}_{\mathbf{a} \setminus \{a_1\}} + \frac{1}{2}ns_{0_1}\alpha^{10} > \tilde{L}_{\mathbf{a} \setminus \{a_1, a_2\}} + \frac{1}{2}ns_{0_2}\alpha^{10} + \frac{1}{2}ns_{0_1}\alpha^{10} > \dots \\ &\dots > \sum_{i \in [t]} \frac{1}{2}ns_{0_i}\alpha^{10} = \frac{1}{2}n\hat{s}\alpha^{10} \end{aligned}$$

Now because $\tilde{L}_{\mathbf{a}} < n\tilde{m}(\ln(\ln(n)) + 1)$ it's straightforward to see that $\hat{s} < \alpha^{-1}n < n/4$ and because we fix at least $\lg(n)$ variable at each step (recall the definition of the distances s_r) we must have that $t \leq \hat{s}/\lg(n)$, thus after a finite number of steps we obtain the desired contradiction. Moreover note that $\hat{m} > n(1 - \alpha^{-1})$.

Recall now that $|Y| = \prod_i^t \Pr(a_i)^{-1} |\tilde{Y}_t|$, and by Claim 10

$$\Pr(\mathbf{a}) = \Pr(a_1)\Pr(a_2)\dots\Pr(a_t) > n^{-(3+s_{0_1})}n^{-(3+s_{0_2})}\dots n^{-(3+s_{0_t})} = n^{-3t}n^{-\hat{s}} > 2^{-3\hat{s}}n^{-\hat{s}}$$

By the contradiction on the upper bound of the potential function we have that, for $\hat{x}$ distributed uniformly on $\tilde{Y}_t$, $H(\hat{x}) \leq \hat{m}(\lg(n) - \lg(\alpha))$, but on the other hand by the uniform distribution of $\hat{x}$ we have by the entropy bound that $H(\hat{x}) = \lg(|\tilde{Y}_t|)$ thus we have that

$$|\tilde{Y}_t| \leq \left(\frac{n}{\alpha}\right)^{\hat{m}}$$

Therefore

$$|Y| \leq 2^{3\hat{s}}n^{\hat{s}} \left(\frac{n}{\alpha}\right)^{\hat{m}} < 2^{3\hat{s}}\alpha^{\hat{s}} \left(\frac{n}{\alpha}\right)^m$$

Now $(2^3)^{\hat{s}} \alpha^{\hat{s}} < \alpha^{2\hat{s}}$ (for n large enough) and $\hat{s} = m - \hat{m} < m/4$ therefore

$$|Y| < \left(\frac{n}{\sqrt{\alpha}} \right)^m$$

This concludes the proof of the lemma. □

4. Representation theory of finite groups

Representation Theory might appear a bit daunting at first, but its underlying ideas are in fact quite simple, and can be seen as coming from very familiar results.

As Joseph J. Rotman put it "A group is a cloud...", an abstract object composed of a set and rules that dictate how the elements of that set interact with each other, as such it can be quite hard to work with and gain insight into its structure. Much of group theory is dedicated to understanding the general structure of groups, and characterizing them in the simplest of ways, for instance it's known that finite abelian groups are nothing more than a product of cyclic groups, and these factors are uniquely determined up to order. For non-abelian groups, things are not so bright, but there is something we can always say about a group. Let S_n be the *symmetric group*; of permutations on a set with n elements.

Theorem 4.0.1 (Cayley). *Every finite group G can be imbedded as a subgroup of a Symmetric group, in particular if $|G| = n$ then G can be imbedded in S_n .*

The reader should be familiar with this result from elementary group theory and its proof can be seen on any book on the subject, it fundamentally considers the group G as acting on itself by left multiplication thus resulting in any $g \in G$ defining a permutation on the elements of G .

What this tells us is that given any finite group we can always regard it as being a subgroup of a Symmetric group, and so its elements are something we are quite familiar with. The imbedding alluded above gives us the capability to represent any group as something more concrete and that we (hopefully) have more tools to apply in its study.

To this end let us not stop here, although we know a lot about permutations, we can do better. As is known any permutation can be represented by a matrix, aptly named a permutation matrix, and so any group can be regarded as a group of permutations matrices, and this is very good because now we have the whole theory of linear algebra at our disposal to tackle the problem.

Let's put things in perspective, we started with some group, recast it as a group of permutations, that as we know act on a set of, say n elements, and then we recast it again as a group of permutation matrices, that can be seen as acting on vector space of dimension n . So we can ask the question *How does this group act on a vector space?*

Much information can be gained from a group by studying how it acts on suitable structures and vector spaces provide the best examples. So put simply here in representation theory (of finite groups) we'll study how groups act on vector spaces.

With that being said, we start off with some definitions and see where that takes us.

4.1. Representations

4.1.1 Definitions

First and foremost, some notation. We denote by $\mathbb{F}$ a field of characteristic 0, usually we take $\mathbb{F} = \mathbb{C}$ for reasons that will be made clear, but other field can also be considered. Building representation theory upon these fields gives rise to what is called *ordinary representation theory*, if we were to consider finite fields, denoted by $\mathbb{F}_q$ where q is its order, then it would be called *modular representation theory*.

Also, for $n \in \mathbb{N}$ where $\mathbb{N}$ is the set of natural numbers, we'll denote a set of natural numbers $\{1, 2, \dots, n\}$ by $[n]$, this is a nice compact way of referring to such sets.

We'll consider only finite dimensional vector spaces V over a field $\mathbb{F}$ and denote by $\text{GL}(V)$ the group of all invertible linear maps from V to V (do note that this is indeed a group with the operation of composition of maps), this group is called the *general linear group* on V .

We recall a simple fact from linear algebra. If $\dim_{\mathbb{F}}(V) = n < \infty$, we can choose a basis B of V , say $B = (e_1, e_2, \dots, e_n)$ and identify V with $\mathbb{F}^n$, seeing that any element of V is written uniquely as a linear combination of the elements in the basis just map elements of V to the respective list of coefficients and check this gives an isomorphism between V and $\mathbb{F}^n$. With this given any map $\theta \in \text{GL}(V)$ we can construct a matrix $A_\theta \in M_n(\mathbb{F})$, where $M_n(\mathbb{F})$ is the set of all matrices of size $n \times n$ with entries in $\mathbb{F}$, that acts on $\mathbb{F}^n$ the same way θ acts on V . By noting that $\theta(e_j)$ is some element in V written in an unique way in the basis B we have that $A_\theta = (a_{ij})$ is given by:

$$\theta(e_j) = \sum_i a_{ij} e_i$$

It's straightforward to verify than indeed using θ on V or A_θ on $\mathbb{F}^n$ amounts to the same thing, by noting that these are linear maps so θ is uniquely determined by how it acts on the basis B .

Denoting $\text{GL}(\mathbb{F}^n)$ by $\text{GL}_n(\mathbb{F})$ we have that $A_\theta \in \text{GL}_n(\mathbb{F})$ and in fact it's not hard to prove the following.

Proposition 4.1.1. *As groups, $\text{GL}(V) \cong \text{GL}_n(\mathbb{F})$ with the isomorphism given by the map $\theta \mapsto T_\theta$*

Yet do note that choosing a different basis for V produces a different isomorphism, but we have the following fact.

Proposition 4.1.2. *Matrices A_1, A_2 represent the same element of $\text{GL}(V)$, with respect to different basis, if and only if they are conjugate, that is there exists some $B \in \text{GL}_n(\mathbb{F})$ such that*

$$A_1 = BA_2B^{-1}$$

More linear algebra facts we'll use can be seen in Appendix C. We can now proceed to define a representation and in fact we'll define it in several equivalent ways.

Definition 4.1.3 (Representation). *Let V be a finite dimensional vector space over $\mathbb{F}$ and G a finite group. A representation of G on V is a (group) homomorphism*

$$\rho : G \rightarrow \text{GL}(V)$$

We'll usually denote $\rho(g)$ by ρ_g .

Considering ρ is a homomorphism we have the usual properties that $\rho_{gh} = \rho_g\rho_h$ and $\rho_{g^{-1}} = (\rho_g)^{-1}$, moreover for all $g \in G$ we have that $\rho_g \in \text{GL}(V)$ and so ρ_g is in particular a $\mathbb{F}$ -linear map on V , as redundant as it may seem to state this we'll see that this property is in fact highly valuable. And seeing we're going for the obvious it suffices to say that if ρ wasn't a homomorphism we wouldn't have much hopes of learning anything about the structure of G .

Representations as define above are usually called *linear* representations, and it's sometimes useful to make the vector space V upon which we define the representation more evident, by denoting the representation by ρ_V .

You may recall from elementary group theory that, given ρ is a group homomorphism, $\ker(\rho)$ is a normal subgroup of G , thus $G/\ker(\rho)$ is a group and $G/\ker(\rho) \cong \rho(G)$, therefore when $\ker(\rho)$ is the trivial subgroup $\{1_G\}$, that is, composed of only the identity of G , then ρ is one-to-one and $G \cong \rho(G)$, when this happen we give it a name.

Definition 4.1.4. *A faithful representation ρ is a representation where $\ker(\rho) = \{1_G\}$.*

Relating this to Cayley's theorem mentioned previously, here we also get an imbedding of G in $\text{GL}(V)$ and we can see that our definition of a representation is quite more flexible for it needs not to be faithful.

As stated before, we are going to unveil facts about the structure of the group G by studying how it acts on a vector space V , that is, by studying its representations. It is now time to make this statement more concrete.

Note that given a representation ρ_V of G we can use it to define an operation between G and V as

$$\begin{aligned} (\cdot) : G \times V &\rightarrow V \\ (g, v) &\mapsto \rho_g v \end{aligned}$$

Firstly we note that this function is well defined, in the sense that if $(g, v) = (g', v')$ then $g \cdot v = g' \cdot v'$, and $\rho_g v \in V$. Moreover we have that:

- $1_g \cdot v = v$
- $g \cdot (h \cdot v) = (gh) \cdot v$

The first one should be clear and the second comes easily by using the definitions and the fact that ρ is a homomorphism, indeed $g \cdot (h \cdot v) = \rho_g(\rho_h v) = \rho_g \rho_h v = \rho_{gh} v = (gh) \cdot v$. When this happens we say that we have a *Group Action*.

Definition 4.1.5 (Group Action*). *Let G be a group and X a set. We say that G acts on X if there is a map*

$$\begin{aligned} (\cdot) : G \times X &\rightarrow X \\ (g, x) &\mapsto g \cdot x = gx \end{aligned}$$

Such that

- $1_g \cdot x = x$
- $g \cdot (h \cdot x) = (gh) \cdot x$

Under these condition we usually say X is a G -set.

Because ρ_g is an $\mathbb{F}$ –linear map on V , for all $g \in G$, we get a bit more than the above:

- $\rho_g(v_1 + v_2) = \rho_g v_1 + \rho_g v_2$

*Note that a group action defines a homomorphism $G \mapsto S_X$ where S_X is the group of symmetries/symmetric group on X (if $|X| = n$ then $S_x = S_n$), moreover given this homomorphism we can recover the group action thus these two ideas are interchangeable.

- $\rho_g(\lambda v) = \lambda \rho_g(v)$

Definition 4.1.6 (Linear Action). *A group G acts linearly on a vector space V if it acts on V and*

- $g \cdot (v_1 + v_2) = g \cdot v_1 + g \cdot v_2$
- $g \cdot (\lambda v) = \lambda(g \cdot v)$

For all $g \in G$, $v_1, v_2 \in V$ and $\lambda \in \mathbb{F}$.

So given a (linear) representation ρ_V of G we can define a linear action $(\cdot)$ of G on V , going the other way around given a linear action of G on V we can define a representation $\rho_V : G \rightarrow \text{GL}(V)$ by letting

$$\begin{aligned}\rho_g : V &\rightarrow V \\ v &\mapsto g \cdot v\end{aligned}$$

For each $g \in G$. Using the previous definitions it's straightforward to verify that ρ_V is well defined in the sense that $\rho_V(g) \in \text{GL}(V)$ and that it is indeed a (linear) representation. So we can see that representations and linear actions are really two faces of the same coin, i.e. equivalent definitions.

Remark 1. It is interesting to note that, referring to footnote *, while a group action defines a homomorphism $G \mapsto S_X$, and S_X is the group of symmetries of the set X , a linear action defines a homomorphism $G \mapsto \text{GL}(V)$ where $\text{GL}(V)$ can be regarded as the group of symmetries of the vector space V .

Continuing our venture, now that we have a linear action of G on V we can define the following.

Definition 4.1.7. *Let G be a group and V a vector space, then V is called a G -space if there exists a linear action $(\cdot)$ of G on V . Elements of a G -space are of the form*

$$\sum_{g \in G} g v_g$$

with $v_g \in V$. Operations on these elements are given by the ones on V and the linear action $(\cdot)$, thus we have a sum of elements, component wise, as

$$\sum_{g \in G} g v_g + \sum_{g \in G} g w_g = \sum_{g \in G} g(v_g + w_g)$$

and a scalar product by $\mathbb{F}$ and by G . We usually denote a G -space by ${}_G V$ to make it evident that we are referring to a G -space, except when it's clear from context in which case we simply denote it by V .

One way to look at this is that now V has inherited a set of linear maps on it whose structure is given by G .

Note that our linear action allows us to define a left product of elements of G by elements of V (we have mentioned nothing about something like $v \cdot g$) so it would be correct to say that what we have is a *left* G -space.

If the reader is familiar with Module Theory note that considering the additional property $(g + h) \cdot v = g \cdot v + h \cdot v$, that is $(\rho_g + \rho_h)v = \rho_g v + \rho_h v$, which is easy to verify, we could get a module, therefore a G -space is not far from being a module, yet do note that such operation as $+$ on the elements of G has not been defined, we can define it on $\text{GL}(V)$ because it is a ring under composition and addition, but G is just a group, so this is far out of line. If you are not familiar with Module Theory do not despair, we'll bring you up to speed before getting to Chapter 5 on the Okounkov-Vershik Approach, for now start by familiarizing yourself with the definition of a module in Section D.1.2.

For the G -space to be a module it needs to be defined over a ring and not a group, or else the aforementioned missing property that $(g + h) \cdot v = g \cdot v + h \cdot v$ does not make sense. To solve this we first give a definition.

Definition 4.1.8. Let $\mathbb{F}$ be a field and G a finite group of order n , denote the identity of G by 1. The group ring $\mathbb{F}G$ is defined to be the n -dimensional vector space over $\mathbb{F}$ with elements of G as its basis. Elements of $\mathbb{F}G$ can be described as sums of the form

$$\sum_{g \in G} c_g g$$

where $c_g \in \mathbb{F}$. Addition is as usual in a vector space (component wise) and product is given by

$$\left(\sum_{g \in G} a_g g \right) \left(\sum_{h \in G} b_h h \right) = \sum_{k \in G} c_k k$$

where $c_k = \sum_{gh=k} a_g b_h$.

It's straightforward to verify that this is indeed a ring and in fact one can verify that it's in fact an algebra using definition D.2.1, thus one usually refers to $\mathbb{F}G$ as a *Group Algebra*.

Going back to our (left) G -space V , recall that it is equipped with a linear action, hence $g \cdot (\lambda v) = \lambda(g \cdot v)$. We claim that ${}_G V$ is equal to V as a left module over the group algebra

$\mathbb{F}G$ (denoted $\mathbb{F}_G V$) with left action given by extending the action of G on V by linearity to $\mathbb{F}G$. Indeed it is clear any element of ${}_G V$ is in $\mathbb{F}_G V$ and using the fact that $\lambda(g \cdot v) = g \cdot (\lambda v)$ we have the other inclusion, moreover the action of G coincides in both cases.

Using a similar argument to that provided for the equivalence of the definitions of linear actions and representations one can conclude the equivalence of a representation over V and the respective G -space/ $\mathbb{F}G$ -module V . Considering our discussion about studying the group G through its actions on a vector space V it should be clear that we are very much interested in G -spaces/ $\mathbb{F}G$ -modules. In fact much of the representation theory of finite groups can be developed from module theory, and it was Emmy Noether, in 1929, that brought the connection between these two theories to light. As such in practice we'll be more focused on $\mathbb{F}G$ -modules than G -spaces even though G -spaces have a more natural link to representations.

Recalling Proposition 4.1.1 we can give another equivalent definition of a representation, after all if we can represent a linear map by something as concrete as a matrix then we should certainly take advantage of that.

Definition 4.1.9 (Matrix Representation). $R : G \rightarrow \text{GL}_n(\mathbb{F})$ is a matrix representation of G of degree n if R is a (group) homomorphism.

It is clear, by Proposition 4.1.1, that given a representation ρ of G over a vector space V of dimension n , by fixing a basis B of V , we can produce the matrix representation R defined above, on the other hand any matrix representation is in particular a representation as in Definition 4.1.3. This latter definition gives rise to another definition.

Definition 4.1.10. The dimension/degree of a representation ρ_V is the dimension of the vector space V , i.e. $\dim_{\mathbb{F}}(V)$.

When we consider V as an inner product space, that is a vector space equipped with an inner product $\langle \cdot, \cdot \rangle_V$ as defined in C.0.1, then referring to the definitions of unitary and adjoint map as in C.0.9 and C.0.6 respectively, we can consider a special type of representations.

Definition 4.1.11. A unitary representation ρ_V is one that, for all $g \in G$, ρ_g is an unitary operator.

The advantage of this is that now $\rho_{g^{-1}} = \rho_g^{-1} = \rho_g^*$ where ρ_g^* is the adjoint of ρ_g , thus for $v, w \in V$, $\langle \rho_g v, w \rangle_V = \langle v, \rho_{g^{-1}} w \rangle_V$.

In fact, there is an inner product we can always define on an inner product space to make a representation unitary.

Theorem 4.1.12 (Weyl's unitarity trick). *Given a vector space V with inner product $\langle \cdot, \cdot \rangle_0$ and a representation ρ_V of a (finite) group G we can construct an inner product $\langle \cdot, \cdot \rangle$ on V such that ρ_V is unitary.*

Proof. Consider the inner product $\langle \cdot, \cdot \rangle$ on V defined by

$$\langle v, w \rangle = \sum_{g \in G} \langle \rho_g v, \rho_g w \rangle_0$$

It's straightforward to prove that this is indeed an inner product on V , moreover for all $h \in G$ we have that

$$\langle \rho_h v, \rho_h w \rangle = \sum_{g \in G} \langle \rho_{hg} v, \rho_{hg} w \rangle_0 = \sum_{g' \in G} \langle \rho_{g'} v, \rho_{g'} w \rangle_0 = \langle v, w \rangle$$

Therefore ρ_g is unitary for all $g \in G$ and so ρ is unitary. □

When an inner product behaves like the above for the action of G , that is $\langle gv, gw \rangle = \langle v, w \rangle$ we say that the inner product is G -invariant. As a consequence of this theorem, when we are working with a particular representation, we can always consider it to be unitary as long as we use a G -invariant inner product like the aforementioned.

When working over $\mathbb{C}$ the existence of this inner product is guaranteed by the existence of an inner product on V as can be seen in Appendix A, definition C.0.2 and so when working over $\mathbb{C}$ we always consider the representations to be unitary.

4.1.2 Examples

The simplest representation one can conceive is aptly named the *trivial representation*

Definition 4.1.13 (Trivial Representation). *For a group G , let $V = \mathbb{F}$ (this makes sense as $\mathbb{F}$ is a field so in particular can be seen as a vector space over itself) the trivial representation of G is $\iota_G : G \rightarrow \text{GL}(\mathbb{F})$ where each $g \in G$ maps to the identity in $\text{GL}(\mathbb{F})$, as such one can say that for all $g \in G$, $\iota_G(g) = 1_{\mathbb{F}}$ (where $1_{\mathbb{F}}$ is the multiplicative identity in $\mathbb{F}$), seeing that for all $x \in \mathbb{F}$, $1_{\mathbb{F}} x = x$.*

Although quite innocent looking this representation is useful in far from trivial scenarios, as we shall see.

Definition 4.1.14. *Given a representation ρ_V of a group G , so with respective G -space V , and a sub-group H of G we can always construct a representation of H by restricting ρ_V to H , this is denoted by $\rho|_H^G$ or $V|_H^G$. More simply we write $\rho|_H$ or $V|_H$ when the original group G is clear from context.*

We'll be spending a considerable amount of time in a later chapter delving in to the representation theory of the symmetric group S_n , which is quite rich, and intricate. Here we preset a couple examples of representations of S_n .

Example 4.1.15. Consider the symmetric group S_n and recall that given $\pi \in S_n$ we say that π is even/odd if it's a composition of an even/odd number of transpositions (cycles of length 2). The *alternating representation* of S_n is defined over $\mathbb{F}$, just as in the trivial representation, as $\epsilon : S_n \rightarrow \mathbb{F}$ by

$$\epsilon(\pi) = \begin{cases} 1 & \text{if } \pi \text{ is even} \\ -1 & \text{if } \pi \text{ is odd} \end{cases}$$

For each $\pi \in S_n$.

Example 4.1.16. Let V be an n dimensional vector space and fix a basis $B = (e_1, e_2, \dots, e_n)$ of V . Considering S_n acts naturally on the set $\{1, 2, \dots, n\}$ we can define a representation $\rho : S_n \rightarrow GL(V)$ by, for $\pi \in S_n$, setting $\rho_\pi(e_i) = e_{\pi(i)}$ and extend this by linearity to any element $v \in V$. This is called the *permutation representation* of S_n .

This last example can be generalized to any group G acting on a set X .

Definition 4.1.17 (Permutation representation). Consider a group action of G on a finite set X and a field $\mathbb{F}$. Let $\mathbb{F}X$ be, similar to the group ring, a $|X|$ dimensional vector space over $\mathbb{F}$ with basis the elements of X , so its elements are of the form

$$\sum_{x \in X} a_x x$$

With G action given by

$$g \sum_{x \in X} a_x x = \sum_{x \in X} a_x gx$$

One can verify that this determines a linear action of G on $\mathbb{F}X$. The corresponding representation $\lambda : G \rightarrow GL(\mathbb{F}X)$ is called the *permutation representation* of G .*

Remark 2. Given a group action of G on a set X let $x \in X$, we define the *stabilizer* of x under the action of G as $G_x = \{g \in G \mid gx = x\}$. Let then $x \in X$, $K = G_x$ and consider the set of left cosets G/K . If the group action is transitive, that is for any $x, y \in X$ there exists

*We could, just as in example 4.1.16, define a vector space V over $\mathbb{F}$ with a basis **indexed** by the elements of X , i.e., letting $X = \{x_1, x_2, \dots, x_n\}$, we'd use the basis $B = \{e_{x_1}, e_{x_2}, \dots, e_{x_n}\}$ and then the *permutation representation* λ is given by letting $\lambda_g(e_{x_i}) = e_{gx_i}$, for each $g \in G$, and extending this by linearity to any element in V . This is equivalent to the given definition.

$g \in G$ such that $gx = y$, then we can define the map $\theta : X \rightarrow G/K$ by $y \mapsto sK$ such that $sx = y$. The map is well defined as $s_1x = s_2x \Leftrightarrow s_2^{-1}s_1x = x \Leftrightarrow s_2^{-1}s_1 \in K \Leftrightarrow s_2K = s_1K$ and because these are all equivalences we also get that the map is one-to-one. That is onto also follows from the previous argument considering that, for $sK \in G/K$ letting $y = sx$ then $\theta(y) = sK$ and so we have a bijection. Now the interesting part is that, for $g \in G$ and $y \in X$, $\theta(gy) = g\theta(y)$ and so θ preserves the action of G on X and so working with G , X and its group action is the same as working with G , G/K and the operation defined on G . As such, if the group action is transitive, we can choose any $x \in X$, define K , and in the above definition replace X with G/K and just work with the operation on G .

If in the above definition we let $X = G$ and let the action of G on $\mathbb{F}G$ be the determined by the operation defined on the group G then we get one of the most, if not the most, interesting representation of a finite group G . We'll become quite familiar with it; the definition follows.

Definition 4.1.18. *The regular representation of G is a special case of the permutation representation where we let X be G and consider the natural action of G on $\mathbb{F}G$ (by left multiplication). The resulting left module ${}_{\mathbb{F}G}\mathbb{F}G$, i.e. the module $\mathbb{F}G$ over itself, simply denoted $\mathbb{F}G$ when it's clear from context we are referring to it as a module, is called the regular module.*

Going back to the symmetric group we have the following example of a regular representation.

Example 4.1.19. Consider the symmetric group S_n and the respective group algebra $\mathbb{F}S_n$. The regular representation of S_n , $\lambda : S_n \rightarrow \text{GL}(\mathbb{F}S_n)$ is defined by

$$\lambda_\pi\left(\sum_{\eta \in S_n} a_\eta \eta\right) = \sum_{\eta \in S_n} a_\eta (\pi\eta)$$

For each $\pi \in S_n$.

Note that a permutation representation (and consequently its special case the regular representation) simply permutes the basis elements of the vector space over which it's defined, justifying its name of course, but more importantly telling us that its respective matrix representation maps elements of the group G to permutation matrices.

Example 4.1.20. Consider the symmetric group $S_2 = \{1_{S_2}, (1\ 2)\}$ and its natural action on $X = \{1, 2\}$. Let V be a two dimension vector space with basis indexed by X , so

$B = \{e_1, e_2\}$. Then because the associated permutation representation λ is a group homomorphism we must have that λ_{1s_2} is the identity map on V , that is the identity in $GL(V)$ and for the permutation $(1\ 2)$ we have that

$$\lambda_{(1\ 2)}(e_1) = e_2$$

$$\lambda_{(1\ 2)}(e_2) = e_1$$

Therefore, if we fix the canonical basis of $\mathbb{F}^2$, $B' = ((1, 0), (0, 1))$, the matrix representation associated with λ and the basis B' is

$$\lambda_{1s_2} \equiv \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\lambda_{(1\ 2)} \equiv \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

The next example raises the questions to be answered in the next section. It's now a good time for the reader to get familiar with the definition of a homomorphism between modules, Appendix B.1 definition D.1.6.

Example 4.1.21. Consider the cyclic group of order 4, $C_4 = \langle a \rangle$ and a two dimensional vector space V . Let ρ_1, ρ_2 be two representation of C_4 over V defined by

$$\rho_1(a)e_1 = -e_2 ; \quad \rho_1(a)e_2 = e_1$$

$$\rho_2(a)e_1 = e_2 ; \quad \rho_2(a)e_2 = -e_1$$

Over the canonical basis these representation have corresponding matrix representation given by:

$$\rho_1(a) \equiv \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad \rho_2(a) \equiv \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

Note that because C_4 is a cyclic group generated by a and representations are homomorphisms they are uniquely determined by a .

Seemingly these are two different representations, yet if we consider the respective $\mathbb{F}C_4$ -modules V it's easy to find an $\mathbb{F}C_4$ -isomorphism between them; define $\theta : V \rightarrow V$ by $\theta(e_1) = -e_2, \theta(e_2) = -e_1$ and extend this linearly to V , it is clear this is an automorphism on V , moreover it's easy to see that $\theta(ae_1) = a\theta(e_1)$ and $\theta(ae_2) = a\theta(e_2)$ which suffices to conclude that indeed we have an $\mathbb{F}C_4$ -isomorphism. But if the modules are isomorphic

then the representations underlying them must also be in some way "equivalent".

4.1.3 G -homomorphisms and equivalent representations

When studying algebraic structures it's fundamental to be able to find and exploit similarities between any two, and this is done by finding "morphisms" that somehow preserve said structure. Going back to our view of representations as G -spaces, we are interested in morphism from a G -space V to a G -space W that, first preserve the vector space structure and second preserve the action of G .

Definition 4.1.22 (G -Homomorphism). *Let V, W be G -spaces over $\mathbb{F}$. A function $\theta : V \rightarrow W$ is called a G -Homomorphism, from ${}_G V$ to ${}_G W$, if it is an $\mathbb{F}$ -linear map from V to W and for all $g \in G$ and $v \in V$*

$$\theta(gv) = g\theta(v)$$

We denote the set of all such maps by $\text{Hom}_G(V, W)$. Note that this set can be seen as a vector space over $\mathbb{F}$ when equipped with a scalar product and addition.

If θ is bijective then we have a G -isomorphism and consequently the G -spaces are said to be isomorphic.

Note that a G -homomorphism also satisfies the definition of a homomorphism between modules, as should be expected. Thus all said above applies to $\mathbb{F}G$ -modules and vice-versa for a homomorphism between $\mathbb{F}G$ -modules V and W must be in particular $\mathbb{F}$ -linear.

What this means for representations is that, given two representations of G , ρ_V , σ_W , and a G -homomorphism $\theta : V \rightarrow W$, for all $g \in G$

$$\theta(\rho_g v) = \sigma_g \theta(v)$$

This can be seen in a more visual way by observing the following diagram.

$$\begin{array}{ccc} V & \xrightarrow{\rho_g} & V \\ \downarrow \theta & & \downarrow \theta \\ W & \xrightarrow{\sigma_g} & W \end{array}$$

The relation described above simply means that the diagram *commutes*, that is it does not matter what path we choose in the diagram to go from V on the top left to W on the bottom right.

In this setting we say that the map θ *intertwines* ρ_V and σ_W , and we usually refer to $\text{Hom}_G(V, W)$

as $\text{Hom}_G(\rho_V, \sigma_W)$ (or even $\text{Hom}_G(\rho, \sigma)$) when we want to make the representations we're referring to evident.

Definition 4.1.23. Two representations ρ_V, σ_W of a group G are said to be equivalent, denoted $\rho_V \equiv \sigma_W$, if there exists a G -isomorphism between their respective G -spaces/ $\mathbb{F}G$ -modules. In other words there exists a vector space isomorphism $\theta : V \rightarrow W$ such that $\theta(\rho_g v) = \sigma_g \theta(v)$.

Proposition 4.1.24. Equivalence of representations is an equivalence relation on the set of all representations of a group G .

The proof of the above proposition is routine and left to the reader.

Note that because a G -homomorphism $\theta : V \rightarrow W$ is a linear map, upon fixing a basis for V and W , θ can be described by a matrix and so we can still live comfortably in the world of matrix representations. Moreover if θ is a G -isomorphism then V and W are isomorphic as vector spaces, thus they have the same dimension and the matrix of θ is square and non singular (it being in particular a vector space isomorphism). Consequently, if R and S are the implied matrix representation in the respective chosen basis of V and W and T is the matrix of θ for those same basis, we have that

$$R_g = T^{-1} S_g T$$

Recalling Proposition 4.1.2 we can conclude that equivalent matrix representation are constituted by the same linear maps but operating on different basis of the same vector space.

Disambiguation. The symbol $\equiv$ will also be used when we want to say that two linear maps are the same but operating on different, yet of course isomorphic, vector spaces. This is then consistent with the use for representations it's just the case that we don't restrict the use of the symbol just for representations. Nonetheless the intended use will be clear from context.

Bearing in mind all the above, particularly that equivalent representations produce isomorphic modules, we generally don't distinguish between the two, unless stated otherwise.

Example 4.1.25. Let ρ_V be a representation of a group G . Let $B = (e_1, \dots, e_n)$ be a basis of V and R the matrix representation of ρ_V with respect to the basis B . We know of a vector space isomorphism θ between V and F^n , simply let, for $v \in V$, $\theta(v) = (\alpha_1, \dots, \alpha_n)$ such that $v = \sum_{j=1}^n \alpha_j e_j$. If this map preserves the action of G then we can conclude that

ρ_V is equivalent to R , as we expect. Indeed we have, letting $R_g = (a_{i,j})_{i,j \in [n]}$ such that $\rho_g e_j = \sum_{i=1}^n a_{i,j} e_i$, that

$$\theta(\rho_g v) = \sum_{j=1}^n \alpha_j \theta(\rho_g e_j) = \left(\sum_{j=1}^n \alpha_j a_{1,j}, \dots, \sum_{j=1}^n \alpha_j a_{n,j} \right) = R_g \theta(v)$$

Thus the conclusion follows.

4.1.4 G -subspaces and direct sums

Considering a G -space V , it's not hard to imagine what a G -subspace W of V should be. It must be a subset of V that is a G -space in it's own right when considering the restriction of the linear action of G on V to W .

Definition 4.1.26. Given a G -space V , a G -subspace W of V is a subspace of V that is closed for the action of G , that is $gw \in W$ for all $g \in G$ and $w \in W$.

This of course is consistent with definition D.1.4 of a submodule. The concepts of a G -subspace gives rise to the idea of a *subrepresentation*, so to this end let's see what a G -subspace means in terms of representations.

Theorem 4.1.27. Let ρ_V be a representation of G and W a subspace of V . W is a G -subspace of ${}_G V$ if and only if the restriction of ρ_V to W , that is restricting each $\rho_g : V \rightarrow V$ to the subspace W (denoted $\rho_g|_W$), is a representation $\rho_W : G \rightarrow \text{GL}(W)$.

Proof. If W is a G -subspace of ${}_G V$ then $\rho_g w \in W$ for all $g \in G$ and $w \in W$, thus it is clear that $\rho_g|_W \in \text{GL}(W)$ for all $g \in G$, moreover $\rho_g|_W \rho_h|_W = (\rho_g \rho_h)|_W = \rho_{gh}|_W$ and clearly $\rho_{1_G} = id_W$ so we have a homomorphism $\rho_W : G \rightarrow \text{GL}(W)$. The other direction follows immediately considering that W is then a subspace of V closed for the action of G . $\square$

Definition 4.1.28. When the condition of theorem 4.1.27 are met we say that ρ_W is a *subrepresentation* of ρ_V , and that W is a G -invariant subspace of V .

When thinking in terms of matrix representations the above results in a neat observation.

Proposition 4.1.29. *Let ρ_V be a representation of G and W a G -subspace of ${}_G V$. If $B = \{e_1, \dots, e_n\}$ is a basis of V containing a basis $B' = \{e_1, \dots, e_m\}$ of W , with $0 < m < n$, then the matrix representation of ρ_g , for $g \in G$, with respect to the basis B has the block upper triangular form*

$$\rho_g \equiv \begin{bmatrix} P_g & Q_g \\ \mathbf{0} & T_g \end{bmatrix}$$

Where $P_g \in \mathbf{GL}_n(\mathbb{F})$, $Q_g \in M_{n \times (n-m)}(\mathbb{F})$ and $T_g \in M_{(n-m) \times (n-m)}(\mathbb{F})$. Moreover P_g , for $g \in G$, give the matrix representation of ρ_W with respect to the basis B' .

Ideally we'd like the block matrix above to be block diagonal, yet there's no reason *a priori* for this to be the case, as it would imply that the complement of W in V is also G -invariant.

A G -space is always a G -subspace of itself, along with $\{0\}_G$; these are the trivial examples. When these are the only G -subspaces of ${}_G V$ we have the following.

Definition 4.1.30. *A G -space is simple or irreducible if it has no non-trivial G -subspaces. In this case we also say that the respective representation is simple or irreducible. We denote the set of all irreducible representations of G by $\text{Irr}(G)$.*

By Proposition 4.1.24 we define $\widehat{G}$ to be the set of equivalence classes of irreducible representation of G . We think of this as a set of fixed non equivalent irreducible representations of G , in the sense of choosing a set of representatives of each class.

Because Definition 4.1.26 is consistent with definition D.1.4 of a submodule we can conclude that a simple G -space is a simple $\mathbb{F}G$ -module, and clearly if a G -space is not simple then the corresponding $\mathbb{F}G$ -module is not simple, so we have an equivalence. Thus considering proposition D.1.9 we have the following.

Theorem 4.1.31. *${}_G V$ is simple if and only if $\mathbb{F}G V = \mathbb{F}G v$ for any $v \in V$, that is for any $v, w \in V$ there exists $\alpha \in \mathbb{F}G$ such that $\alpha v = w$.*

Proof. This follows immediately from proposition D.1.9. □

Thus a simple G -space is just a cyclic $\mathbb{F}G$ -module. We could just as in module theory define the G -subspace of ${}_G V$ generated by some element $v \in V$ but because a G -subspace needs to be in particular a subspace of V then we have to define it as the **space generated by** $Gv = \{w \in V \mid w = gv \text{ for some } g \in G\}$, yet this coincides with the cyclic $\mathbb{F}G$ -module $\mathbb{F}Gv$.

From this observation we can conclude that a 1-dimensional representation is always irreducible, as it determines a cyclic $\mathbb{F}G$ -module, or equivalently a G -space ${}_G V$ where V is a one dimensional vector space over $\mathbb{F}$. The following example shows that the converse is not necessarily true.

Example 4.1.32. Consider the symmetric group S_n and the permutation representation determined by the natural action of S_n on $X = \{1, 2, \dots, n\}$. Let $\mathbb{F} = \mathbb{C}$ and λ be the associated matrix representation with respect to the canonical basis. Thus, given $v \in \mathbb{C}^n$ and $\pi \in S_n$, λ_π simply permuted the entries of v according to the permutation π , that is letting $v = (v_1, v_2, \dots, v_n)$, then $\lambda_\pi v = (v_{\pi(1)}, v_{\pi(2)}, \dots, v_{\pi(n)})$.

With this it's clear that $W = \{(t, t, \dots, t) \in \mathbb{C}^n \mid t \in \mathbb{C}\} = \mathbb{C}(1, 1, \dots, 1)$ and $U = \{v \in \mathbb{C}^n \mid \sum_{i=1}^n v_i = 0\}$ are G -invariant subspaces of V . Moreover because W is one dimensional we have that $\lambda|_W$ (the restriction of λ to the subspace W) is an irreducible subrepresentation of λ .

For U note that $B = \{(1, -1, 0, \dots, 0), (1, 0, -1, 0, \dots, 0), \dots, (1, 0, 0, \dots, -1)\}$ is a basis of this space. We're going to show that for any non-zero vector $v \in U$, the module generated by v , $\mathbb{C}S_n v$ contains B , thus $\mathbb{C}S_n v = \mathbb{C}S_n U$, that is $\mathbb{C}S_n v$ is the whole S_n -space U , and so U is simple.

Let then $v = (v_1, v_2, \dots, v_n) \in U$ be non-zero, hence for some $i \neq j$, $v_i \neq v_j$. We can easily find a permutation π in S_n that simply sends i to 1 and j to 2, thus $\pi v = (v_i, v_j, v'_3, \dots, v'_n) \in U$, now consider the permutation $(1\ 2)$ and note that

$$\begin{aligned} \pi v - (1\ 2)\pi v &= (v_i, v_j, v'_3, \dots, v'_n) - (v_j, v_i, v'_3, \dots, v'_n) = (v_i - v_j, v_j - v_i, 0, \dots, 0) = \\ &= (v_i - v_j)(1, -1, 0, \dots, 0) \end{aligned}$$

As $(v_i - v_j) \neq 0$ by assumption we can divide the above by $(v_i - v_j)$ to get the basis vector of U , $(1, -1, 0, \dots, 0)$ and now it's only a matter of applying permutations of the type $(2\ i)$ for $i \in \{3, \dots, n\}$ to it to get all the basis vectors of U , so in conclusion U is simple/irreducible.

There's something else, something very important that we need to pay attention to in the above example. First note that $B' = B \cup \{(1, 1, \dots, 1)\}$ is basis of $\mathbb{C}^n$ (where $\{(1, 1, \dots, 1)\}$ is the basis of W), this is easily seen by noting the we have $n = \dim_{\mathbb{C}}(\mathbb{C}^n)$ elements and with them we can generate any element of the canonical basis, thus any element of $\mathbb{C}^n$. So any element of $\mathbb{C}^n$ can be written uniquely as a sum of these elements, moreover because W and U are S_n -invariant, any element of ${}_S \mathbb{C}^n$ can be written uniquely as a sum of an

element of $_{S_n}W$ and an element of $_{S_n}U$. Indeed, if it wasn't unique then for some $v \in _{S_n}\mathbb{C}^n$ we'd have that $v = g_1w + g_2u = h_1w' + h_2u'$ with $(g_1w, g_2u) \neq (h_1w', h_2u')$, but then $g_1w = h_1w' + (h_2u' - g_2u)$ therefore either $g_1w \notin W$ and so we get the contradiction that W is not S_n -invariant, or we get a contradiction of the assumption $(g_1w, g_2u) \neq (h_1w', h_2u')$. In conclusion we have that $_{S_n}\mathbb{C}^n = _{S_n}W + _{S_n}U$ and $_{S_n}W \cap _{S_n}U = \{0\}$, and this is easily seen to be equivalent to the statement that $_{S_n}\mathbb{C}^n$ can be written uniquely as a sum of an element of $_{S_n}W$ and an element of $_{S_n}U$. Note that with this we can easily produce an S_n -isomorphism $_{S_n}\mathbb{C}^n \rightarrow _{S_n}W \times _GU$ by simply letting $v \mapsto (g_1w, g_2u)$ and letting S_n act on $W \times U$ by $g(w, u) = (gw, gu)$. This shouldn't be completely new to the reader as it is just an extension of the concept of direct sums of vector spaces, extended to the concept of a G -space. All this remains true for the implied $\mathbb{C}S_n$ -modules and so we refer the reader to Appendix D.1 for the general definition and properties of direct sums of modules.

Definition 4.1.33. A G -space V is decomposable if there are proper G -invariant subspaces W, U of V such that $V = W \oplus U$. If no such decomposition exists we say that $_GV$ is indecomposable.

It should be clear that a simple G -space is necessarily indecomposable. As we shall see in Section 4.2 it turns out that, over a field of characteristic 0, the converse is also true.

Definition 4.1.34. Let ρ_V and σ_W be representations of G . The direct sum of ρ_V and σ_W is the representation

$$\rho \oplus \sigma : G \rightarrow \text{GL}(V \oplus W)$$

Given by

$$(\rho \oplus \sigma)(g)(v, w) = (\rho_g v, \sigma_g w)$$

If the direct sum is internal then we could instead say that $(\rho \oplus \sigma)(g)(v + w) = \rho_g v + \sigma_g w$.

Considering the above, if a G -space V is decomposable, with $V = W \oplus U$, then we have that $\rho_V \equiv \rho_W \oplus \rho_U$. Now if we again consider matrix representations we have the following.

Proposition 4.1.35. Let ρ_V be a decomposable representation of G , with $V = W \oplus U$; let B_1 and B_2 be basis of W and U respectively and let $B = B_1 \cup B_2$ be the corresponding basis of V , then

$$\rho_g \equiv \begin{bmatrix} R_g & \mathbf{0} \\ \mathbf{0} & S_g \end{bmatrix}$$

Where R and S are the matrix representations of ρ_W and ρ_U respectively in the basis B_1 and B_2 .

Example 4.1.36. Continuing example 4.1.32, for $n = 2$, we have that $W = \mathbb{C}(1, 1)$ and $U = \mathbb{C}(1, -1)$. Let $B = ((1, 1), (1, -1))$ be a basis of $\mathbb{C}^2$. It is clear that λ_W acts as the identity on W as $\lambda_g|_W(1, 1) = (1, 1)$ for $g \in S_2$. For λ_U we have trivially that $\lambda_{1_{S_2}}|_U(1, -1) = (1, -1)$, and $\lambda_{(12)}|_U(1, -1) = (-1, 1)$, therefore we can conclude that, in the basis B

$$\lambda_{1_{S_2}} \equiv \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \lambda_{(12)} \equiv \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

The interesting thing to note here is that λ_W coincides with the trivial representation of S_2 and λ_U coincides with the alternating representation of S_2 , given in example 4.1.15. Therefore we have that, for S_2 and its natural action on $X = \{1, 2\}$, $\lambda \equiv \iota \oplus \epsilon$ and this is a decomposition into a direct sum of irreducible representations, or equivalently, the respective $\mathbb{C}S_2$ -module decomposes into a direct sum of simple modules.

Note also that the change of basis matrix from the canonical basis to the basis B defines the map required to prove the equivalence between this representation and the representation given in example 4.1.20, as expected.

4.1.5 The dual representation

Let V, W be vector spaces over $\mathbb{F}$, we denote the set of all linear maps from V to W by $\text{Hom}_{\mathbb{F}}(V, W)$. Note that this set is itself a vector space over $\mathbb{F}$. When $W = \mathbb{F}$ we get the vector space $\text{Hom}_{\mathbb{F}}(V, \mathbb{F})$.

Definition 4.1.37. Let V be a vector space over $\mathbb{F}$. The dual space of V , denoted V^* , is the vector space $\text{Hom}_{\mathbb{F}}(V, \mathbb{F})$.

It's an elementary result in linear algebra that $V \cong V^*$, and it's proof can be seen on most introductory book on the subject.

If V is a vector space over $\mathbb{C}$ then, as can be seen definition C.0.2, we can define an inner product on it. As a result we have the following theorem.

Theorem 4.1.38 (Riesz representation theorem). Let V be a vector space over $\mathbb{C}$ with inner product $\langle \cdot, \cdot \rangle_V$ and V^* its respective dual space. Then the map $\Phi : V \rightarrow V^*$ given by $v \mapsto \theta_v$ where $\theta_v(w) = \langle w, v \rangle_V$ is a bijection.

Proof. It's clear that $\theta_v \in V^*$ for all $v \in V$, moreover if $\theta_v = \theta_w$ then for an orthonormal basis $B = \{e_1, \dots, e_n\}$ of V we have $\theta_v(e_i) = \theta_w(e_i)$ for all i and so we must have, by the

definition of these maps, that $v = w$. Thus the map is one-to-one, to see that it's onto note that, given $f \in V^*$ and $w \in V$, because f is a linear map then f is uniquely determined by its image on the basis B . If $v \in V$ is such that $\langle e_i, v \rangle_V = f(e_i)$ for all i then it is clear that $\langle v, w \rangle_V = f(w)$, completing the proof. The existence of v is straightforward to assert, just let $v = \overline{f(e_1)}e_1 + \dots + \overline{f(e_n)}e_n$ where $\overline{f(e_i)}$ is the complex conjugate of $f(e_i)$. $\square$

The map Φ given above is not a vector space isomorphism, because it's not linear, in fact it's what we call antilinear, in the sense that $\Phi(\alpha v + \beta w) = \overline{\alpha}\Phi(v) + \overline{\beta}\Phi(w)$, yet it gives us a very strong association between elements of V and V^* , telling us that every element of V^* is of the form θ_v for some $v \in V$. Moreover, if $B = \{e_1, \dots, e_n\}$ is an orthonormal basis of V then $B^* = \{\theta_{e_1}, \dots, \theta_{e_n}\}$ is a basis of V^* , indeed given $f \in V^*$ and $v = \alpha_1 e_1 + \dots + \alpha_n e_n \in V$

$$f(v) = \alpha_1 f(e_1) + \dots + \alpha_n f(e_n) = f(e_1)\theta_{e_1}(v) + \dots + f(e_n)\theta_{e_n}(v)$$

So B^* generates V^* and because we have n elements and $\dim_{\mathbb{F}}(V^*) = \dim_{\mathbb{F}}(V) = n$, by $V \cong V^*$, then B^* is a basis of V^* . (actually this is, quite likely, part of the argument you'd see in a proof of $V \cong V^*$)

More generally if V is an inner product space over $\mathbb{F}$ then we define an inner product on V^* by

$$\langle \theta_v, \theta_w \rangle_{V^*} = \langle w, v \rangle_V$$

Note that if the inner product on V is G -invariant then so is the resulting inner product on V^* , also if $B = \{e_1, \dots, e_n\}$ is an orthonormal basis of V then $B^* = \{\theta_{e_1}, \dots, \theta_{e_n}\}$ is also orthonormal.

What we want here is to define a representation over the dual space V^* from a given representation over V , thus producing a concept of *dual* representation. In other words we want to use the linear action of G on V to define a linear action of G on V^* . To this end, recalling $V^* = \text{Hom}_{\mathbb{F}}(V, \mathbb{F})$, we can do the following.

Definition 4.1.39. Let V be a vector space over $\mathbb{F}$ and ρ_V a representation of G . The dual/adjoint representation ρ° of G on V^* is given by

$$[\rho_g^\circ f](v) = f(\rho_{g^{-1}} v)$$

For $g \in G$, $f \in V^*$ and $v \in V$.

One might wonder why the g gets inverted, but this becomes evident once we set out to prove that ρ° is indeed a representation of G . It is clear that $\rho_{1_G}^\circ f = f$ for all $f \in V^*$ thus $\rho_{1_G}^\circ$ is the identity on $GL(V^*)$, moreover

$$[\rho_{gh}^\circ f](v) = f(\rho_{h^{-1}}\rho_{g^{-1}}v) = [\rho_h^\circ f](\rho_{g^{-1}}v) = [\rho_g^\circ[\rho_h^\circ f]](v)$$

From this it's evident why the g must be inverted. In conclusion we have the $\mathbb{F}G$ -module V^* , but how does it relate to the $\mathbb{F}G$ -module V ?

Consider again V^* when V is a vector space over $\mathbb{C}$. We have that

$$[\rho_g^\circ \theta_w](v) = \theta_w(\rho_{g^{-1}}v) = \langle \rho_{g^{-1}}v, w \rangle_V$$

If we consider the inner product on V given by Theorem 4.1.12 then we have that

$$\langle \rho_{g^{-1}}v, w \rangle_V = \langle v, \rho_g w \rangle_V = \theta_{\rho_g w}(v)$$

therefore $\rho_g^\circ \theta_w = \theta_{\rho_g w}$.

By noting that W is a subspace of V if and only if $\{\theta_w \mid w \in W\}$ is a subspace of V^* (because $\theta_{\alpha v + \beta w} = \bar{\alpha}\theta_v + \bar{\beta}\theta_w$) we can conclude that ρ° is irreducible if and only if ρ is irreducible. Thus we have a one to one correspondence between simple $\mathbb{C}G$ -modules V and simple $\mathbb{C}G$ -modules V^* .

Definition 4.1.40. A representation ρ_V (where V is over a field $\mathbb{F}$) of a group G is self-dual if $V \cong V^*$ as G -spaces.

Recall that we always have that $V \cong V^*$ as vector spaces, but this is not necessarily true if we require the isomorphism to preserve the action of G , that is be a G -isomorphism. Note that in this case the two $\mathbb{F}G$ modules are isomorphic and so the implied representations are equivalent.

4.1.6 Tensor products

We've already seen a bit about representations of a finite group, but what if we now combine two groups by considering their cartesian product, that is $G \times H = \{(g, h) \mid g \in G, h \in H\}$ for groups G, H with component wise operation given by each group respectively. This is of course still a group but how do its representations relate to the representations of each group independently, if they relate at all.

As it turns out they are very much related, and the ideal notion of a representation of $G \times H$

is given by considering the tensor product of vector spaces equipped with an appropriate $G \times H$ action. For now we define these concepts and later on in Section 4.3 we'll get the answers we seek.

There are several different definitions of tensor product of vector spaces we can give, all varying in level of abstraction, but producing fundamentally the same object, we start with a constructive approach and then see how this relates to other definitions.

Definition 4.1.41. *Let V, W, T be vector spaces over $\mathbb{F}$, a map $\theta : V \times W \rightarrow T$ is $\mathbb{F}$ -bilinear (or just bilinear when the field is clear from context) if:*

$$\theta(\alpha v_1 + \beta v_2, w) = \alpha \theta(v_1, w) + \beta \theta(v_2, w)$$

$$\theta(v, \alpha w_1 + \beta w_2) = \alpha \theta(v, w_1) + \beta \theta(v, w_2)$$

$$\theta(\alpha v, w) = \theta(v, \alpha w) = \alpha \theta(v, w)$$

For all $\alpha, \beta \in \mathbb{F}$, $v \in V$ and $w \in W$.

The way we build the tensor product of V and W is by considering the cartesian product $V \times W$ and enforcing the concept of bilinearity on the elements (v, w) .*

Definition 4.1.42. *Let V, W be vector spaces over $\mathbb{F}$ and consider the vector space E over $\mathbb{F}$ of formal linear combination of elements $(v, w) \in V \times W$, so elements in this space are linear combinations where the only possible interaction between them is $\alpha(v, w) - \alpha(v, w) = 0$, for $\alpha \in \mathbb{F}$, commutativity and all the other properties that define a vector space. Let K be the subspace of E generated by the elements:*

$$(v_1 + v_2, w) - (v_1, w) - (v_2, w)$$

$$(v, w_1 + w_2) - (v, w_1) - (v, w_2)$$

$$(\alpha v, w) - (v, \alpha w)$$

$$(\alpha v, w) - \alpha(v, w)$$

The tensor product space of V with W , denoted $V \otimes_{\mathbb{F}} W$ is the quotient vector space E/K .[†] Considering this we have that elements of $V \otimes W$ are of the form

$$\sum_{i=1}^n \alpha(v_i, w_i) + K$$

*The idea we'll use is analogous to the concept of presentations in group theory, refer to [Rot02, sec. 5.5] for more on this.

[†]This essentially enforces the properties $(v_1 + v_2, w) = (v_1, w) + (v_2, w)$, $(v, w_1 + w_2) = (v, w_1) + (v, w_2)$, $(\alpha v, w) = (v, \alpha w)$ and $(\alpha v, w) = \alpha(v, w)$ on the elements of E .

We denote $(v, w) + K$ as $v \otimes w$ and so we write any element of $V \otimes_{\mathbb{F}} W$ as

$$\sum_{i=1}^n \alpha(v_i \otimes w_i)$$

This definition usually constitutes part of the argument used to prove the existence of a tensor product of modules, according to definition D.1.18 in Appendix D.1, although here we specialize it for vector spaces. When the scalar product of $\mathbb{F}$ defined on V and W also makes sense on the right of these vector spaces then the definition above is consistent with definition D.1.18 in Appendix D.1 and in fact we can always assume this to be the case because as the vector spaces we consider are finite dimensional then if the dimension of V is n then $V \cong \mathbb{F}^n$ and because $\mathbb{F}$ is a field then in particular it's a commutative ring and so the right scalar product of $\mathbb{F}$ on $\mathbb{F}^n$ coincides with the left scalar product.

Since $v \otimes 0 = v \otimes (0 + 0) = v \otimes 0 + v \otimes 0$ we have that $v \otimes 0 = 0_{V \otimes W}$, that is the neutral element of the vector space $V \otimes_{\mathbb{F}} W$, and the same goes for $0 \otimes w$, for any $v \in V$ or $w \in W$. Moreover by proposition D.1.21 and its subsequent arguments, if V has dimension n and W has dimension m then $V \otimes W$ has dimension nm as so if $B_V = (v_1, v_2, \dots, v_n)$ and $B_W = (w_1, w_2, \dots, w_m)$ are basis of V and W respectively then by the bilinearity of elements in $V \otimes W$ it's easy to see that $\{v_i \otimes w_j \mid v_i \in B_V, w_j \in B_W\}$ is a basis of $V \otimes_{\mathbb{F}} W$.

Note that in a tensor product space not all elements can be written as $v \otimes w$, for instance consider $v \otimes w + v' \otimes w'$ where $v \neq v'$ and $w \neq w'$, the bilinear properties give us no way to combine these elements.

There's another definition of tensor product space we can give that enforces the bilinear properties on $V \times W$ in a more implicit way.

Definition 4.1.43. Let V, W be vector spaces over $\mathbb{F}$ with $\dim_{\mathbb{F}}(V) = n$ and $\dim_{\mathbb{F}}(W) = m$. Let $B_V = (v_1, v_2, \dots, v_n)$ and $B_W = (w_1, w_2, \dots, w_m)$ be a basis of V and W respectively, then the tensor product space $V \otimes_{\mathbb{F}} W$ is the nm -dimensional vector space over $\mathbb{F}$ with basis given by the formal elements

$$\{v_i \otimes w_j \mid v_i \in B_V, w_j \in B_W\}$$

As such the elements of $V \otimes_{\mathbb{F}} W$ are linear combination of these elements, that is

$$V \otimes_{\mathbb{F}} W = \left\{ \sum_{i,j} \alpha_{ij} (v_i \otimes w_j) \mid \alpha_{ij} \in \mathbb{F} \right\}$$

Considering this addition and scalar products are defined as usual and we additionally define, for $v = \sum_i \alpha_i v_i$ and $w = \sum_j \beta_j w_j$

$$v \otimes w = \sum_{i,j} \alpha_i \beta_j (v_i \otimes w_j)$$

The last expression on the above definition gives us precisely the bilinear properties we need. Moreover it's straightforward to prove that the definition is independent of the basis B_V and B_W we choose, any basis of these vector spaces will produce a set of generators for $V \otimes_{\mathbb{F}} W$ and as there are nm of them we must then have a basis.

In both of the given definitions of tensor product space we end up with a nm -dimensional vector space over $\mathbb{F}$ and thus both vector spaces must be isomorphic.

We are now ready to define the tensor product of representations.

Definition 4.1.44. Let G, H be finite groups and ρ_V, σ_W respective representations, we define the outer tensor representation $\rho \otimes \sigma : G \times H \rightarrow \mathbf{GL}(V \otimes_{\mathbb{F}} W)$ by

$$[(\rho \otimes \sigma)(g, h)] \left(\sum_{i,j} \alpha_{ij} (v_i \otimes w_j) \right) = \sum_{i,j} \alpha_{ij} (\rho_g v_i \otimes \sigma_h w_j)$$

If $G = H$ then we define the inner tensor representation $\rho \otimes \sigma : G \rightarrow \mathbf{GL}(V \otimes_{\mathbb{F}} W)$ by

$$[(\rho \otimes \sigma)(g)] \left(\sum_{i,j} \alpha_{ij} (v_i \otimes w_j) \right) = \sum_{i,j} \alpha_{ij} (\rho_g v_i \otimes \sigma_g w_j)$$

The difference between the *inner* and *outer* representations is that one defines an action of $G \times H$ on $V \otimes_{\mathbb{F}} W$ and the other an action of G on $V \otimes_{\mathbb{F}} W$.

To make sure the definition makes sense we must verify that indeed we have a representation in each case. Indeed

$$\begin{aligned} [(\rho \otimes \sigma)((g, h)(g', h'))] \left(\sum_{i,j} \alpha_{ij} (v_i \otimes w_j) \right) &= [(\rho \otimes \sigma)(gg', hh')] \left(\sum_{i,j} \alpha_{ij} (v_i \otimes w_j) \right) = \\ &= \sum_{i,j} \alpha_{ij} (\rho_{gg'} v_i \otimes \sigma_{hh'} w_j) = \sum_{i,j} \alpha_{ij} (\rho_g \rho_{g'} v_i \otimes \sigma_h \sigma_{h'} w_j) = \\ &= [(\rho \otimes \sigma)((g, h))] \left([(\rho \otimes \sigma)(g', h')] \left(\sum_{i,j} \alpha_{ij} (v_i \otimes w_j) \right) \right) \end{aligned}$$

The inner representation case is analogous.

If V and W have inner products $\langle \cdot, \cdot \rangle_V$ and $\langle \cdot, \cdot \rangle_W$ then we can define an inner product on $V \otimes_{\mathbb{F}} W$ by

$$\langle v \otimes w, v' \otimes w' \rangle_{V \otimes W} = \langle v, v' \rangle_V \langle w, w' \rangle_W$$

Thus if B_V, B_W are orthonormal basis of V and W respectively then the resulting basis of $V \otimes_{\mathbb{F}} W$ is also orthonormal. Moreover if the inner products are G -invariant (H -invariant) then the resulting inner product on $V \otimes_{\mathbb{F}} W$ is also G -invariant ($G \times H$ -invariant).

4.1.7 Matrix coefficients

The matrix coefficients of each matrix in a matrix representation, when unitary, have very interesting properties, one of them being as we shall see that they produce a basis for the space of complex functions $f : G \rightarrow \mathbb{C}$ where G is a group.* Here we give some notation and simple properties of the matrix coefficients.

Definition 4.1.45. Given a matrix representation R over $\mathbb{F}^n$, let $R_g = (\varphi_{i,j})$ for $i, j \in [n]$. Then we can define for each i, j a function $\varphi_{ij} : G \rightarrow \mathbb{F}$ by $\varphi_{ij}(g) = \varphi_{i,j}$ where $\varphi_{i,j}$ is the respective matrix coefficient of R_g . We call the set of functions $\{\varphi_{ij} \mid i, j \in [n]\}$ the matrix coefficients of R .

Of course given any representation ρ_V , by considering an equivalent matrix representation, we can always get the corresponding matrix coefficients, yet do note that they depend on the chosen basis of V . We usually write φ_{ij}^ρ to make it clear what representation we're referring to, when required.

If V is an inner product space and $B = (v_1, v_2, \dots, v_n)$ is an orthonormal basis of V we can get the corresponding matrix coefficients by noting that, referring to Appendix A, proposition C.0.5

$$\varphi_{ij}(g) = \langle \rho_g v_j, v_i \rangle_V$$

Remark 3. We stress for one last time that when we're talking about an inner product space we always assume the inner product used is the G -invariant one given in Theorem 4.1.12.

Proposition 4.1.46. Let V be an inner product space and ρ_V a representation of G , then the matrix coefficients of the adjoint representation $\rho_{V^*}^\circ$ with respect to the orthonormal dual basis $B^* = \{\theta_{v_1}, \dots, \theta_{v_n}\}$ are

$$\varphi_{ij}^{\rho^\circ}(g) = \langle \rho_g^\circ \theta_{v_j}, \theta_{v_i} \rangle_{V^*} = \overline{\varphi_{ji}^\rho(g)}$$

*This effectively produces a *Fourier Analysis* for noncommutative groups.

Where φ_{ij}^ρ is a matrix coefficient of ρ_V .

Proof. This is immediate as

$$\langle \rho_g^\circ \theta_{v_j}, \theta_{v_i} \rangle_{V^*} = \langle \theta_{\rho_g v_j}, \theta_{v_i} \rangle_{V^*} = \langle v_i, \rho_g v_j \rangle_V = \overline{\langle \rho_g v_j, v_i \rangle_V} = \overline{\varphi_{ij}^\rho(g)}$$

□

For the tensor product we have the following.

Proposition 4.1.47. *Let V, W be inner product spaces and ρ_V, σ_W representations of G and H respectively, then for an orthonormal basis $B_V = (v_1, v_2, \dots, v_n)$ of V and an orthonormal basis $B_W = (w_1, w_2, \dots, w_m)$ of W the matrix coefficients of the tensor product representation $\rho \otimes \sigma$ with respect to the basis $(v_i \otimes w_j \mid v_i \in B_V, w_j \in B_W)$ are*

$$\varphi_{ij, i' j'}^{\rho \otimes \sigma}(g, h) = \langle (\rho_g \otimes \sigma_h)(v_i \otimes w_j), v_{i'} \otimes w_{j'} \rangle_{V \otimes W} = \varphi_{ii'}^\rho(g) \varphi_{jj'}^\sigma(h)$$

Proof. Again immediate considering

$$\begin{aligned} \langle (\rho_g \otimes \sigma_h)(v_i \otimes w_j), v_{i'} \otimes w_{j'} \rangle_{V \otimes W} &= \langle (\rho_g v_i \otimes \sigma_h w_j), v_{i'} \otimes w_{j'} \rangle_{V \otimes W} = \\ &= \langle \rho_g v_i, v_{i'} \rangle_V \langle \sigma_h w_j, w_{j'} \rangle_W = \varphi_{ii'}^\rho(g) \varphi_{jj'}^\sigma(h) \end{aligned}$$

□

4.2. Complete reducibility

4.2.1 Maschke's theorem

As we've seen in example 4.1.36 the permutation representation λ of S_2 decomposes as a direct sum of two irreducible representations ι, ϵ , or equivalently the respective $\mathbb{C}S_2$ -module V decomposes into a direct sum of two simple $\mathbb{C}S_2$ -modules W and U . More generally when a representation ρ of a group G can be decomposed into a direct sum of irreducible representations we say that the representation ρ is *semisimple/completely reducible*, this is then consistent with the respective definitions for modules in D.1.16 and D.1.22 considering corollary D.1.24. As it turns out, because we are considering representations over fields of characteristic 0, this is always the case and it's one of the reasons we choose to work over such fields.

Theorem 4.2.1 (Maschke). *Every finite dimensional representation ρ_V of a finite group G with V over a field $\mathbb{F}$ of characteristic 0 is completely reducible.*

By corollary D.1.24 completely reducible is interchangeable with semisimple and so every finite dimensional representation ρ_V of a finite group G can be written as a direct sum of irreducible representations. If a representation is irreducible then it's trivially semisimple/completely reducible. The above theorem is a particular case of the more general Maschke's theorem [Bea99, 2.3.8] and follows by induction with the following lemma.

Lemma 4.2.2. *Let ρ_V be a representation of a finite group G with V over a field $\mathbb{F}$ of characteristic 0. If W is a G -subspace of V then there exists U G -subspace of V such that $V = W \oplus U$.*

Proof. Considering V is a vector space and W a subspace we can define a projection $p : V \rightarrow W$ from V onto W , that is, letting W' be the orthogonal complement of W so that $V = W \oplus W'$ then any $v \in V$ can be written uniquely as a sum of an element of W and an element of W' , say $v = w + w'$, then $p(v) = w$. The map p is $\mathbb{F}$ -linear and it's straightforward to see that $V = \text{Im}(p) \oplus \ker(p)$ and $\text{Im}(p) = W$, our objective will be to produce a G -homomorphism $\hat{p} : V \rightarrow W$ such that ${}_G V = {}_G \text{Im}(\hat{p}) \oplus {}_G \ker(\hat{p})$ with $\text{Im}(\hat{p}) = W$ thus decomposing the G -space V into the direct sum of two G -subspaces $\text{Im}(\hat{p})$ and $\ker(\hat{p}) = U$ *. To that end let $|G| = n$ and consider the map

$$\hat{p}(v) = \frac{1}{n} \sum_{g \in G} gp(g^{-1}v)$$

for $v \in V$. Because $\mathbb{F}$ has characteristic 0 then $1/n$ is well defined for in this case we have that $n \neq 0$, in fact this will be the case as long as the characteristic of $\mathbb{F}$ is not a divisor of $n = |G|$ so we could relax that assumption (this is the case in the more general Maschke's theorem). First let's check this is a G -homomorphism, indeed for αh with $\alpha \in \mathbb{F}$ and $h \in G$

$$\begin{aligned} \hat{p}(\alpha hv) &= \frac{1}{n} \sum_{g \in G} gp(g^{-1}\alpha hv) = \frac{\alpha}{n} \sum_{g \in G} h(h^{-1}g)p((h^{-1}g)^{-1}v) = \\ &= \frac{\alpha h}{n} \sum_{g \in G} (h^{-1}g)p((h^{-1}g)^{-1}v) = \frac{\alpha h}{n} \sum_{g' \in G} g'p(g'^{-1}v) = \alpha h\hat{p}(v) \end{aligned}$$

The other properties follows as easily, moreover if $w \in W$ then $g^{-1}w \in W$ for all $g \in G$ thus

$$\hat{p}(w) = \frac{1}{n} \sum_{g \in G} gp(g^{-1}w) = \frac{1}{n} \sum_{g \in G} gg^{-1}w = \frac{1}{n} \sum_{g \in G} w = w$$

and it's straightforward to see that for $v \in V$, $\hat{p}(v) \in W$ considering p projects to W , hence $\text{Im}(\hat{p}) = W$ and we consider $\hat{p} : V \rightarrow W$.

Now if $i : W \rightarrow V$ is the inclusion of W in V , that is $i(w) = w$ for $w \in W$, then it's

*A module homomorphism in these conditions is said to be split, refer to [Bea99, 2.2.9] for more on this.

clear that $\hat{p}i = id_W$ and this is sufficient to prove that $\mathbb{F}G V = \mathbb{F}G \text{Im}(i) \oplus \mathbb{F}G \ker(\hat{p})$. Noting $\text{Im}(i) = \text{Im}(\hat{p})$ the conclusion follows, yet for completeness we prove this assertion. Indeed if $\hat{p}i = id_W$ then because i and $\hat{p}$ are $\mathbb{F}G$ -homomorphisms then $\text{Im}(i)$ and $\ker(\hat{p})$ are $\mathbb{F}G$ -submodules of V , moreover if $v \in \text{Im}(i) \cap \ker(\hat{p})$ then $v = i(w)$ for some $w \in W$ and $w = \hat{p}i(w) = \hat{p}(v) = 0$ so $w = 0$ hence $v = i(w) = 0$, therefore $\text{Im}(i) \cap \ker(\hat{p}) = \{0\}$. Finally for $v \in V$ consider $v - i\hat{p}(v)$, then $\hat{p}(v - i\hat{p}(v)) = \hat{p}(v) - \hat{p}i\hat{p}(v) = \hat{p}(v) - \hat{p}(v) = 0$ thus $v - i\hat{p}(v) \in \ker(\hat{p})$ and clearly $v = i\hat{p}(v) + (v - i\hat{p}(v))$ and $i\hat{p}(v) \in \text{Im}(i)$, therefore $V = \text{Im}(i) + \ker(\hat{p})$, so in conclusion $V = \text{Im}(i) \oplus \ker(\hat{p}) = \text{Im}(\hat{p}) \oplus \ker(\hat{p})$ as $\mathbb{F}G$ -modules and thus the same goes for the respective equivalent G -spaces. $\square$

To complete the proof of theorem 4.2.1 all we have to do is note that V is a finite dimensional vector space, so if the representation has dimension 1 then it's trivially irreducible, if it has dimension greater than 1 then it either is already irreducible and so semisimple, or it has a G -subspace and so we can apply the previous lemma to decompose it as the direct sum of two invariant subspaces that by the induction hypothesis are semisimple.

When we consider inner product spaces then the proof of lemma 4.2.2 becomes even simpler, indeed by adopting a G -invariant inner product on V , $\langle \cdot, \cdot \rangle$ and letting W be some G -subspace of V , then the orthogonal complement $W^\perp$ is also a G -subspace, to see this note that for $w \in W$, $w' \in W^\perp$ and $g \in G$

$$\langle w, gw' \rangle = \langle g^{-1}w, w' \rangle = 0$$

considering $g^{-1}w \in W$, hence $gw' \in W^\perp$. Therefore $V = W \oplus W^\perp$ as G -spaces/ $\mathbb{F}G$ -modules.

Example 4.2.3. As stated before the regular representation λ of a group G over a field $\mathbb{F}$ of characteristic 0, as in definition 4.1.18, is of particular interest, one of the reasons being that any irreducible representation ρ_V of G is a subrepresentation of the regular representation of G , this will later follow from a more general result via character theory but we can easily prove it now using lemma 4.2.2.

Take any $0 \neq v \in V$ and define $q : \mathbb{F}G \rightarrow V$ by

$$q \left(\sum_{g \in G} \alpha_g g \right) = \sum_{g \in G} \alpha_g (gv)$$

It's straightforward to see that this is a G -homomorphism, moreover because $\text{Im}(q)$ is a G -subspace of V and V is an irreducible G -space we must have that $\text{Im}(q) = V$ considering

$q(1) = v \neq 0$, therefore q is onto. Also by q being a G -homomorphism we must have that $\ker(p)$ is a G -subspace of $\mathbb{F}G$ (it is a submodule), now by lemma 4.2.2 $\ker(p)$ has a complementary G -subspace U such that $\mathbb{F}G = \ker(p) \oplus U$ as $\mathbb{F}G$ -modules and by the isomorphism theorem for modules D.1.7 we have that

$$V = \text{Im}(q) \cong \mathbb{F}G / \ker(p) \cong U$$

Therefore V is isomorphic to the submodule U of $\mathbb{F}G$, or equivalently ρ_V is a subrepresentation of λ (technically ρ_V is *equivalent* to a subrepresentation of λ , but as stated before we generally don't distinguish between equivalent representations).

In conclusion we have the following.

Corollary 4.2.4. *Let ρ_V be a representation of G over a field of characteristic 0, then the G -space V decomposes into a direct sum*

$$V = \bigoplus_{i \in [m]} V_i$$

where the G -spaces V_i are simple.

By definition 4.1.34 and its subsequent statement we can also equivalently say that the representation ρ_V decomposes as the direct sum

$$\rho_V \equiv \bigoplus_{i \in [m]} \sigma_{V_i}$$

Where then the representations σ_{V_i} are the implied representations in the G -spaces V_i , if the direct sum is internal then we could just write

$$\rho_V \equiv \bigoplus_{i \in [m]} \rho_{V_i}$$

Where ρ_{V_i} is the restriction of ρ_V to the G -subspace V_i . Of course considering the equivalence of inner and outer direct sums of modules we'll always have $\rho_{V_i} \equiv \sigma_{V_i}$.

Note that in all of the above we mention nothing about how distinct the factors in the decomposition are, in the sense that two or more factors V_i might well be isomorphic G -spaces, that is the respective representations σ_{V_i} might well be equivalent. In fact this is generally the case as many simple examples show, the simplest perhaps being the trivial representation of a group G on a vector space V of dimension n defined as $\iota^n : G \rightarrow \text{GL}(V)$ where $\iota_g^n = id_V$ for all $g \in G$. Given a basis $B = (v_1, \dots, v_n)$ of V it's

clear that any v_i generates a G -invariant subspace of V , moreover because these are one dimensional then they are simple G -spaces, also for any $i, j \in [n]$ with $i \neq j$ we have that $V_i \cap V_j = \{0\}$ and $V = V_1 \oplus \dots \oplus V_n$ as G -spaces, therefore this is a decomposition of V into simple G -subspaces, now note that the representations implied in each G -subspaces V_i are equivalent to the trivial representation $\iota : G \rightarrow \mathbb{F}$ as in definition 4.1.13, indeed each space V_i is the space generated by v_i , denoted $\langle v_i \rangle$, and the map $\theta : \langle v_i \rangle \rightarrow \mathbb{F}$ defined as $\theta(\alpha v_i) = \alpha$, $\alpha \in \mathbb{F}$, gives a G -isomorphism from V_i to $\mathbb{F}$ (where $\mathbb{F}$ is a G -space by the trivial representation ι). Therefore

$$\iota^n \equiv \bigoplus_{i \in [n]} \iota$$

As such one usually writes the decomposition in corollary 4.2.4 as

$$V = \bigoplus_{j \in [n]} n_j V_j$$

where we group together the isomorphic G -spaces into a single G -space denoted $n_j V_j$, so that $n_j V_j$ is the direct sum of the n_j G -spaces isomorphic to V_j , that is

$$n_j V_j = \bigoplus_{i \in I_j} V_i$$

where I_j is the set of indices in $[m]$ for which $V_i \cong V_j$, so $|I_j| = n_j$ (note that $n_j \in \mathbb{N}$). The space V_j is chosen as a representative of those n_j isomorphic G -spaces, but because they're all isomorphic any such other would do, suffice to say for any $i \in I_j$

$$n_j V_j \cong \underbrace{V_i \oplus \dots \oplus V_i}_{n_j}$$

and of course $j \in I_j$ which justifies the notation.

4.2.2 Schur's lemma

We know now that representations as we consider them are semisimple but we have yet to determine if such a decomposition is unique or not and what information can we gain about it. The next theorem, known as *Schur's lemma*, will do the trick. Although simple to prove it produces a set of very useful observations and for that, together with Maschke's theorem, it largely justifies why we give preference to setting $\mathbb{F} = \mathbb{C}$.

Theorem 4.2.5 (Schur). *Let ρ_V and σ_W be two irreducible representations of G over a field $\mathbb{F}$.*

- a) *Any G -homomorphism $\theta : V \rightarrow W$ is either an isomorphism or identically 0.*

b) If $\mathbb{F}$ is an algebraically closed field * then $\text{Hom}_G(\rho_V, \rho_V) = \mathbb{F}id_V$ where id_V is the identity map on V , in particular $\text{Hom}_G(\rho_V, \rho_V) \cong \mathbb{F}$

Proof. a) Let $\theta \in \text{Hom}_G(V, W)$, then because $\ker(\theta)$ and $\text{Im}(\theta)$ are G -subspaces of V and W respectively, and these are simple, we must have that either $\ker(\theta) = \{0\}$ and so $\text{Im}(\theta) = W$ thus θ is a G -isomorphism, or $\ker(\theta) = V$ and so θ is identically 0.

b) Let $\theta \in \text{Hom}_G(V, V)$, because $\mathbb{F}$ is algebraically closed and θ is an $\mathbb{F}$ -linear map from V to V then it must have an eigenvalue in $\mathbb{F}$, for its characteristic polynomial is in $\mathbb{F}[x]$ so it must have a root in $\mathbb{F}$, call it α . Consider the map $\theta - \alpha id_V$, it is clearly a G -homomorphism and so by a) we must have that either $\theta - \alpha id_V$ is an isomorphism or identically 0. As $\ker(\theta - \alpha id_V)$ contains the eigenvectors associated with the eigenvalue α we must have that $\ker(\theta - \alpha id_V) = V$ and so $\theta - \alpha id_V = 0$ hence $\theta = \alpha id_V$, the conclusion follows by noting that for any $\alpha \in \mathbb{F}$, $\alpha id_V \in \text{Hom}_G(V, V)$.

□

Recall that a homomorphism from a space V onto itself is usually called an endomorphism on V , and the space $\text{Hom}(V, V)$ is then simply denoted as $\text{End}(V)$.

Corollary 4.2.6. Let ρ_V and σ_W be two irreducible representation of G over an algebraically closed field $\mathbb{F}$. Then

$$\dim_{\mathbb{F}}(\text{Hom}_G(\rho_V, \sigma_W)) = \begin{cases} 1 & \rho \equiv \sigma \\ 0 & \text{otherwise} \end{cases}$$

Proof. Let $\theta, \gamma \in \text{Hom}_G(V, W)$, then by part a) of Schur's lemma (theorem 4.2.5) θ and γ are isomorphisms and so have inverses, thus we can have $\theta^{-1}\gamma \in \text{End}_G(V)$ and by part b) of Schur's lemma we must have that $\theta^{-1}\gamma = \alpha id_V$ for some $\alpha \in \mathbb{F}$ therefore $\gamma = \alpha\theta$. For the other case if ρ and σ are non equivalent then θ can't be an isomorphism and so by part a) of Schur's lemma it must be identically 0. □

As a first application of the minute representation theory we've developed so far we have the following.

*An algebraically closed field $\mathbb{F}$ is a field such that any polynomial in the polynomial ring $\mathbb{F}[x]$ has a root in $\mathbb{F}$.

Proposition 4.2.7. *If G has a faithful irreducible representation over an algebraically closed field $\mathbb{F}$ then its center $Z(G)$ is cyclic, in particular if additionally G is abelian then G is cyclic.*

Proof. Let ρ_V be such representation of G , then because for $z \in Z(G)$ and $g \in G$ $zg = gz$ we have that $\rho_z \rho_g = \rho_g \rho_z$ therefore $\rho_z \in \text{End}_G(V)$ and so by part b) of Schur's lemma $\rho_z = \alpha_z \text{id}_V$ for some $\alpha_z \in \mathbb{F}$. Now let $\mathbb{F}^\times$ denote the multiplicative group in $\mathbb{F}$, that is $\mathbb{F}$ with just the multiplicative operation defined on it, which indeed is a group as $\mathbb{F}$ is a field, and consider the map $\mu : Z(G) \rightarrow \mathbb{F}^\times$ defined by $\mu(z) = \alpha_z$, it's straightforward to see that this defines a group homomorphism and because ρ is faithful, that is injective, then μ also is and so $\mu(Z(G))$ is a finite subgroup of $\mathbb{F}^\times$. As is known finite subgroups of the multiplicative group of a field are necessarily cyclic [Rot02, 3.30], therefore $\mu(Z(G)) \cong Z(G)$ is cyclic.

If G is abelian then $G = Z(G)$ and the conclusion follows by the previous argument. $\square$

Another demonstration of the power of Schur's lemma is that it allows us to characterize the irreducible representations of abelian groups over the field $\mathbb{C}$. We start with the following simple observation.

Proposition 4.2.8. *Two one dimensional representations are equivalent if and only if they are equal.*

Proof. Clearly if they are equal then they are equivalent. Going the other way around if two one dimensional representations ρ, σ are equivalent as they are one dimensional we can consider $\rho_g, \sigma_g \in \mathbb{F}$ for all $g \in G$, as such if θ is a G -isomorphism that exists by the equivalence of the representations we get, for $x \in \mathbb{F}$, $\sigma_g \theta(x) = \theta(\rho_g x) = \rho_g \theta(x)$ because θ is in particular $\mathbb{F}$ -linear and so by θ being bijective, for all $g \in G$ and $x \in \mathbb{F}$, $\sigma_g x = \rho_g x$ therefore $\sigma_g = \rho_g$ for all $g \in G$ and so $\rho = \sigma$. $\square$

Proposition 4.2.9. *All irreducible representations of an abelian group G over an algebraically closed field are one dimensional.*

The converse of this theorem is also true as we'll later see in Section 4.4.

Proof. If G is abelian then given an irreducible representation ρ_V we have that for all $g \in G$, $\rho_g \in \text{Hom}_G(V, V)$ considering $\rho_g \rho_h = \rho_h \rho_g$ for all $h \in G$, so by part b) of Schur's lemma it must be the case that for all $g \in G$, $\rho_g = \alpha_g \text{id}_V$ for some $\alpha_g \in \mathbb{F}$, but then given $0 \neq v \in V$ we have that for all $g \in G$, $\rho_g v = \alpha_g \text{id}_V v = \alpha_g v$ thus the space generated by v , denoted $\langle v \rangle$, is G -invariant (and consequently, being it one dimensional, irreducible).

But because V is already irreducible it must be the case that $V = \langle v \rangle$ and so ρ_V is one dimensional. $\square$

Proposition 4.2.10. *If G is abelian then it has precisely $|G|$ irreducible representations over the field $\mathbb{C}$.*

Proof. Considering G is an abelian finite group then it is equal to a direct product of cyclic groups, that is

$$G = \bigotimes_{i \in [m]} C_{n_i}$$

where C_{n_i} is a cyclic group of order $n_i \in \mathbb{N}$. As such let C_{n_i} be generated by $g_i \in C_{n_i}$, then $C_{n_i} = \langle g_i \rangle$ where $\langle g_i \rangle$ is the group generated by g_i . As any irreducible representation of G must be one dimensional we have that such a representation is a homomorphism $\rho : G \rightarrow \mathbb{C}^\times$ where $\mathbb{C}^\times$ is the multiplicative group of the field $\mathbb{C}$ (this considering that $\text{GL}(\mathbb{C}) = \mathbb{C}^\times$). Now let

$$\rho((1_{C_{n_1}}, \dots, 1_{C_{n_{j-1}}}, g_j, 1_{C_{n_{j+1}}}, \dots, 1_{C_{n_m}})) = \lambda_j$$

Because ρ is a homomorphism it must then be the case that $\lambda^{n_j} = 1$ and so λ is a n_j th root of unity, moreover because the groups C_{n_i} are cyclic ρ is uniquely determined by the values λ_i . Indeed

$$\rho((g_1^{k_1}, \dots, g_m^{k_m})) = \lambda_1^{k_1} \lambda_2^{k_2} \dots \lambda_m^{k_m}$$

On the other hand if we let λ_i be a n_i th root of unity, for each $i \in [m]$, and define $\sigma : G \rightarrow \mathbb{C}^\times$ as

$$\rho((g_1^{k_1}, \dots, g_m^{k_m})) = \lambda_1^{k_1} \lambda_2^{k_2} \dots \lambda_m^{k_m}$$

for $k_i \in \mathbb{Z}$ then it is clear this defines a group homomorphism and so a one dimensional representation of G that consequently must be irreducible. By proposition 4.2.8 these representation must be non equivalent thus we have a bijection between irreducible representations of G and m -tuples $(\lambda_1, \dots, \lambda_m)$ where λ_i is an n_i th root of unity. As there are $n_1 n_2 \dots n_m$ such tuples and $|G| = |C_{n_1}| |C_{n_2}| \dots |C_{n_m}| = n_1 \dots n_m$ we have that there are $|G|$ irreducible representations of $|G|$, or in other words $|\widehat{G}| = |G|$. $\square$

The proof of the above proposition explicitly gives us the sought after characterization of the irreducible representations over $\mathbb{C}$ of abelian groups, although it requires the decomposition of the group G into cyclic groups.

Example 4.2.11. Consider the cyclic group of order n , C_n and let it be generated by g , thus $C_n = \langle g \rangle$. By the above proposition C_n has n irreducible representations ρ^j , for $j \in [n]$. Those are defined as

$$\rho^j(g^k) = e^{\frac{2\pi i}{n}jk}$$

for $k \in \mathbb{Z}$. If we now consider the group $G = C_n \otimes C_m$ then letting h be a generator of C_m we have nm irreducible representations ρ^{jl} with $j \in [n]$ and $l \in [m]$, which are defined as

$$\rho^{jl}((g^{k_1}, h^{k_2})) = e^{\frac{2\pi i}{n}jk_1} e^{\frac{2\pi i}{m}lk_2}$$

for $k_1, k_2 \in \mathbb{Z}$.

4.2.3 Multiplicities and isotypic components

We now assume for the remainder of this section that the field $\mathbb{F}$ we are working with is of characteristic 0 and algebraically closed.

Recalling corollary 4.2.4, given a representation ρ_V , or equivalently a G -space V , we can decompose it into a direct sum of simple G -spaces as

$$V = \bigoplus_{i \in [n]} n_i V_i \tag{4.1}$$

where then $n_i V_i$ is the direct sum of the n_i G -spaces isomorphic to V_i , that is

$$n_i V_i = \bigoplus_{j \in I_i} V_j$$

where I_i is the set of indices in $[m]$ (as in corollary 4.2.4) for which $V_j \cong V_i$, so $|I_i| = n_i$.

Now for each $i \in [n]$ and $j \in I_i$ consider a G -isomorphism $\Gamma_{ij} : V_i \rightarrow V_j$, then we have that

$$n_i V_i = \bigoplus_{j \in I_i} \Gamma_{ij}(V_i)$$

Of course we already knew $n_i V_i \cong \overbrace{V_i \oplus \cdots \oplus V_i}^{n_i}$, but the isomorphisms Γ_{ij} are of particular interest.

Proposition 4.2.12. *Considering the above, the G -isomorphisms $\{\Gamma_{ij} \mid j \in I_i\}$ form a basis of the space $\text{Hom}_G(V_i, V)$, in particular $\dim_{\mathbb{F}}(\text{Hom}_G(V_i, V)) = n_i$.*

Proof. By proposition D.1.14 and corollary 4.2.4 we have that

$$\text{Hom}_G(V_i, V) \cong \bigoplus_{j \in [m]} \text{Hom}_G(V_i, V_j)$$

where each V_j , $j \in [m]$, is a simple G -space. As such because V_i is also simple by Schur's lemma $\dim_{\mathbb{F}}(\text{Hom}_G(V_i, V_j))$ is either 1 or 0 depending on whether the G -spaces are isomorphic or not. But then by 4.1 we know that $\dim_{\mathbb{F}}(\text{Hom}_G(V_i, V_j)) = 1$ if $j \in I_i$ and 0 otherwise hence

$$\text{Hom}_G(V_i, V) \cong \bigoplus_{j \in I_i} \text{Hom}_G(V_i, V_j)$$

Now we have $\Gamma_{ij} \in \text{Hom}_G(V_i, V_j)$ therefore $\text{Hom}_G(V_i, V_j) = \mathbb{F} \Gamma_{ij}$ and so

$$\text{Hom}_G(V_i, V) \cong \bigoplus_{j \in I_i} \mathbb{F} \Gamma_{ij}$$

Thus it's clear that $\{\Gamma_{ij} \mid j \in I_i\}$ is a basis of the space $\text{Hom}_G(V_i, V)$. □

Note that the G -isomorphisms Γ_{ij} were chosen arbitrarily, any choice will produce a basis of the respective space $\text{Hom}_G(V_i, V)$.

Theorem 4.2.13. *Let V be a G -space with decomposition as in corollary 4.2.4 and W a simple G -space, then $\dim_{\mathbb{F}}(\text{Hom}_G(W, V)) = |\{i \in [m] \mid W \cong V_i\}|$.*

The proof of the above theorem is analogous to the one of the preceding proposition and follows easily considering that by Schur's lemma $\dim_{\mathbb{F}}(\text{Hom}_G(W, V_i)) = 1$ if $W \cong V_i$ and 0 otherwise. This is precisely the result we need to assert the uniqueness of the decomposition in corollary 4.2.4.

Theorem 4.2.14. *Let V be a G -space with decomposition as in corollary 4.2.4, then this decomposition is unique up to order and isomorphism of factors.*

Proof. Indeed if V has two decompositions into simple factor $V_1 \oplus \cdots \oplus V_n$ and $V'_1 \oplus \cdots \oplus V'_m$ then by theorem 4.2.13

$$|\{j \in [m] \mid V'_i \cong V'_j\}| = \dim_{\mathbb{F}}(\text{Hom}_G(V'_i, V)) = |\{j \in [n] \mid V_i \cong V_j\}|$$

and in the same way

$$|\{j \in [m] \mid V_i \cong V'_j\}| = \dim_{\mathbb{F}}(\text{Hom}_G(V_i, V)) = |\{j \in [n] \mid V_i \cong V_j\}|$$

thus both decompositions must coincide up to order and isomorphism of factors. □

With this we can have the following definition knowing it will be well defined.

Definition 4.2.15. Let V be a G -space with decomposition as in corollary 4.2.4, then we can write

$$V = \bigoplus_{i \in [n]} n_i V_i$$

If we let $n_i V_i = \overbrace{V_i \oplus \cdots \oplus V_i}^{n_i}$ then the G -space $n_i V_i$ is called the isotypic component of V_i in V and n_i the multiplicity of V_i in V . This decomposition of V is called the canonical decomposition or the decomposition into isotypic components of V .

When $n_i = 1$ for all $i \in [n]$ we say the G -space V , or the implied representation ρ_V , is multiplicity free.

4.2.4 The structure of $\text{End}_G(V)$

The G -isomorphisms Γ_{ij} in the previous subsection can be used to construct another set of G -homomorphism also of great interest. Recall we've considered the maps Γ_{ij} for $i \in [n]$ and $j \in I_i$ (per the decomposition in 4.1), this way we indeed have G -isomorphism, yet more generally we can define the G -homomorphism $\mathcal{T}_{ij} : V_i \rightarrow V_j$ for any $i, j \in [m]$ (per the decomposition in corollary 4.2.4) and by Schur's lemma these maps are either G -isomorphisms or the null map. For $i \in [m]$ let $i' \in [n]$ be such that $i \in I_{i'}$, it should be clear that i' is unique in these conditions, then we define the maps $\mathcal{T}_{ij}$ as

$$\mathcal{T}_{ij} = \begin{cases} \Gamma_{i'j} \Gamma_{i'i}^{-1} & \text{if } j \in I_{i'} \\ 0 & \text{otherwise} \end{cases}$$

It is clear that $\mathcal{T}_{ij}(V_i) \subseteq V_j$ and that these are G -homomorphism considering the null map is and $\Gamma_{i'j} \Gamma_{i'i}^{-1}$ is a composition of G -homomorphism thus one itself.

Theorem 4.2.16. Let V be a G -space with decomposition as in corollary 4.2.4 and consider also the decomposition 4.1, then the set $\{\mathcal{T}_{ij} \mid i, j \in I_{i'}, i' \in [n]\}$ produces a basis of $\text{End}_G(V)$, in particular

$$\dim_{\mathbb{F}}(\text{End}_G(V)) = \sum_{i' \in [n]} n_{i'}^2$$

That is $\dim_{\mathbb{F}}(\text{Hom}_G(V, V))$ is equal to the sum of the squares of the multiplicities as in definition 4.2.15.

Proof. By proposition D.1.14 and corollary 4.2.4 we have that

$$\text{End}_G(V) \cong \bigoplus_{i, j \in [m]} \text{Hom}_G(V_i, V_j)$$

and because V_i, V_j are simple G -spaces then by Schur's lemma $\dim_{\mathbb{F}}(\text{Hom}_G(V_i, V_j))$ is either 1 or 0 depending on whether $V_i \cong V_j$, therefore considering the decomposition 4.1 we have

$$\text{End}_G(V) \cong \bigoplus_{i' \in [n]} \bigoplus_{i, j \in I_{i'}} \text{Hom}_G(V_i, V_j)$$

so here we always have $V_i \cong V_j$, therefore because $\mathcal{T}_{ij} \in \text{Hom}_G(V_i, V_j)$ we have

$$\text{Hom}_G(V_i, V_j) = \mathbb{F}\mathcal{T}_{ij}$$

hence

$$\text{End}_G(V) \cong \bigoplus_{i' \in [n]} \bigoplus_{i, j \in I_{i'}} \mathbb{F}\mathcal{T}_{ij}$$

So indeed we get a basis of $\text{End}_G(V)$ and the conclusion follows considering $|I_{i'}| = n_{i'}$. □

This theorem produces a couple of straightforward corollaries. First note that

$$\text{End}_G(V) \cong \bigoplus_{i \in [n]} M_{n_i}(\mathbb{F})$$

as vector spaces, where $M_{n_i}(\mathbb{F})$ is the space of square matrices of size $n_i \times n_i$ and entries in the field $\mathbb{F}$, this is immediate considering that by the above theorem both are vector spaces over $\mathbb{F}$ with the same dimension. The isomorphism above can easily be made explicit, essentially what we're saying is that, considering theorem 4.2.16, we can write any element of $\text{End}_G(V)$ as a linear combination of elements of the given basis, as such we can then organize the factor of that linear combination in matrices, one matrix for each $i' \in [n]$ so that for $T \in \text{End}_G(V)$ such that

$$T \equiv \bigoplus_{i' \in [n]} \bigoplus_{i, j \in I_{i'}} \alpha_{ij}^{i'} \mathcal{T}_{ij} \tag{4.2}$$

with $\alpha_{ij}^{i'} \in \mathbb{F}$ we define

$$T \mapsto \bigoplus_{i' \in [n]} (\alpha_{ij}^{i'})_{i, j \in I_{i'}} \tag{4.3}$$

Considering $I_{i'} \subset \mathbb{N}$ we have a natural order defined on these sets and so there's no ambiguity in constructing the matrices M_{n_i} by imposing that same order on its rows and columns.

If we consider the isomorphic spaces above as algebras (with composition of maps as the product operation) then they'll also be isomorphic algebras. To see this we first transform the outer direct sum in 4.2 into an inner direct sum in the usual way by considering

projection and inclusion maps $p_i : V \rightarrow V_i$ and $i : V_i \rightarrow V$ and defining then for $v \in V$

$$\mathcal{T}'_{ij}(v) = i_j(\mathcal{T}_{ij}(p_i(v)))$$

so that $\mathcal{T}'_{ij} : V \rightarrow V$ and

$$T = \sum_{i' \in [n]} \sum_{i, j \in I_{i'}} \alpha_{ij}^{i'} \mathcal{T}'_{ij}$$

Note that the maps $\mathcal{T}'_{ij}$ form a basis of $\text{End}_G(V)$ and here the product of two maps is as usual, that is if we let

$$T_1 = \sum_{i, j \in [m]} \alpha_{ij} \mathcal{T}'_{ij}, \quad T_2 = \sum_{l, k \in [m]} \beta_{lk} \mathcal{T}'_{lk}$$

then

$$T_1 T_2 = \sum_{i, j, l, k \in [m]} \alpha_{ij} \beta_{lk} \mathcal{T}'_{ij} \mathcal{T}'_{lk}$$

The conclusion then follows by noting that

$$\mathcal{T}'_{ij} \mathcal{T}'_{lk} = \delta_{ki} \delta_{l'i'} \mathcal{T}'_{lj}$$

where i', l' are such that $i \in I_{i'}$ and $j \in I_{j'}$ (as previously defined), δ_{ki} is the Kronecker delta, such that $\delta_{ki} = 1$ if $k = i$ and 0 otherwise, and using the definition of matrix multiplication together with the isomorphism defined in 4.3. (it is clear that it preserves the identity, that is $id_V \mapsto \bigoplus_{i \in [n]} id_{\mathbb{F}^{n_i}}$, we usually omit this argument when proving isomorphisms of algebras as it is usually clear).

Corollary 4.2.17. *Let V be a G -space with decomposition as in 4.1, then as algebras*

$$\text{End}_G(V) \cong \bigoplus_{i \in [n]} M_{n_i}(\mathbb{F})$$

where n_i is the multiplicity of the isotypic component V_i , in particular $M_{n_i}(\mathbb{F}) \cong \text{End}_G(n_i V_i)$.

The algebra $\text{End}_G(V)$ is usually called the *commutant* of $_G V$ and corollary 4.2.17 gives the following characterization of multiplicity free representations.

Corollary 4.2.18. *A G -space V is multiplicity free if and only if $\text{End}_G(V)$ is a commutative ring.*

Proof. Indeed if V is multiplicity free then $n_i = 1$ for all i and so $\text{End}_G(V) \cong \mathbb{F}^n$ which is a commutative ring (considering $\mathbb{F}$ is a field so commutative). On the other hand if $\text{End}_G(V)$ is commutative then because $\text{End}_G(V) \cong \bigoplus_{i \in [n]} M_{n_i}(\mathbb{F})$ we must have that each $M_{n_i}(\mathbb{F})$ is commutative which is only possible if $n_i = 1$ thus V is multiplicity free. $\square$

4.3. Character theory

Although we present character theory after introducing concepts of representation theory, historically character theory came first. Initial ideas can be traced back to Gauss, yet the birth of character theory is attributed to F.G. Frobenius and R. Dedekind, in a series of letters exchanged by the two in April of 1896, only then did representation theory started to flourish [Lam98]. Yet these two theories are very much intertwined and the very birth of representation theory of finite groups is also attributed to those same letters.

From now on we let $\mathbb{F} = \mathbb{C}$ and let then V be a vector space over $\mathbb{C}$ unless stated otherwise. To understand what we have to gain we ask a couple still unanswered questions.

- For an abelian group G we know it must have $|G|$ non equivalent irreducible representations, but what about for a non-abelian group?
- Given a representation ρ_V of a group G we know it decomposes into a direct sum of irreducible representations, but how can we determine its decomposition?

Character theory will give us a systematic way of answering these question for any group G , and more. We start with an example, an attempt to in particular answer the first question above for the group S_3 , by brute force.

Example 4.3.1. Let G be the symmetric group S_3 , as can be easily verified this group is generated by the permutations $r = (1\ 2\ 3)$ and $s = (1\ 2)$. Our objective here is to determine the irreducible representations of S_3 . We already know of three such representations, those being the trivial representation, the alternating representation as in example 4.1.15, and by example 4.1.32 we also have a two dimensional irreducible representation λ_U , it being the restriction of the natural permutation representation of S_3 to the space $U = \langle (1, -1, 0), (1, 0, -1) \rangle$. These are clearly seen to be non equivalent as one of them has dimension 2 and the others are one dimensional but not equal. The representation λ_U is equivalent to the representation σ_W , with $W = \mathbb{C}^2$, defined as

$$\sigma_r = \begin{bmatrix} w & 0 \\ 0 & w^2 \end{bmatrix}, \quad \sigma_s = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

where w is the third root of unity $e^{2\pi i/3}$. This can be seen by exploiting the diagonalization of the matrix representation of λ_r where w and w^2 are its eigenvalues. Intuitively σ_W can be seen as the action of S_3 on $\mathbb{R}^2$ as rotations and reflections of the plane that correspond to the symmetries of the triangle, this comes from the fact that S_3 is isomorphic to the

dihedral group of order 6.

Now let ρ_V be a decomposable representation of S_3 . As is known a diagonalizable linear map on a vector space V produces a decomposition of V into a direct sum of subspaces each of which is associated with the eigenvalues of said map, that is given a diagonalizable linear map $\theta : V \rightarrow V$ with eigenvalues $\{\lambda_1, \dots, \lambda_k\}$ the eigenspaces of each eigenvalue λ_i , denoted $V_{\lambda_i} = \{v \in V \mid \theta(v) = \lambda_i v\}$ are direct summands of V and

$$V = \bigoplus_{i \in [k]} V_{\lambda_i}$$

In our case we have that V is not just a vector space but also a G -space, so what we're going to do is exploit the above decompositions for the maps ρ_r and ρ_s in order to gain insight into the decomposition of V as a G -space.

With the above in mind we first note that r has order 3 hence ρ_r has also order 3 considering ρ is decomposable (so not the trivial or alternating representation) and 3 is prime, therefore by proposition C.0.12 and corollary C.0.14 ρ_r is diagonalizable and can have eigenvalues 1, w , w^2 where w is the third root of unity $e^{2\pi i/3}$, as before. thus V decomposes, as a vector space, into eigenspaces as

$$V = V_1 \oplus V_w \oplus V_{w^2}$$

For now let's assume this is the decomposition of V into eigenspaces of ρ_r . Note that these spaces must be invariant to the action of r . For the action of s , since $s = r^{-1}sr^{-1}$ we have:

- if $v \in V_1$ then $\rho(s)v = \rho(r^{-1}sr^{-1})v = \rho(r^{-1}s)v$ considering $\rho(r^{-1})v = \rho(r)^{-1}v = v$, thus $\rho(r^{-1})\rho(s)v = \rho(s)v$ and so $\rho(s)v \in V_1$.
- if $v \in V_w$ then $\rho(s)v = \rho(r^{-1}sr^{-1})v = \rho(r^{-1}s)w^{-1}v \Leftrightarrow \rho(r^{-1})\rho(s)v = w\rho(s)v = w^{-2}\rho(s)v$ thus $\rho(s)v \in V_{w^2}$.

The remaining case where we consider $v \in V_{w^2}$ is analogous and produces $\rho(s)v \in V_w$, hence $\rho(s)$ preserves V_1 but interchanges V_w and V_{w^2} . This tells us that if w is an eigenvalue of ρ_r then w^2 must also be one. Therefore we can conclude that V_1 and $V_w \oplus V_{w^2}$ are G -invariant spaces, we now need to further decompose these spaces into irreducible G -spaces.

Starting with V_1 we now decompose it into $\rho(s)$ eigenspaces, again because s has order 2 then the same goes for $\rho(s)$ and so by corollary C.0.14 it can have eigenvalues 1, -1 ,

hence considering r act trivially on V_1 this space can only decompose into a direct sum of trivial and alternating representations (that is for $v \in V_1$, $\rho(r)v = v$ and either $\rho(s)v = v$ giving the trivial representation or $\rho(s)v = -v$ giving the alternating representation). Hence if w and w^2 are not eigenvalues of ρ_r then ρ decomposes into a direct sum of trivial and alternating representations, depending also on whether ρ has eigenvalues 1 and -1 or just one of them.

For the G -space $V_w \oplus V_{w^2}$ note that because $\rho_s \in GL(V)$ then in particular it is an injective linear map and so by proposition C.0.11 if we assume $\dim_{\mathbb{C}}(V_w) > \dim_{\mathbb{C}}(V_{w^2})$ then a basis of V_w is mapped to a linear independent set in V_{w^2} which can't be the case by hypothesis, on the other hand if $\dim_{\mathbb{C}}(V_w) < \dim_{\mathbb{C}}(V_{w^2})$ then a basis of V_{w^2} is mapped to a linear independent set in V_w which again is a contradiction, hence we must have that $\dim_{\mathbb{C}}(V_w) = \dim_{\mathbb{C}}(V_{w^2})$ and if $(v_1, \dots, v_n)$ is a basis of V_w then $(\rho(s)v_1, \dots, \rho(s)v_n)$ is a basis of V_{w^2} . Consider the spaces $V'_i = \langle v_i, \rho(s)v_i \rangle$, for $i \in [n]$. These are clearly G -invariant spaces, moreover the matrix representations associated with these G -spaces, seen as simply the restriction of ρ to these spaces, are easily seen to be, for $i \in [n]$

$$\rho_{V'_i}(r) = \begin{bmatrix} w & 0 \\ 0 & w^2 \end{bmatrix}, \quad \rho_{V'_i}(s) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

hence $\rho_{V'_i} = \sigma_W$. In conclusion the representation ρ decomposes into a direct sum of the initially observed irreducible representation of S_3 , and as ρ is any decomposable representation we have that those are the only non equivalent irreducible representations of S_3 .

As fun as the above example may have been it's no way to live, and indeed things can be better.

4.3.1 Group characters

The objects in focus in character theory are *group characters*, for abelian groups the group characters are simply homomorphism from G to $\mathbb{C}$ so they can be seen as one dimensional representations of G . In the work of Dedekind and Frobenius the emergence of these objects had seemingly nothing to do with representations, in fact no such thing as representation theory had yet been invented, but it didn't take long for Frobenius to lay out the definition of a finite group representation once he fully grasped group characters. To define what a character is we first recall the definition of the *trace* of a linear map.

Definition 4.3.2. Let V be a vector space over $\mathbb{F}$. The trace on V is a linear map $\text{tr} : \text{End}(V) \rightarrow \mathbb{F}$ (where we consider $\text{End}(V)$ as a vector space over $\mathbb{F}$) such that:

- a) $\text{tr}(AB) = \text{tr}(BA)$ for all $A, B \in \text{End}(V)$
- b) $\text{tr}(\text{id}_V) = \dim_{\mathbb{F}}(V)$

It can be proven that under these condition the map *trace* is unique [Gre75, sec. 7], moreover if $\theta \in \text{End}(V)$ and T_θ is its matrix form then $\text{tr}(\theta) = \text{tr}(T_\theta)$ where the trace on the right hand side is defined on the space $\mathbb{F}^n$ where $n = \dim_{\mathbb{F}}(V)$. This latter statement makes sense considering that by property a) above the trace is independent of the chosen basis, indeed $\text{tr}(ATA^{-1}) = \text{tr}(TA^{-1}A) = \text{tr}(T)$. In more concrete term if θ is a linear map and $T_\theta = (\alpha_{ij})_{i,j \in [n]}$ then the trace of θ can be computed as

$$\text{tr}(\theta) = \text{tr}(T_\theta) = \sum_{i \in [n]} \alpha_{ii}$$

That is the trace of θ is the sum of the elements in the diagonal of T_θ . It's straightforward to verify that this is consistent with the above definition.

Note that if V is an inner product space with inner product $\langle \cdot, \cdot \rangle$ and $B = (v_1, \dots, v_n)$ is an orthonormal basis of V then

$$\text{tr}(\theta) = \sum_{i \in [n]} \langle \theta(v_i), v_i \rangle$$

We can already remark a very interesting property of the trace in the setting of representation theory. Considering that the trace is invariant for conjugation of maps in $\text{End}(V)$ then given two equivalent representation ρ, σ of a group G it must be the case that $\text{tr}(\rho_g) = \text{tr}(\sigma_g)$ for all $g \in G$. As it turn out the converse is also true, yet let's not get ahead of ourselves here, we are now ready to define characters.

Definition 4.3.3. Let ρ_V be a representation of G , the character of ρ is a function $\chi_\rho : G \rightarrow \mathbb{C}$ defined as

$$\chi_\rho(g) = \text{tr}(\rho_g)$$

By this one usually says that the representation ρ affords the character χ_ρ . If we consider V as the corresponding G -space of ρ then we can also denote the character of ρ by χ_V to make the representation/ G -space evident. One usually also says that χ_ρ is a character of G .

Considering the definition above and our initial statement that the group characters of abelian groups are simply homomorphism from G to $\mathbb{C}$, an explanation is in order. In the

initial work of Dedekind with abelian groups, group characters were defined as homomorphism from G to the multiplicative group $\mathbb{C}$, yet in our setting these characters pertain to irreducible representations of G , indeed as we've seen in proposition 4.2.9 irreducible representations of an abelian group are one dimensional, therefore the character of these representation coincides with the representation, thus are homomorphism from G to $\mathbb{C}$ (considering we've stated in the beginning of this section that we assume V to be a vector space over $\mathbb{C}$, unless stated otherwise). As such when we've said group character what we really meant was the following.

Definition 4.3.4. *Let ρ_V be a representation of G , if ρ is irreducible then the character of ρ is said to be irreducible.*

As we've observed before, character are constant in the classes of equivalent representations, as such they naturally transpire our desire to not distinguish between equivalent representations.

Definition 4.3.5. *We denote the set of all irreducible characters of a group G by $\widehat{G}_\chi$. By the properties of the trace we have that $|\widehat{G}_\chi| \leq |\widehat{G}|$.*

Not only do two equivalent representations afford the same character, but as we've said before the converse will also be proven to be true. This result will essentially show that character theory and representation theory are really two faces of the same coin, by showing that characters uniquely determine representations up to equivalence, as such many results about representations are obtained by focusing solely on characters. For this some terminology about representations is transported to characters (or perhaps the other way around historically) as in definition 4.3.4.

Definition 4.3.6. *Let ρ_V be a representation, then the degree of its character χ_ρ is $\dim_{\mathbb{F}}(V)$.*

That is the *degree* of a character equals the degree/dimension of its underlying representation. As stated before a character of degree 1 is a homomorphism $\chi : G \rightarrow \mathbb{C}$, these are usually called *linear* characters.

Definition 4.3.7. *A character χ_ρ is faithful if ρ is a faithful representation.*

The same way we have the trivial, alternating and regular representations we also have the corresponding trivial, alternating and regular characters, these of course depend only on the group we're working with.

Recall that a matrix is diagonalizable if it is similar to a diagonal matrix, that is a matrix A is diagonalizable if $A = PBP^{-1}$ for some non singular matrix P and diagonal matrix B .

Lemma 4.3.8. *Let ρ be a matrix representation of a group G , then ρ_g is a diagonalizable matrix for all $g \in G$. In particular ρ_g is similar to a diagonal matrix whose entries are roots of unity.*

Proof. Considering G is a finite group then every element of G has finite order and so for all $g \in G$, $\rho(g)$ has finite order, hence by proposition C.0.12 and corollary C.0.14 $\rho(g)$ is diagonalizable and its eigenvalues are roots of unity. As $\rho(g)$ is then similar to a diagonal matrix whose entries are the eigenvalues of $\rho(g)$ (repeated according to their multiplicity) the result follows. $\square$

Proposition 4.3.9. *Let ρ_V be a representation of G affording character χ_ρ .*

- a) *Let $\theta \in \text{GL}(V)$, then for all $g \in G$, $\chi_\rho(g) = \text{tr}(\rho_g) = \text{tr}(\theta^{-1}\rho_g\theta)$. In particular for all $h \in G$, $\chi_\rho(hgh^{-1}) = \chi_\rho(g)$*
- b) *For all $g \in G$, $\chi_\rho(g)$ is equal to the sum of its eigenvalues, counting with multiplicities. That is if ρ_g has eigenvalues $\{\lambda_1, \dots, \lambda_k\}$ with corresponding multiplicities $\{m_1, \dots, m_k\}$ then*

$$\chi_\rho(g) = \sum_{i \in [k]} m_i \lambda_i$$

In particular $\chi_\rho(g)$ is the sum of n th roots of unity where n is the order of ρ_g in $GL(V)$.

- c) *For all $g \in G$, $\chi_\rho(g^{-1})$ equals the complex conjugate of $\chi_\rho(g)$, that is $\chi_\rho(g^{-1}) = \overline{\chi_\rho(g)}$.*

Proof. a) This follows immediately from the definition 4.3.2 of trace.

b) Immediate considering the previous lemma and part a) of this proposition.

c) Considering the eigenvalues of $\rho_{g^{-1}}$ are the inverses of the ones of ρ_g and inverses of roots of unity are equal to complex conjugates we have by b) that

$$\chi_\rho(g^{-1}) = \sum_{i \in [k]} m_i \lambda_i^{-1} = \sum_{i \in [k]} m_i \overline{\lambda_i} = \overline{\sum_{i \in [k]} m_i \lambda_i} = \overline{\chi_\rho(g)}$$

$\square$

Corollary 4.3.10. *Let ρ_V be a representation of G affording character χ_ρ . Then for $g \in G$*

$$|\chi_\rho(g)| \leq \chi_\rho(1_G)$$

and we have equality if and only if $\rho_g = \lambda id_V$ for some $\lambda \in \mathbb{C}$ a root of unity. Moreover

$$\ker(\rho) = \{g \in G \mid \chi_\rho(g) = \chi_\rho(1_G)\}$$

Proof. Considering part b) of the previous proposition and the triangle inequality for complex numbers we have that

$$|\chi_\rho(g)| = \left| \sum m_i \lambda_i \right| \leq \sum m_i |\lambda_i| = \sum m_i = \dim_{\mathbb{C}}(V) = \chi_\rho(1_G)$$

and we have equality if and only if all λ_i are equal to, say λ which must be a root of unity by the λ_i being so. Considering this if $\chi_\rho(g) = \chi_\rho(1_G)$ then it must be the case that $\lambda = 1$ and so $\rho_g = id_V$, that is $g \in \ker(\rho)$, and the other direction follows immediately. $\square$

By the above corollary we can conclude that the character of a representation fully characterizes its kernel, which is a normal subgroup of G , thus characters can be a valuable tool in studying the normal subgroups of a group G .

As characters are functions that map to the field $\mathbb{C}$, we can define operations on them.

Definition 4.3.11. On the set of all characters of a group G we define two operations, addition and multiplication, defined as, for χ_ρ, χ_σ

$$[\chi_\rho + \chi_\sigma](g) = \chi_\rho(g) + \chi_\sigma(g), \quad [\chi_\rho \cdot \chi_\sigma](g) = \chi_\rho(g)\chi_\sigma(g)$$

This results in the following.

Proposition 4.3.12. Let χ_ρ, χ_σ be character of a group G . Then

- a) $\chi_\rho + \chi_\sigma = \chi_{\rho \oplus \sigma}$, that is $\chi_\rho + \chi_\sigma$ is the character of $\rho \oplus \sigma$.
- b) $\chi_{\rho^\circ} = \overline{\chi_\rho}$ where ρ° is the adjoint representation of ρ . Moreover if χ_ρ is irreducible then so is χ_{ρ° .
- c) $\chi_\rho \chi_\sigma = \chi_{\rho \otimes \sigma}$, that is $\chi_\rho \chi_\sigma$ is the character of $\rho \otimes \sigma$, where here we consider the inner tensor product.
- d) If $G = G_1 \times G_2$ and $\rho = \rho_1 \otimes \rho_2$ with ρ_1, ρ_2 representations of the groups G_1 and G_2 respectively, so that here we're considering an outer tensor product, then for $(g_1, g_2) \in G$, $\chi_{\rho_1 \otimes \rho_2}(g_1, g_2) = \chi_{\rho_1}(g_1)\chi_{\rho_2}(g_2)$.

Proof. a) This follows immediately from proposition 4.1.35.

b) This is also clear considering proposition 4.1.46 and the irreducibility follows from our discussion preceding definition 4.1.40.

c) By proposition 4.1.47, and adopting the same notation, we have that

$$\chi_{\rho \otimes \sigma}(g) = \sum_{i,j} \varphi_{ij,ij}^{\rho \otimes \sigma}(g) = \sum_{i,j} \varphi_{ii}^{\rho}(g) \varphi_{jj}^{\sigma}(g) = \chi_{\rho}(g) \chi_{\sigma}(g)$$

d) Again by proposition 4.1.47 we have that

$$\chi_{\rho_1 \otimes \rho_2}(g_1, g_2) = \sum_{i,j} \varphi_{ij,ij}^{\rho_1 \otimes \rho_2}(g_1, g_2) = \sum_{i,j} \varphi_{ii}^{\rho_1}(g_1) \varphi_{jj}^{\rho_2}(g_2) = \chi_{\rho_1}(g_1) \chi_{\rho_2}(g_2)$$

□

The above proposition and corollary 4.2.4 produce a most interesting observation.

Corollary 4.3.13. *A character χ_{ρ} of a group G decomposes into a sum of irreducible characters of G according to the decomposition of ρ into irreducible representations.*

As such the set $\widehat{G}_{\chi}$ will be of major focus to us, as the set $\widehat{G}$ is in representation theory. When the group G is abelian then $\widehat{G}_{\chi}$ together with the product operation previously defined on characters gains an abelian group structure. This is straightforward to verify as associativity and commutativity follow from the field $\mathbb{C}$, by part c) of proposition 4.3.12 the product of two characters of G is also a character of G and the inverse of a character χ is $\bar{\chi}$ considering that if $\chi : G \rightarrow \mathbb{C}$ is a homomorphism then by G being finite it must be the case that $\chi(g)$ is a root of unity for all $g \in G$. The identity of this group is clearly the trivial character.

Another interesting property of irreducible characters of abelian groups is that, for χ a non trivial such character

$$\sum_{g \in G} \chi(g) = 0 \tag{4.4}$$

To see this note that, letting $h \in G$

$$\chi(h) \sum_{g \in G} \chi(g) = \sum_{g \in G} \chi(h) \chi(g) = \sum_{g \in G} \chi(hg) = \sum_{g' \in G} \chi(g')$$

where the second to last equality follows from χ being a homomorphism. Hence $(\chi(h) - 1) \sum_{g \in G} \chi(g) = 0$ and as $\mathbb{C}$ is a field it must be the case that either $(\chi(h) - 1) = 0$ or $\sum_{g \in G} \chi(g) = 0$. As h is any in G and χ is non trivial it must then be the case that $\sum_{g \in G} \chi(g) = 0$.

It is also possible to define an inner product of characters of a group G , this follows from the fact that such characters are elements of the space of complex function $f : G \rightarrow \mathbb{C}$, denoted $\mathbb{C}[G]$. It's straightforward to verify that this is indeed a vector space over $\mathbb{C}$. The notation $\mathbb{C}[G]$ alludes to the fact that $\mathbb{C}[G]$ is isomorphic to the group algebra $\mathbb{C}G$ seen simply as a vector space, and it's possible to define a product on $\mathbb{C}[G]$ analogous to the product on $\mathbb{C}G$ such that $\mathbb{C}[G]$ is then an algebra isomorphic to $\mathbb{C}G$.

Definition 4.3.14. *On the space $\mathbb{C}[G]$ of complex functions from G to $\mathbb{C}$ we define an inner product $\langle \cdot, \cdot \rangle_G$ as*

$$\langle f_1, f_2 \rangle_G = \frac{1}{|G|} \sum_{g \in G} f(g) \overline{f_2(g)}$$

for $f_1, f_2 \in \mathbb{C}[G]$.

Note that for the particular case of characters, by part b) and c) of proposition 4.3.12 we have that $\langle \chi_\rho, \chi_\sigma \rangle_G = \langle \chi_{\rho \otimes \sigma^\circ}, \chi_i \rangle_G$, where χ_i is the trivial character and ρ, σ are representations of G .

It's easy to see that the space $\mathbb{C}[G]$ has dimension $|G|$, indeed any element of this space can be seen as a $|G|$ -tuple, or equivalently a vector in the space $\mathbb{C}^{|G|}$, and any vector in $\mathbb{C}^{|G|}$ can be seen as a function from G to $\mathbb{C}$. This correspondence produces a vector space isomorphism in a straightforward way. From this it becomes clear that the above inner product differs only by a constant from the usual inner product used on $\mathbb{C}^n$.

By part a) of proposition 4.3.9 we remark that characters are what is usually called *central/class functions*, these function are invariant under conjugation in G of the argument, or in other words they are constant in the conjugacy classes of G . The set of all class functions from G to $\mathbb{C}$ form a subspace of $\mathbb{C}[G]$ we denote by $C(G)$, thus

$$C(G) = \{f \in \mathbb{C}[G] \mid f(hgh^{-1}) = f(g) \text{ for all } g, h \in G\}$$

As such the set $\widehat{G}_\chi$ is a subset of $C(G)$.

Proposition 4.3.15. *If G is abelian then the set $\widehat{G}_\chi$ is an orthonormal basis of $\mathbb{C}[G]$.*

Proof. Letting $\|f\| = \sqrt{\langle f, f \rangle_G}$ then it's clear that for $\chi \in \widehat{G}_\chi$, $\|\chi\| = 1$ from the definition of inner product and the fact that $\chi(g)$ is a root of unity so has norm 1 as a complex number. Now note that for $\xi \in \widehat{G}_\chi$ different than χ

$$\langle \chi, \xi \rangle_G = \sum_{g \in G} \chi(g) \overline{\xi(g)}$$

By part c) of proposition 4.3.12 $\chi\bar{\xi}$ is a character of G and so by 4.4, considering χ and ξ are distinct and so $\chi\bar{\xi}$ is non trivial, we have that $\langle \chi, \xi \rangle_G = 0$.

To finish the proof note that $\mathbb{C}[G]$ must have dimension equal to G and by proposition 4.2.10 this is also the cardinality of $\widehat{G}_\chi$, hence the conclusion follows. $\square$

This essentially produces the Fourier Analysis for complex functions $\mathbb{C}[G]$ where G is abelian, and the main objective is to write any function in $\mathbb{C}[G]$ as a linear combination of elements of $\widehat{G}_\chi$, which follows easily from the above proposition. The basis $\widehat{G}_\chi$ is hailed for having very good properties. More generally for non-abelian groups G we'll see that although $\widehat{G}_\chi$ is no longer a basis of $\mathbb{C}[G]$, it will still be a basis of the space of class functions of $C(G)$ and instead it will be the matrix coefficients of representations of G that produce a basis of $\mathbb{C}[G]$, thus producing a Fourier Analysis for non-abelian groups. This will follow easily from our exposition of Character Theory.

Example 4.3.16. We compute explicitly here the character of the regular representation λ of a group G . This character and the corresponding representations are very important in representation theory as we'll see, so it's very useful to have it at hand. Although useful these characters are very simple and easy to compute. First we recall that the regular representation is determined by the action of G on itself by left multiplication, specifically it's given by the action of G on a vector space V of dimension $|G|$ with basis B indexed by the elements of G , $B = \{e_g\}_{g \in G}$ (which can be seen as being the group algebra $\mathbb{F}G$), such that $ge_h = e_{gh}$.

Now as we know the matrix representation of λ is composed of permutation matrices. For 1_G the matrix of $\lambda(1_G)$ is trivially the identity on $\mathbb{C}^{|G|}$ and so $\chi_\lambda(1_G) = |G|$. For the other elements of G note that, letting then $g \in G \setminus \{1_G\}$, if the respective matrix of $\lambda(g)$ as a 1 in a certain position in its diagonal (it must be a 1 considering it's a permutation matrix) then there exists $h \in G$ such that $gh = g$, but then because G is a group we must have that $g = 1_G$ and so $\chi_\lambda(g) = 0$.

The above example can be generalized to permutation representations in an analogous way.

Proposition 4.3.17. *Let a group G act on a finite set X and consider the corresponding permutation representation λ as defined in 4.1.17, then the corresponding character χ_λ is given by*

$$\chi_\lambda(g) = |\{x \in X \mid gx = x\}|$$

for $g \in G$.

Proof. All we have to do is note that the matrix representation of λ according to the basis X of the space $\mathbb{C}X$ is composed of permutation matrices, and for a $g \in G$ the matrix of $\lambda(g)$ can only have a 1 in some position in its diagonal if for the corresponding $x \in X$, $gx = x$, hence

$$\chi_\lambda(g) = |\{x \in X \mid gx = x\}|$$

□

4.3.2 Orthogonality of group characters

We've seen that when G is an abelian group then the group characters $\widehat{G}_\chi$ is a set of orthogonal elements in the space $\mathbb{C}[G]$ of complex functions. One now wonders what can we say about $\widehat{G}_\chi$ when G is non-abelian. As it turn out it will still be a set of orthogonal elements in $\mathbb{C}[G]$. We'll prove this in this subsection in the simplest of ways and then explore its consequences.

To prove orthogonality of group characters we first prove a stronger result regarding the matrix coefficients of irreducible representations, then the rest will follow easily. Recall definition 4.1.45 of matrix coefficients, in our case where $\mathbb{F} = \mathbb{C}$ it is clear the matrix coefficients are complex functions in $\mathbb{C}[G]$ so the inner product defined in 4.3.14 makes sense here.

Theorem 4.3.18. *Let ρ_V, σ_W be two irreducible representation of a group G and R, S the respective matrix representations according to some basis of V and W . Denoting the matrix coefficients of R by r_{ij} and similarly for S , s_{ij} then*

- a) *If ρ and σ are non equivalent then $\langle r_{ij}, s_{lk} \rangle_G = 0$ for all i, j, k, l .*
- b) *Let $d = \dim_{\mathbb{C}}(V)$, then*

$$\langle r_{ij}, r_{lk} \rangle_G = \begin{cases} \frac{1}{d} & \text{if } (i, j) = (k, l) \\ 0 & \text{otherwise} \end{cases}$$

Proof. Let $\theta : V \rightarrow W$ be a linear map and consider the map

$$\tilde{\theta} = \frac{1}{|G|} \sum_{g \in G} \sigma_g^{-1} \theta \rho_g$$

It's clear this is also a linear map/homomorphism from V to W , moreover it's straightforward to verify that this is also a G -homomorphism. Considering V and W are irreducible G -spaces then by Schur's lemma we must have that, if ρ is not equivalent to σ then

$$\tilde{\theta} = 0 \tag{4.5}$$

On the other hand if $\sigma = \rho$ and so $V = W$ then again by Schur's lemma it must be the case that $\tilde{\theta} = cid_V$ for some $c \in \mathbb{C}$. It's easy to determine the constant c , indeed we have that $\text{tr}(\tilde{\theta}) = cd$ where $d = \dim_{\mathbb{C}}(V)$, but by the definition of $\tilde{\theta}$ we also have that

$$\text{tr}(\tilde{\theta}) = \frac{1}{|G|} \sum_{g \in G} \text{tr}(\sigma_g^{-1} \theta \rho_g) = \frac{1}{|G|} \sum_{g \in G} \text{tr}(\theta) = \text{tr}(\theta)$$

Hence we have that $c = \text{tr}(\theta)/d$ and so

$$\tilde{\theta} = \frac{\text{tr}(\theta)}{d} id_V \tag{4.6}$$

a) Let T and $\tilde{T}$ be the matrix form of the maps θ and $\tilde{\theta}$ respectively according to the initially selected basis for V and W . If ρ and σ are not equivalent then we are in the situation 4.5 and so

$$0 = \tilde{T}_{ij} = \frac{1}{G} \sum_{g \in G} \sum_{\substack{p \in [b], \\ q \in [d]}} s_{ip}(g^{-1}) T_{pq} r_{qj}(g)$$

where $d = \dim_{\mathbb{C}}(W)$.

Note that we are free to choose the map θ as long as it is a linear map, as such let then $T_{ij} = 1$ if $(i, j) = (k, l)$ for some indices k, l and 0 otherwise. Then we get

$$0 = \tilde{T}_{ij} = \frac{1}{G} \sum_{g \in G} s_{ik}(g^{-1}) r_{lj}(g)$$

Hence because i, j, k, l are any indices we get that for all i, j, k, l

$$\frac{1}{G} \sum_{g \in G} s_{ik}(g^{-1}) r_{lj}(g) = 0$$

Now recall that for vector spaces over $\mathbb{C}$ we can consider the matrix representations as being unitary (theorem 4.1.12) and so $s_{ik}(g^{-1}) = \overline{s_{ki}(g)}$ therefore we get

$$0 = \frac{1}{G} \sum_{g \in G} \overline{s_{ki}(g)} r_{lj}(g) = \langle r_{lj}, s_{ki} \rangle_G$$

- b) When $\sigma = \rho$ and so $V = W$ we are in situation 4.6 and following the proof of part a), if we let $T_{ij} = 1$ if $(i, j) = (k, k)$ for some index $k \in [d]$ and 0 otherwise we get

$$\frac{1}{G} \sum_{g \in G} \overline{r_{ki}(g)} r_{kj}(g) = \frac{\text{tr}(\theta)}{d} = \frac{1}{d}$$

if $i = j$ and

$$\frac{1}{G} \sum_{g \in G} \overline{r_{ki}(g)} r_{kj}(g) = 0$$

if $i \neq j$. Applying the definition of the inner product 4.3.14 as in part a) of this theorem the conclusion follows immediately.

□

Corollary 4.3.19. *Let ρ_V, σ_W be two irreducible representation of a group G affording characters χ_ρ, χ_σ respectively, then*

$$\langle \chi_\rho, \chi_\sigma \rangle_G = \begin{cases} 1 & \text{if } \rho \equiv \sigma \\ 0 & \text{otherwise} \end{cases}$$

Proof. All we have to do here is note that, using notation from the previous theorem, $\chi_\rho = \sum_{i \in [d]} r_{ii}$ and similarly goes for χ_σ , thus by the linearity of inner products

$$\langle \chi_\rho, \chi_\sigma \rangle_G = \sum_{\substack{i \in [d], \\ j \in [b]}} \langle r_{ii}, s_{jj} \rangle_G$$

So if the representations are non equivalent then by the previous theorem $\langle r_{ii}, s_{jj} \rangle_G = 0$ for all i, j so $\langle \chi_\rho, \chi_\sigma \rangle_G = 0$. If the representations are equivalent then because characters of equivalent representations coincide we have $\chi_\sigma = \chi_\rho$ and $\langle r_{ii}, s_{jj} \rangle_G = \langle r_{ii}, r_{jj} \rangle_G$ so by the previous theorem we get

$$\langle \chi_\rho, \chi_\sigma \rangle_G = \sum_{i \in [d]} \langle r_{ii}, r_{ii} \rangle_G = d \frac{1}{d}$$

□

With this we can easily get a similar result to 4.4 for non-abelian groups.

Corollary 4.3.20. ** Let χ_ρ be a non trivial irreducible character of a group G , then*

$$\sum_{g \in G} \chi_\rho(g) = 0$$

*Later on in theorem 4.7.19 we prove a more general result not requiring ρ to be irreducible.

Proof. Indeed

$$\sum_{g \in G} \chi_\rho(g) = |G| \langle \chi_\rho, \chi_\iota \rangle_G$$

Where χ_ι is the character of the trivial representation ι , thus the conclusion follows from the previous corollary. □

Corollary 4.3.21. *Let ρ_V be a representation of a group G with canonical decomposition*

$$V = \bigoplus_{i \in [n]} n_i V_i$$

and χ_{V_i} be the character of the G -space V_i , $i \in [n]$, then

$$\langle \chi_{V_i}, \chi_\rho \rangle_G = n_i = \dim_{\mathbb{C}}(\text{Hom}_G(V_i, V))$$

More generally if σ_W is another representation of G then

$$\langle \chi_\sigma, \chi_\rho \rangle_G = \dim_{\mathbb{C}}(\text{Hom}_G(W, V))$$

In particular

$$\langle \chi_\rho, \chi_\rho \rangle_G = \sum_{i \in [n]} n_i^2 = \dim_{\mathbb{C}}(\text{Hom}_G(V, V))$$

The poof of the above corollary is immediate considering proposition 4.2.12, theorem 4.2.16, corollary 4.3.13 and corollary 4.3.19. Ultimately the above results produce the following remarkable theorem, making good on our promises. We recall that so far our work is true for linear representations over finite dimensional vector spaces over $\mathbb{C}$ of finite groups, as in the next theorem.

Theorem 4.3.22. *Two representations over $\mathbb{C}$ are equivalent if and only if their characters are equal. In particular $|\widehat{G}| = |\widehat{G}_\chi|$ and representations over $\mathbb{C}$ are characterized by their characters.*

Proof. As we've already remarked prior to definition 4.3.3 if two representations are equivalent then they have the same character. On the other hand let ρ_V and σ_W be the two representations, by corollary 4.3.13 their characters decompose into a sum of irreducible characters of G according to the decomposition of ρ and σ into irreducible representations, respectively. As such, if their characters are equal then letting ξ_{V_i} be an irreducible character in the decomposition of χ_ρ we have by the previous corollary that

$$n_i = \langle \xi, \chi_\rho \rangle_G = \langle \xi, \chi_\sigma \rangle_G$$

where n_i is the multiplicity of V_i in the decomposition of ρ into irreducibles. As such considering corollary 4.3.19 the decomposition of ρ and σ must coincide, up to equivalence of representations and order, and so both representations must be equivalent. $\square$

By the above theorem given a representation ρ_V of G and an irreducible representation $\sigma \in \widehat{G}$ we can determine the multiplicity of σ in ρ just by computing the inner product $\langle \chi_\sigma, \chi_\rho \rangle_G$, thus we can determine the decomposition into irreducibles of V from $\widehat{G}$. Or equivalently determine the decomposition into irreducibles of χ_ρ from $\widehat{G}_\chi$.

The theorem also provides another proof of the uniqueness of the decomposition of a representation into irreducibles, and the proof of this is just a rephrasing of the proof given for the theorem above.

Note that for the case of infinite groups the theorem fails, indeed if we let $G = \mathbb{Z}$, the group of integers, then the representations defined by

$$1 \mapsto \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad 1 \mapsto \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

are not equivalent, yet have the same character.

Example 4.3.23. We can now go back to example 4.3.1 and apply the character theory developed so far to determine the decomposition of a representation of S_3 .

Recall we had determined that the group S_3 has 3 irreducible representations, those being the trivial and alternating representations ι and ϵ respectively and a two dimension representation we denoted by σ . Let's consider the regular representation λ of $S_3 = \{1_{S_3}, (12), (13), (23), (123), (132)\}$, to simplify notation let $S_3 = \{g_1, g_2, \dots, g_6\}$ where each g_i pertains to the element in position $i \in [6]$ in the aforementioned list of elements of S_3 . By corollary 4.3.21 we need only to find the characters of the aforementioned representations and the rest follows by simply computing inner products.

The matrix representation of λ is composed of permutation matrices and it's straightforward to see that the character of λ is $\chi_\lambda = (6, 0, 0, 0, 0, 0)$ considering that by example 4.3.16 in the diagonal of said matrices there are only zeros, with the exception of $\lambda(1_{S_3})$ which is the identity map on a space of dimension 6. The character of the trivial representation is of course $\chi_\iota(g_i) = 1$ for all $i \in [6]$ and for the alternating representation we have the character $\chi_\epsilon = (1, -1, -1, -1, 1, 1)$.

For the representation σ recall that by example 4.3.1 we had $r = (123)$ and $s = (12)$ thus $(13) = rs$, $(23) = sr$, $(132) = r^2$ and with the identity we have all the elements of

S_3 , moreover we had already given the matrices of σ_s and σ_r thus we now need only to compute products of matrices to get the matrix representation of σ and then compute the sum of the element in the diagonal of said matrices to get the character, which results in $\chi_\sigma = (2, 0, 0, 0, w + w^2, w^2 + w^4) = (2, 0, 0, 0, -1, -1)$. We can organize this irreducible characters in a table as

	g_1	g_2	g_3	g_4	g_5	g_6
χ_ι	1	1	1	1	1	1
χ_ϵ	1	-1	-1	-1	1	1
χ_σ	2	0	0	0	-1	-1

Now by corollary 4.3.21 we have only to compute inner products of characters to get the answers we seek. This is straightforward so we omit said computations, the result is

$$\langle \chi_\iota, \chi_\lambda \rangle_G = 1$$

$$\langle \chi_\epsilon, \chi_\lambda \rangle_G = 1$$

$$\langle \chi_\sigma, \chi_\lambda \rangle_G = 2$$

Note that each irreducible representation has multiplicity equal to its dimension, as we'll see next this is no coincidence for a regular representation. In conclusion we have that λ decomposes as

$$\lambda = \iota \oplus \epsilon \oplus 2\sigma$$

We still relied on example 4.3.1 to know what all the irreducible representations of S_3 were, in the next subsection we'll break free from this.

4.3.3 Consequences

Several remarkable results emerge from the orthogonality of group characters, the following proposition shows why we are so interested in regular representations.

Proposition 4.3.24. *The decomposition into irreducibles of the regular representation λ of a group G contains every irreducible representation of G with multiplicity equal to its dimension.*

Proof. By example 4.3.16 we know that $\chi_\lambda(g) = |G|$ if $g = 1_G$ and 0 otherwise hence for $\xi \in \widehat{G}_\chi$

$$\langle \xi, \chi_\lambda \rangle_G = \frac{1}{|G|} \xi(1_G) \chi_\lambda(1_G) = \xi(1_G)$$

Now by part b) of definition 4.3.2 and corollary 4.3.21 the conclusion follows immediately. $\square$

Considering the above proposition and corollary 4.2.4, in particular that any representation decomposes into a finite direct sum, we have the following immediate corollary.

Corollary 4.3.25. *For a group G the set $\widehat{G}$, and consequently $\widehat{G}_\chi$ is finite.*

Corollary 4.3.26. *Let G be a group and consider its group characters $\widehat{G}_\chi$, then*

$$|G| = \sum_{\xi \in \widehat{G}_\chi} \xi(1_G)^2$$

Proof. By the previous proposition the regular representation of G decomposes as $\lambda = \bigoplus_{\rho \in \widehat{G}} d_\rho \rho$ where d_ρ is the dimension of the irreducible representation $\rho \in \widehat{G}$, thus by corollary 4.3.13 we have that, for $g \in G$

$$\chi_\lambda(g) = \sum_{\rho \in \widehat{G}} d_\rho \chi_\rho(g)$$

Now by example 4.3.16 $\chi_\lambda(1_G) = |G|$ and considering that $d_\rho = \chi_\rho(1_G)$ then

$$|G| = \chi_\lambda(1_G) = \sum_{\rho \in \widehat{G}} d_\rho \chi_\rho(1_G) = \sum_{\rho \in \widehat{G}} \chi_\rho(1_G)^2 = \sum_{\xi \in \widehat{G}_\chi} \xi(1_G)^2$$

$\square$

Theorem 4.3.27. *The matrix coefficients $\{\varphi_{ij}^\rho \mid \rho \in \widehat{G}; i, j \in \dim_{\mathbb{C}}(\rho)\}$ form a basis of the space of complex functions $\mathbb{C}[G]$.*

Proof. Indeed by the previous corollary we know that

$$|\{\varphi_{ij}^\rho \mid \rho \in \widehat{G}; i, j \in [d_\rho]\}| = \sum_{\rho \in \widehat{G}} d_\rho^2 = |G|$$

where d_ρ is the dimension of the irreducible representation $\rho \in \widehat{G}$, hence by theorem 4.3.18 because the matrix coefficients are orthogonal elements of $\mathbb{C}[G]$ and there are $|G|$ of them the conclusion follows by noting that the dimension of $\mathbb{C}[G]$ is also $|G|$. $\square$

The irreducibility of a representation can be characterized in terms of its character.

Theorem 4.3.28. *Let ρ_V be a representation of a group G affording the character χ_ρ , then ρ is irreducible if and only if $\langle \chi_\rho, \chi_\rho \rangle_G = 1$.*

Proof. The sufficient implication is given by corollary 4.3.19, for the necessary implication, by corollary 4.3.13 we can decompose χ_ρ according to the canonical decomposition of ρ as

$$\chi_\rho = \sum_{i \in [n]} n_i \chi_{V_i}$$

Hence by corollary 4.3.21 we get

$$\langle \chi_\rho, \chi_\rho \rangle_G = \sum_{i \in [n]} n_i^2$$

But then if $\langle \chi_\rho, \chi_\rho \rangle_G = 1$ considering each $n_i \in \mathbb{N}_0$ it must be the case that $n_i = 1$ for some $i \in [n]$ with the remaining $n_j = 0$ and so ρ must be irreducible. $\square$

Self-dual representations, as defined in 4.1.40, can also be characterized by their characters.

Proposition 4.3.29. *A representation ρ is self-dual if and only if $\chi_\rho = \overline{\chi_\rho}$, i.e. χ_ρ is a real-valued function.*

Proof. By definition a representation ρ is self-dual if $\rho \equiv \rho^\circ$, that is the representation is equivalent to its dual, thus their characters must coincide, so by part b) of proposition 4.3.12, $\chi_\rho = \chi_{\rho^\circ} = \overline{\chi_\rho}$. The other direction follows again by part b) of proposition 4.3.12 and by theorem 4.3.22. $\square$

Considering this and what we know about the regular characters, that they are real functions, the following corollary follows easily.

Corollary 4.3.30. *The regular representation of a group G is self-dual.*

Corollary 4.3.21 allows us to prove a very fundamental result in representation theory quite easily.

Theorem 4.3.31. *Let V, W be G -spaces then*

$$\text{Hom}_G(V, W) \cong \text{Hom}_G(W, V)$$

as vector spaces over $\mathbb{C}$.

Proof. Indeed by the aforementioned corollary $\dim_{\mathbb{C}}(\text{Hom}_G(V, W)) = \langle \chi_V, \chi_W \rangle_G$, moreover by the properties of inner products

$$\langle \chi_V, \chi_W \rangle_G = \overline{\langle \chi_W, \chi_V \rangle_G} = \overline{\dim_{\mathbb{C}}(\text{Hom}_G(W, V))}$$

As $\dim_{\mathbb{C}}(\text{Hom}_G(W, V))$ is a real number we conclude both spaces have the same dimension and so are isomorphic vector spaces. $\square$

We now arrive at the crown jewel of this subsection. We've already seen that $\widehat{G}_\chi$ is a finite set, and as we've stated before the set $\widehat{G}_\chi$ is a basis of the space of class functions $C(G)$, so once we prove this fact it'll be clear what $|\widehat{G}_\chi|$ is.

We start by noting that, as $C(G)$ is composed of functions constant in the conjugacy classes of G , its dimension must be equal to the number of conjugacy classes of G . If we let $\{J_1, J_2, \dots, J_k\}$ be the conjugacy classes of G then we can easily produce a basis of $C(G)$ by considering the functions $\delta_i : G \rightarrow \mathbb{C}$, $i \in [k]$, defined as

$$\delta_i(g) = \begin{cases} 1 & \text{if } g \in J_i \\ 0 & \text{otherwise} \end{cases} \quad (4.7)$$

Indeed it is clear that any function in $C(G)$ can be expressed as a linear combination of these functions and no less. These functions are called the *characteristic functions* of the conjugacy classes of G , the value k is called the *conjugacy number* of G .

Theorem 4.3.32. *For a group G the group characters $\widehat{G}_\chi$ form a basis of $C(G)$, in particular $|\widehat{G}|$ is equal to the number of conjugacy classes of G , that is the conjugacy number of G .*

Proof. Consider the group characters $\widehat{G}_\chi$, as we know these are elements of $C(G)$, moreover as $C(G)$ is a vector space the space generated by $\widehat{G}_\chi$ must have an orthogonal complement in $C(G)$, we will prove that this orthogonal space is in fact trivial, thus $\widehat{G}_\chi$ spans $C(G)$ and as the group characters are orthogonal it must be the case that $\widehat{G}_\chi$ is a basis for $C(G)$.

To that end let $f \in C(G)$ such that

$$\langle f, \xi \rangle_G = 0 \quad (4.8)$$

for all $\xi \in \widehat{G}_\chi$, we need to prove that $f = 0$. Consider an irreducible representation ρ_V of G affording character $\chi \in \widehat{G}_\chi$, with it we construct the linear map from V to V

$$\theta = \frac{1}{|G|} \sum_{g \in G} \overline{f(g)} \rho_g$$

We note that θ is a G -homomorphism, indeed it is clearly a linear map by ρ_g being so for all $g \in G$, moreover

$$\rho_{h^{-1}} \theta \rho_h = \frac{1}{|G|} \sum_{g \in G} \overline{f(g)} \rho(h^{-1}gh) = \frac{1}{|G|} \sum_{g \in G} \overline{f(h^{-1}gh)} \rho(h^{-1}gh) = \theta$$

where the second to last equality comes from f being a class function. Because V is a simple G -space then by Schur's lemma we must have that $\theta = cid_V$ for some $c \in \mathbb{C}$, to find c note that

$$c \dim_{\mathbb{C}}(V) = \text{tr}(\theta) = \text{tr} \left(\frac{1}{|G|} \sum_{g \in G} \overline{f(g)} \rho_g \right) = \frac{1}{|G|} \sum_{g \in G} \overline{f(g)} \chi(g) = \langle \chi, f \rangle_G = 0$$

considering $\chi \in \widehat{G}_\chi$ and $\langle \chi, f \rangle_G = \overline{\langle f, \chi \rangle_G}$ so by 4.8 $\langle \chi, f \rangle_G = 0$, hence $c = 0$ thus $\theta = 0$. As this is true for any irreducible representation then by corollary 4.2.4 it is true for any representation of G . If we consider in particular the regular representation λ of G then we have that

$$0 = \theta(1_G) = \frac{1}{|G|} \sum_{g \in G} \overline{f(g)} \lambda_g(1_G) = \frac{1}{|G|} \sum_{g \in G} \overline{f(g)} g$$

Now as G is a basis of $\mathbb{C}G$ then it must be the case that $\overline{f(g)} = 0$ for all $g \in G$ and so $f = 0$.

Therefore $\widehat{G}_\chi$ is a basis of $C(G)$, now as we remarked before the space $C(G)$ has dimension equal to the conjugacy number of G , hence it must be the case that $|\widehat{G}_\chi|$ is equal to the conjugacy number of G . □

Knowing this, if we now go back to example 4.3.1, once we had found three non equivalent irreducible representations of S_3 we could immediately conclude these were all the irreducible representations of S_3 by simply noting that S_3 has three conjugacy classes (this comes from the well known fact that the conjugacy classes of the symmetric group S_n are in bijection with possible cycle structures of the permutations in S_n [Rot95, 3.5]).

Corollary 4.3.33. *Let $g_1, g_2 \in G$, then these elements are conjugate in G if and only if $\chi(g_1) = \chi(g_2)$ for all $\chi \in \widehat{G}_\chi$.*

Proof. If g_1 and g_2 are conjugate then because characters are class function it must be the case that $\chi(g_1) = \chi(g_2)$ for all $\chi \in \widehat{G}_\chi$.

On the other hand consider the characteristic functions of the classes of G defined in 4.7. As these are class functions by the previous theorem we can express them as a linear combination of elements of $\widehat{G}_\chi$ as

$$\delta = \sum_{\chi \in \widehat{G}_\chi} \alpha_\chi \chi$$

Let then δ be the characteristic function of the conjugacy class of which g_1 is a member, then

$$\delta(g_2) = \sum_{\chi \in \widehat{G}_\chi} \alpha_\chi \chi(g_2) = \sum_{\chi \in \widehat{G}_\chi} \alpha_\chi \chi(g_1) = \delta(g_1) = 1$$

Therefore g_2 must be conjugate to g_1 . $\square$

The information gathered so far is enough to characterize the irreducible representations of direct products of groups from the irreducibles of each group, and justifies why for such direct products we defined tensor representations, together with the fact that products of characters are characters of tensor representations.

Theorem 4.3.34. *Let G_1, G_2 be groups and consider the direct product $G_1 \times G_2$, then*

$$\widehat{G_1 \times G_2} = \{\rho_1 \otimes \rho_2 \mid \rho_1 \in \widehat{G_1}, \rho_2 \in \widehat{G_2}\}$$

Proof. By part d) of proposition 4.3.12 we know that $\chi_{\rho_1 \otimes \rho_2} = \chi_{\rho_1} \chi_{\rho_2}$, hence if $\rho_1, \sigma_1 \in \widehat{G_1}$ and $\rho_2, \sigma_2 \in \widehat{G_2}$ we have that

$$\begin{aligned} \langle \chi_{\rho_1 \otimes \rho_2}, \chi_{\sigma_1 \otimes \sigma_2} \rangle_{G_1 \times G_2} &= \frac{1}{|G_1 \times G_2|} \sum_{\substack{g_1 \in G_1, \\ g_2 \in G_2}} \chi_{\rho_1 \otimes \rho_2}(g_1, g_2) \overline{\chi_{\sigma_1 \otimes \sigma_2}(g_1, g_2)} = \\ &= \frac{1}{|G_1| |G_2|} \sum_{\substack{g_1 \in G_1, \\ g_2 \in G_2}} \chi_{\rho_1}(g_1) \chi_{\rho_2}(g_2) \overline{\chi_{\sigma_1}(g_1) \chi_{\sigma_2}(g_2)} = \langle \chi_{\rho_1}, \chi_{\sigma_1} \rangle_{G_1} \langle \chi_{\rho_2}, \chi_{\sigma_2} \rangle_{G_2} \end{aligned}$$

Therefore by corollary 4.3.19

$$\langle \chi_{\rho_1 \otimes \rho_2}, \chi_{\sigma_1 \otimes \sigma_2} \rangle_{G_1 \times G_2} = \begin{cases} 1 & \text{if } \rho_1 \equiv \sigma_1 \text{ and } \rho_2 \equiv \sigma_2 \\ 0 & \text{otherwise} \end{cases}$$

Considering additionally theorem 4.3.28 the representations $\{\rho_1 \otimes \rho_2 \mid \rho_1 \in \widehat{G_1}, \rho_2 \in \widehat{G_2}\}$ are non-equivalent and irreducible, moreover the conjugacy number of $G_1 \times G_2$ is equal to the product of the conjugacy numbers of G_1 and G_2 , that is by theorem 4.3.32 $|\widehat{G_1 \times G_2}| = |\widehat{G_1}| |\widehat{G_2}|$, hence we have exhausted all irreducible representations of $G_1 \times G_2$. $\square$

Considering a group G has a number of irreducible characters equal to its conjugacy number and the characters are constant in the conjugacy classes, we can organize the characters in a square matrix, called the *character table* G .

Definition 4.3.35. Given a group G with conjugacy number k we define its character table to be the square matrix

$$\mathfrak{C} = (\chi(g_i))_{\substack{\chi \in \overline{G}_\chi, \\ i \in [k]}}$$

Although we define it as a matrix we do usually think of it as a table, for example:

Example 4.3.36. The character table of S_3 was already alluded to in example 4.3.23, we need only to summarize the columns. If we let $g_1 = 1_G$, $g_2 = (1\ 2)$ and $g_3 = (1\ 2\ 3)$ be the representatives of each class and denote them by 1 , J_2 , J_3 respectively then we get the character table

	1	J_2	J_3
χ_1	1	1	1
χ_ϵ	1	-1	1
χ_σ	2	0	-1

It is not unusual to add the sizes of each conjugacy class to the table for easy reference, in that case we'd modify the table to

	1	J_2	J_3
χ_1	1	1	1
χ_ϵ	1	-1	1
χ_σ	2	0	-1
	1	3	2

Although it is known that character tables are not enough to characterize groups, as examples have been found in the past of different group with the same character table, they still provide a lot of information and so are very much worth studying in group theory.

In the context of character tables corollary 4.3.19 is usually referred to as row orthogonality of characters, there is a similar result that established the column orthogonality of characters. To that end we first note that the inner product of class functions $C(G)$ can be simplified by using the size of the conjugacy classes of G . Let $\{J_i\}_{i \in [k]}$ be the conjugacy classes of G with respective representatives $\{g_i\}_{i \in [k]}$ and let $f, f' \in C(G)$, then

$$\langle f, f' \rangle_G = \frac{1}{|G|} \sum_{g \in G} f(g) \overline{f'(g)} = \frac{1}{|G|} \sum_{i \in [k]} |J_i| f(g_i) \overline{f'(g_i)}$$

We define the *centralizer* of an element g of G as $C_G(g) = \{h \in G \mid hg = gh\}$, as is known from group theory [Rot95, 3.2]

$$\frac{|G|}{|C_G(g)|} = |J_i|$$

where J_i is the conjugacy class of G that contains g , as such we have

$$\langle f, f' \rangle_G = \sum_{i \in [k]} \frac{1}{|C_G(g_i)|} f(g_i) \overline{f'(g_i)}$$

Theorem 4.3.37. *Let G be a group and consider $\widehat{G}_\chi$, additionally let $\{g_i\}_{i \in [k]}$ be a set of representatives of the conjugacy classes of G , then for $g_i, g_j \in \{g_i\}_{i \in [k]}$*

$$\sum_{\chi \in \widehat{G}_\chi} \chi(g_i) \overline{\chi(g_j)} = \delta_{ij} |C_G(g_i)|$$

Where δ_{ij} is the Kronecker delta, such that $\delta_{ij} = 1$ if $i = j$ and 0 otherwise.

Proof. Consider the characters table $\mathfrak{C}$ of G , as we've seen for $\chi_i, \chi_j \in \widehat{G}_\chi$

$$\delta_{ij} = \langle \chi_i, \chi_j \rangle_G = \sum_{l \in [k]} \frac{1}{|C_G(g_l)|} \chi_i(g_l) \overline{\chi_j(g_l)}$$

therefore if we consider the diagonal matrix

$$Z = \begin{bmatrix} C_G(g_1) & 0 & \cdots & 0 \\ 0 & C_G(g_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & C_G(g_k) \end{bmatrix}$$

it's straightforward to verify that $\mathfrak{C}Z^{-1}\mathfrak{C}^*$, where $\mathfrak{C}^*$ is the conjugate transpose of the matrix $\mathfrak{C}$, is the identity matrix of size $k \times k$. Since $\mathfrak{C}$ is a square matrix then $Z^{-1}\mathfrak{C}^*$ is the inverse of $\mathfrak{C}$ hence $\mathfrak{C}^*\mathfrak{C} = Z$ and the conclusion follows immediately. $\square$

Considering that $C_G(1_G) = G$ another proof of corollary 4.3.26 follows immediately from the above theorem.

4.4. The group algebra

In the previous section we developed Character Theory focusing on the field $\mathbb{C}$ and using mostly Linear Algebra, it's possible to develop this theory in greater generality using the famous *Artin-Wedderburn* theorem [Bea99, 3.3.2 (3)] and taking a more module/ring theoretic approach [Bea99, sec 4.2], but letting go of the field $\mathbb{C}$ won't necessarily grant us the

full capabilities of Schur's lemma and Maschke's theorem, which is something we don't want to let go. Yet recall that Schur's lemma requires only that the field we work with be algebraically closed, and the more general Maschke's theorem [Bea99, 2.3.8] requires only that the field's characteristic be not a divisor of the order of the group in question.

In this section we'll prove a particular case of the *Artin-Wedderburn* theorem for the group algebra which will give us its structure in terms of matrix rings, it will be very useful later on and this result is regarded as a fundamental theorem in Representation and Character Theory.

As in the previous section we'll try to do this in the simplest of ways, avoiding the necessity to go down a great escapade through ring and module theory (yet we do invite the reader on doing so on a later date with [Bea99] or [Hup98]). As such we'll remain tied to the field $\mathbb{C}$ to make use of the theory developed so far.

With all the work done previously the results here will follow quite easily.

Proposition 4.4.1. *Consider the algebras $\mathbb{C}G$ and $\text{End}_G(\mathbb{C}G)$, then*

$$\mathbb{C}G \cong \text{End}_G(\mathbb{C}G)$$

as algebras.

Proof. Indeed for $\theta \in \text{End}_G(\mathbb{C}G)$ and $f = \sum_{g \in G} \alpha_g g \in \mathbb{C}G$

$$\theta(f) = \sum_{g \in G} \alpha_g \theta(g) = \sum_{g \in G} \alpha_g g \theta(1_G) = f \theta(1_G)$$

Thus the G -homomorphism θ is completely determined by its image of 1_G , so if we consider the map $\Phi : \text{End}_G(\mathbb{C}G) \rightarrow \mathbb{C}G$ defined as, for $\theta \in \text{End}_G(\mathbb{C}G)$ if $\theta(1_G) = \sum_{g \in G} \alpha_g g$

$$\Phi(\theta) = \sum_{g \in G} \alpha_g g^{-1}$$

then it is clear this is an injective linear map, moreover it's straightforward to verify that for any $f \in \mathbb{C}G$ if we define $\theta(f') = f'f$, for $f' \in \mathbb{C}G$, then we get a G -homomorphism so Φ is a bijection. Finally if $\theta_1, \theta_2 \in \text{End}_G(\mathbb{C}G)$ with $\theta_1(1_G) = \sum_{g \in G} \alpha_g g$ and $\theta_2(1_G) = \sum_{h \in G} \beta_h h$ then $\theta_1(\theta_2(1_G)) = \theta_2(1_G)\theta_1(1_G)$ thus

$$\Phi(\theta_1\theta_2) = \sum_{k \in G} \sum_{hg=k} \beta_h \alpha_g k^{-1} = \sum_{k \in G} \sum_{g^{-1}h^{-1}=k} \alpha_g \beta_h k =$$

$$= \left(\sum_{g \in G} \alpha_g g^{-1} \right) \left(\sum_{h \in G} \beta_h h^{-1} \right) = \theta_1(1_G) \theta_2(1_G) = \Phi(\theta_1) \Phi(\theta_2)$$

Therefore Φ is an algebra isomorphism. $\square$

Note the above theorem is independent of the G -action considered in $\text{End}_G(\mathbb{C}G)$, that is the considered action of G on itself. The reason for in the above proof defining Φ the way we did and not simply $\Phi(\theta) = \theta(1_G)$ is because in this case we'd get $\Phi(\theta_1 \theta_2) = \theta_2(1_G) \theta_1(1_G)$ and so it wouldn't be a ring isomorphism and consequently an algebra isomorphism, though it would still be a vector space isomorphism.

The above proposition is all we needed to prove the structure theorem for the group algebra $\mathbb{C}G$. We recall that the set of square matrices of size $n \times n$, $n \in \mathbb{N}$, with entries in $\mathbb{C}$ is a vector space over $\mathbb{C}$ with the natural addition and scalar product, moreover it's an algebra with the natural product, we denote it by $M_n(\mathbb{C})$.

Theorem 4.4.2. *The group algebra $\mathbb{C}G$ decomposes as*

$$\mathbb{C}G \cong \bigoplus_{i \in [k]} M_{n_i}(\mathbb{C})$$

where k is the conjugacy number of G . Moreover each matrix algebra M_{n_i} is isomorphic, and uniquely corresponds to $\text{End}_G(n_i V_i)$ for some $i \in [k]$ where $n_i V_i$ is an isotypic component of the regular representation of G , hence each n_i is equal to the dimension of a distinct non equivalent representation of G .

Proof. Considering the regular representation of G it's an immediate consequence of proposition 4.4.1 and corollary 4.2.17 that $\mathbb{C}G$ decomposes into a finite direct sum of matrix algebras, each corresponding to the G -endomorphisms of a distinct isotypic component of the regular representation. Now by proposition 4.3.24 $k = |\widehat{G}|$ and $n_i = \dim_{\mathbb{C}}(\rho_i)$ for $\rho_i \in \widehat{G}$, theorem 4.3.32 finishes the proof by asserting that k is the conjugacy number of G . $\square$

The previous theorem is usually stated in a different way.

Corollary 4.4.3.

$$\mathbb{C}G \cong \bigoplus_{\rho \in \widehat{G}} \text{End}(V_\rho)$$

where $\text{End}(V_\rho)$ is the algebra of linear maps on V_ρ .

Proof. This is straightforward as considering the previous theorem for each $M_{n_i}(\mathbb{C})$, n_i is the dimension of a corresponding irreducible representations ρ_{V_i} of G and so $M_{n_i}(\mathbb{C}) \cong \text{End}(V_i)$ by $V_i \cong \mathbb{C}^{n_i}$, that is $\text{End}(V_i) \cong \text{End}_G(n_i V_i)$ (this also follows by Schur's lemma). $\square$

Corollary 4.4.4. *A group G is abelian if and only if every irreducible representation of G is one dimensional.*

Proof. The first implication was done in proposition 4.2.9. For the other direction if every irreducible representation is one dimensional then by the previous corollary we have that

$$\mathbb{C}G \cong \mathbb{C}^{|\widehat{G}|}$$

as $\text{End}(V_i) \cong \mathbb{C}$ hence $\mathbb{C}G$ is commutative so in particular G must be commutative. $\square$

An interesting and concrete isomorphisms for corollary 4.4.3 can be constructed in a straightforward way by exploiting theorem 4.3.27, namely that the matrix coefficients of non equivalent irreducible representations form a basis of $\mathbb{C}G$, and the fact that $\mathbb{C}G$ and $\mathbb{C}[G]$ are isomorphic vector spaces (and indeed isomorphic algebras).

Consider an irreducible representation $\rho : G \rightarrow \text{GL}(V)$, as G is a basis of $\mathbb{C}[G]$, ρ can be extended uniquely to a homomorphism $\hat{\rho} : \mathbb{C}G \rightarrow \text{End}(V)$ by letting

$$\hat{\rho} \left(\sum_{g \in G} \alpha_g g \right) = \sum_{g \in G} \alpha_g \rho(g) \quad (4.9)$$

The homomorphism $\hat{\rho}$ is usually called the *Fourier transform* of G at ρ and it's clearly a linear map, more generally we have the following.

Definition 4.4.5. *Let G be a group, $\mathcal{E}(G) = \bigoplus_{\rho_{V_i} \in \widehat{G}} \text{End}(V_i)$ and for $\rho \in \widehat{G}$ consider its extension to $\mathbb{C}G$, $\hat{\rho}$ as above. Then we define the Fourier transform of G to be the map*

$$\begin{aligned} \mathcal{F} : \mathbb{C}G &\rightarrow \mathcal{E}(G) \\ f &\mapsto \bigoplus_{\rho \in \widehat{G}} \hat{\rho}(f) \end{aligned}$$

Considering that for $f \in \mathbb{C}G$ and $\rho_{V_i} \in \widehat{G}$, $\hat{\rho}(f)$ is a linear combination of maps in $\text{GL}(V_i)$ it is clear that $\hat{\rho}(f) \in \text{End}(V_i)$ so the map $\mathcal{F}$ is well defined.

Theorem 4.4.6. *For a group G , its Fourier transform $\mathcal{F}$ is an isomorphism of algebras.*

Proof. Indeed considering that each $\hat{\rho}$ is a linear map then $\mathcal{F}$ is a linear map so we need only to prove that it's a bijective homomorphism of rings. First note that for $f_1 = \sum_{g \in G} \alpha_g g$, $f_2 = \sum_{h \in G} \beta_h h \in \mathbb{C}G$

$$\mathcal{F}(f_1 f_2) = \bigoplus_{\rho \in \widehat{G}} \hat{\rho}(f_1 f_2)$$

and

$$\hat{\rho}(f_1 f_2) = \hat{\rho} \left(\sum_{k \in G} \sum_{gh=k} \alpha_g \beta_h k \right) = \sum_{k \in G} \sum_{gh=k} \alpha_g \beta_h \rho(k)$$

As $\rho(k) = \rho(gh) = \rho(g)\rho(h)$ we have that

$$\hat{\rho}(f_1 f_2) = \left(\sum_{g \in G} \alpha_g \rho(g) \right) \left(\sum_{h \in G} \beta_h \rho(h) \right) = \hat{\rho}(f_1) \hat{\rho}(f_2)$$

Thus $\mathcal{F}(f_1 f_2) = \mathcal{F}(f_1) \mathcal{F}(f_2)$ and as it's clear that $\hat{\rho}(1_G) = id_{V_i}$ we indeed have a ring homomorphism.

To prove it's a bijection consider the matrix coefficients basis of $\mathbb{C}[G]$, $B = \{\varphi_{ij}^\rho \mid \rho_{V_i} \in \widehat{G}; i, j \in \dim_{\mathbb{C}}(V_i)\}$, this basis can be easily mapped to a basis of $\mathbb{C}G$ by considering the map

$$f \mapsto \sum_{g \in G} f(g)g \quad (4.10)$$

for $f \in \mathbb{C}[G]$. This is clearly an injective linear map and so by proposition C.0.11 the image of B is a basis of $\mathbb{C}G$, we'll denote this basis with the same terminology, the usage will always be clear from context.

As the conjugation map in $\mathbb{C}$, $\alpha \mapsto \bar{\alpha}$ is an isomorphism that can be straightforwardly extended to an automorphism of $\mathbb{C}G$ we have that $\{\bar{\varphi}_{ij}^\rho \mid \rho_{V_i} \in \widehat{G}; i, j \in \dim_{\mathbb{C}}(V_i)\}$ is also a basis of $\mathbb{C}G$, now note that

$$\begin{aligned} \mathcal{F} \left(\frac{1}{|G|} \bar{\varphi}_{ij}^\sigma \right) &= \bigoplus_{\rho \in \widehat{G}} \hat{\rho} \left(\frac{1}{|G|} \bar{\varphi}_{ij}^\sigma \right) = \\ &= \bigoplus_{\rho \in \widehat{G}} \frac{1}{|G|} \sum_{g \in G} \rho(\bar{\varphi}_{ij}^\sigma) \end{aligned}$$

If we consider the representations as matrix representations (in the same basis as we had to choose to get the matrix coefficients), then for the matrix $P = \frac{1}{|G|} \sum_{g \in G} \rho(\bar{\varphi}_{ij}^\sigma)$ its entries are

$$P_{kl} = \frac{1}{|G|} \sum_{g \in G} \bar{\varphi}_{ij}^\sigma(g) \varphi_{kl}^\rho(g) = \langle \varphi_{kl}^\rho, \varphi_{ij}^\sigma \rangle_G = \delta_{\sigma\rho} \delta_{ij,kl} \frac{1}{d}$$

where d is the dimension of σ , hence

$$\mathcal{F}\left(\frac{1}{|G|}\bar{\varphi}_{ij}^\sigma\right) \equiv \bigoplus_{\rho \in \widehat{G}} P_\rho$$

Where $P_\rho = 0$ for $\rho \neq \sigma$ and for $\rho = \sigma$, P_σ is a matrix of all zeros except in the position (i, j) where it has $1/d$. From this it is clear that for each $\rho_{V_i} \in \widehat{G}$, $\{\frac{1}{|G|}\bar{\varphi}_{ij}^\rho \mid i, j \in \dim_{\mathbb{C}}(V_i)\}$ gets mapped to a basis of $M_{n_i}(\mathbb{F})$ and consequently as $M_{n_i} \cong \text{End}(V_i)$, $\{\frac{1}{|G|}\bar{\varphi}_{ij}^\rho \mid \rho_{V_i} \in \widehat{G}; i, j \in \dim_{\mathbb{C}}(V_i)\}$ gets mapped to a basis of $\mathcal{E}(G)$. Hence $\mathcal{F}$ is a bijection by being a linear map. $\square$

The Fourier transform is actually normally defined not on the algebra $\mathbb{C}G$ but on the algebra $\mathbb{C}[G]$ of complex functions, which is indeed an algebra with the product

$$[f_1 * f_2](k) = \sum_{gh=k} f_1(g)f_2(h)$$

for $f_1, f_2 \in \mathbb{C}[G]$ and $k \in G$. This product is called the *involution* on $\mathbb{C}[G]$, note that it's basically the product on $\mathbb{C}G$ and the map 4.4 preserves this product, turning it into an isomorphism of algebras. From this, under our setting, if we let θ be the map 4.4 then the Fourier transform would be defined as $\mathcal{F}' = \mathcal{F}\theta$, so it's clear that $\mathcal{F}'(f_1 * f_2) = \mathcal{F}'(f_1)\mathcal{F}'(f_2)$. Likewise the Fourier transform of G at ρ becomes $\hat{\rho}$, here we adopt the same terminology and the situation is made clear from context.

Proposition 4.4.7. *Let $f \in C(G)$ be a class function and $\rho_{V_i} \in \widehat{G}$ then for the Fourier transform of G at ρ we have*

$$\hat{\rho}(f) = \frac{|G|}{d_i} \langle f, \bar{\chi}_\rho \rangle_G \text{id}_{V_i}$$

where χ_ρ is the character of ρ and as before d_i is the dimension of V_i .

Proof. Note that $\hat{\rho}(f)$ is a G -endomorphism on V_i , indeed for $g \in G$

$$\begin{aligned} \hat{\rho}(f)\rho(g) &= \sum_{h \in G} f(h)\rho(hg) = \\ &= \sum_{h \in G} f(ghg^{-1})\rho(gh) = \sum_{h \in G} f(h)\rho(gh) = \rho(g)\hat{\rho}(f) \end{aligned}$$

where the second to last equality comes from f being a class function. Hence as ρ is irreducible by Schur's lemma $\hat{\rho}(f) = c \text{id}_{V_i}$ for some $c \in \mathbb{C}$, we can find this value by computing traces as

$$\text{tr}(\hat{\rho}(f)) = \sum_{g \in G} f(g)\text{tr}(\rho(g)) = \sum_{g \in G} f(g)\chi_\rho(g) = |G|\langle f, \bar{\chi}_\rho \rangle_G$$

and $\text{tr}(cid_{V_i}) = c d_i$, thus $c = \frac{|G|}{d_i} \langle f, \overline{\chi}_\rho \rangle_G$. □

Corollary 4.4.8. *Let $\rho_V, \sigma_W \in \widehat{G}$, then*

$$\rho(\overline{\chi}_\sigma) = \begin{cases} \frac{|G|}{d} id_V & \text{if } \rho \equiv \sigma \\ 0 & \text{otherwise} \end{cases}$$

where $d = \dim_{\mathbb{C}}(V)$.

4.5. Normal subgroups and lifting

As is known given a group G if N is a normal subgroup then G/N is a group, one now wonders if there's any relation between the irreducible representation of G/N and the irreducible representations of G . Indeed it's possible to obtain certain irreducible representations of G from the irreducible ones of G/N , and vice versa. As G/N is a simpler group finding its irreducible representations might be quite easier, and under this setting we can choose N to our advantage.

In what follows we denote the subgroup relation by $\leq$ thus when H is a subgroup of a group G we write $H \leq G$.

Theorem 4.5.1. *Let N be a normal subgroup of a group G and $\hat{\rho}_V$ a representation of G/N . Consider the natural projection map $i : G \rightarrow G/N$ such that $i(g) = gN$, then the composition $\rho = \hat{\rho}i$ is a representation of G , moreover both representations have the same dimension and*

- a) *If $\chi_{\hat{\rho}}$ is the character afforded by $\hat{\rho}$ then the character afforded by ρ is $\chi_\rho = \chi_{\hat{\rho}}i$.*
- b) *ρ is irreducible if and only if $\hat{\rho}$ is irreducible.*
- c) *For $\hat{\rho}, \hat{\sigma}$ representations of G/N consider $\rho = \hat{\rho}i$ and $\sigma = \hat{\sigma}i$, then σ, ρ are equivalent if and only if $\hat{\rho}, \hat{\sigma}$ are equivalent.*

Proof. It's easy to see that ρ is a representation of G as the map i is a group homomorphism and so as ρ is a composition of homomorphism it must be itself a homomorphism.

a) Indeed $\chi_\rho(g) = \text{tr}(\rho(g)) = \text{tr}(\hat{\rho}(i(g))) = \chi_{\hat{\rho}}(i(g))$ hence $\chi_\rho = \chi_{\hat{\rho}}i$.

b) Just note that

$$\langle \chi_\rho, \chi_\rho \rangle_G = \frac{1}{|G|} \sum_{g \in G} \chi_\rho(g) \overline{\chi_\rho(g)} = \frac{1}{|G|} \sum_{g \in G} \chi_{\hat{\rho}}(i(g)) \overline{\chi_{\hat{\rho}}(i(g))}$$

$$= \frac{1}{|G|} |N| \sum_{gN \in G/N} \chi_{\hat{\rho}}(gN) \overline{\chi_{\hat{\rho}}(gN)} = \langle \chi_{\hat{\rho}}, \chi_{\hat{\rho}} \rangle_{G/N} = 1$$

Hence the conclusion follows by theorem 4.3.28.

c) Immediate by part a) and theorem 4.3.22. □

Definition 4.5.2. Under the conditions of the previous theorem, we say $\hat{\rho}$ lifts to ρ , and the same goes for the corresponding characters.

Corollary 4.5.3. The lifting operation produces a bijection between irreducibles of G/N and irreducibles of G with N contained in their kernel, that is letting $\mathfrak{l} : \widehat{G/N} \rightarrow \widehat{G}$ be the lifting operator then for $\hat{\rho} \in \widehat{G/N}$, $\mathfrak{l}(\hat{\rho}) = \rho \in \widehat{G}$ such that $N \subseteq \ker(\rho)$, $\mathfrak{l}$ is injective and for $\rho \in \widehat{G}$ such that $N \subseteq \ker(\rho)$, $\mathfrak{l}^{-1}(\rho) \in \widehat{G/N}$.

Proof. Indeed it is clear that lifting $\hat{\rho}_V \in \widehat{G/N}$ to $\rho_V \in \widehat{G}$ gives for $k \in N$, $\rho(k) = \hat{\rho}(i(k)) = \hat{\rho}(kN) = \hat{\rho}(N) = id_V$ hence $N \subseteq \ker(\rho)$. Moreover the map $\mathfrak{l}$ is clearly injective and if $\rho : G \rightarrow \text{GL}(V) \in \widehat{G}$ such that $N \subseteq \ker(\rho)$ then we can define $\rho' : G/N \rightarrow \text{GL}(V)$ as $\rho'(gN) = \rho(g)$, we note that if $gH = kH$ then $k^{-1}g \in H$ and so $\rho(k^{-1}g) = id_V \Leftrightarrow \rho'(gH) = \rho(g) = \rho(k) = \rho'(kH)$ hence the function is well defined and it's easy to verify this is a group homomorphism and $\rho = \rho' \circ i$ where $i : G \rightarrow G/N$ is the natural projection, hence by part b) of the previous theorem $\rho' \in \widehat{G/N}$. □

By the above corollary we can produce all non faithful irreducible representation of a group G from irreducibles of the quotient groups G/N_i for $N_i \neq \{1_G\}$ normal subgroups of G . If we want faithful representations then we'd have to lift irreducible representations of $G/\{1_G\} = G$ so there's nothing to be gained in this situation from lifting.

Some normal subgroups will affect the structure of G/N more than other, the simplest case for us in representation theory is when G/N is abelian, as we've seen (example 4.2.11) in this case we can easily describe all irreducible representations of G/N and as they're one dimensional we can lift them to one dimension representations of G .

Lemma 4.5.4. Let G be a group, then the commutator/derived subgroup of G is the group G' generated by

$$\{aba^{-1}b^{-1} \mid ab \in G\}$$

the elements of this set are denoted by $[a, b] = aba^{-1}b^{-1}$ and $[a, b]$ is called the commutator of a and b .

Thus

$$G' = \langle [a, b] : a, b \in G \rangle$$

The commutator subgroup G' is a normal subgroup of G , moreover if N is a normal subgroup of G then G/N is abelian if and only if $G' \leq N$, hence the commutator is the minimal subgroup of G such that G/N is abelian.

The proof of this lemma can be seen in [Rot95, 2.23] and it's straightforward to demonstrate, for us it means the following.

Theorem 4.5.5. *Let G be a group and G' its commutator subgroup, then G has precisely $n = |G|/|G'|$ (irreducible) representations of dimension one, all obtained by lifting the irreducibles of G/G' .*

Proof. By corollary 4.5.3 lifting gives a bijection between irreducibles of G/G' and irreducibles of G whose kernel contains G' so all we have to prove is that irreducibles of G whose kernel contains G' are necessarily one dimensional considering that for a one dimensional representation $\rho(gh) = \rho(hg) \Leftrightarrow \rho(ghg^{-1}h^{-1}) = id_V$ and so $G' \subseteq \ker(\rho)$.

Just as in the proof of the previous corollary for a representation $\rho : G \rightarrow \text{GL}(V) \in \widehat{G}$ such that $G' \subseteq \ker(\rho)$ we can define the representation $\rho' : G/G' \rightarrow \text{GL}(V)$ such that $\rho = \rho' \circ i$ where $i : G \rightarrow G/G'$ is the natural projection. By part b) of theorem 4.5.1 $\rho' \in \widehat{G/G'}$ hence it is an irreducible representation of an abelian group, so by proposition 4.2.9 this representation is one dimensional and finally again by theorem 4.5.1 ρ must be one dimensional. As G/G' is abelian by proposition 4.2.10 it has $|G/G'|$ irreducible representations, concluding the proof. $\square$

We recall that a group G is said to be simple if it contains no non trivial normal subgroups, in other words any normal subgroup of G is either $\{1_G\}$ of G itself.

Proposition 4.5.6. *Any normal subgroup N of a group G can be expressed as the intersection of kernels of irreducible representations of G , that is*

$$N = \bigcap_{\rho \in I} \ker(\rho)$$

where $I \subseteq \widehat{G}$. In particular G is simple if and only if all irreducible representations of G have trivial kernel.

Proof. We first prove that any normal subgroup of G is the kernel of some representation of G . Indeed consider the regular representation of G/N , λ , this representation is faithful considering that if $\lambda(gN) = \lambda(hN)$ then for all $kN \in G/N$, $gkN = hkN$ and so in particular $gN = hN$. Now if we let $i : G \rightarrow G/N$ be the natural projection then $\rho = \lambda \circ i$ is a representation of G with kernel N .

Now by corollary 4.2.4 the representation ρ is a direct sum of irreducible representations of G , hence if $\rho = \bigoplus_{i \in [n]} \rho_i$ with each $\rho_i \in \widehat{G}$ then $N = \ker(\rho) = \bigcap_{i \in [n]} \ker(\rho_i)$.

Finally, to prove the converse, if $N \neq \{1_G\}$ is a normal subgroup of G then $N = \bigcap_{i \in [n]} \ker(\rho_i)$ for $\rho_i \in \widehat{G}$ and so for some $\rho_i \in \widehat{G}$, $\ker(\rho_i) \neq \{1_G\}$. On the other hand if all irreducible representations of G have trivial kernel then as N is the intersection of kernels of irreducibles of G it must be the case that $N = \{1_G\}$ or $N = G$, hence G is simple. $\square$

With *lifting* in our armory we now tackle the character table of S_4 .

Example 4.5.7. Let $G = S_4$, as is know the conjugacy classes correspond to cycle types, hence we have 5 conjugacy classes and so 5 irreducible characters. As with S_3 we know already 3 irreducible representations; the trivial, the alternating and the 3-dimensional representation σ_3 contained in the permutation representation λ of the natural action of S_4 on $[4]$ (as in example 4.1.32), the characters are clear for the trivial and alternating representation, and for σ noting that $\lambda = \iota_G \oplus \sigma$ then $\chi_\sigma = \chi_\lambda - \chi_\iota$ where ι_G is the trivial representation. Hence using proposition 4.3.17 for the character of λ so far we have the table*

	1	(1 2)	(1 2 3)	(1 2)(3 4)	(1 2 3 4)
χ_ι	1	1	1	1	1
χ_ϵ	1	-1	1	1	-1
χ_σ	3	1	0	-1	-1
χ_4					
χ_5					

By theorem 4.3.34 the product of two irreducible characters is again an irreducible character, hence the product $\chi_\sigma \chi_\epsilon$ given an irreducible character of S_4 and indeed it is different from the ones above, so we get

	1	(1 2)	(1 2 3)	(1 2)(3 4)	(1 2 3 4)
χ_ι	1	1	1	1	1
χ_ϵ	1	-1	1	1	-1
χ_σ	3	1	0	-1	-1
$\chi_\sigma \chi_\epsilon$	3	-1	0	-1	1
χ_5					

*Recall each column pertains to a conjugacy class and not a single element of the group, here we choose a set of representatives of each class to name the columns.

We can find the character χ_5 in several ways, here we'll do it by lifting an irreducible character of S_3 . Let $N = \mathbb{Z}_2 \times \mathbb{Z}_2$, this group is called the *Klein group* of order 4, it has element wise addition as its operation. We can embed this group into S_4 using Cayley's theorem, doing so produces

$$(0, 0) \rightarrow 1_{S_4}$$

$$(0, 1) \rightarrow (1\ 2)(3\ 4)$$

$$(1, 0) \rightarrow (1\ 3)(2\ 4)$$

$$(1, 1) \rightarrow (1\ 4)(2\ 3)$$

Thus we identify N with the respective subgroup of S_4 . We note that $S_4/N \cong S_3$, hence S_4/N is a group and so N must be a normal subgroup. Consider the two dimensional irreducible character of S_3 , $\hat{\sigma}$ as in example 4.3.23, we can lift this character to S_4 per theorem 4.5.1 and it will produce an irreducible character of degree 2 thus indeed a new group character of S_4 . So if $i : G \rightarrow G/N$ is the natural projection map then $\chi_5 = \chi_{\hat{\sigma}}i$ and we get the complete character table of S_4

	1	(1 2)	(1 2 3)	(1 2)(3 4)	(1 2 3 4)
χ_1	1	1	1	1	1
χ_{ϵ}	1	-1	1	1	-1
χ_{σ}	3	1	0	-1	-1
$\chi_{\sigma}\chi_{\epsilon}$	3	-1	0	-1	1
χ_5	2	0	-1	2	-1

4.6. Induced representations

In the previous section we've seen how one can obtain representations of a group G from representations of quotient groups of G , here we'll see how more generally we can produce representations of G from representations of subgroups of G . We'll still make use of the quotient G/H but here as H can be any subgroup we no longer have a quotient group, luckily all we need is the fact that G decomposes as a disjoint union of the cosets in G/H and so each $g \in G$ can belong to one and only one of the cosets.

Definition 4.6.1. Let G be a group and H a subgroup with $\{r_i\}_{i \in [n]}$ representatives of the left coset in G/H , so $n = |G/H|$. Moreover let V be an H -space, then we define the induced G -space of V as

$$V \upharpoonright_H^G = \bigoplus_{i \in [n]} {}_H V \otimes r_i$$

Here we treat r_i as the one dimensional vector space $\mathbb{C}r_i$, thus $V \upharpoonright_H^G$ has dimension $n \dim_{\mathbb{C}}(V)$. Letting $G = \cup_{i \in [n]} r_i H$, if $g \in G$ then $g = rh$ for some unique $h \in H$ and $r \in \{r_i\}_{i \in [n]}$, if we consider $gr_i = r_{gi}h_i(g)$ then we define the action of G on $V \upharpoonright_H^G$ by

$$g(v \otimes r_i) = h_i(g)v \otimes r_{gi}$$

that is we consider h_i , $i \in [n]$, as functions from G to H such that if $gr_i = r_{gi}h$ with $h \in H$ and $r_i \in \{r_i\}_{i \in [n]}$ then $h_i(g) = h$.

Note that we implicitly define an action of G on $[n]$ as $gi = j$ such that $gr_i = r_j h$, moreover as

$$r_{gg'i}h_i(gg') = gg'r_i = gr_{g'i}h_i(g') = r_{g(g'i)}h_{g'i}(g)h_i(g')$$

we have that

$$\begin{aligned} g(g'(v \otimes r_i)) &= g(h_i(g')v \otimes r_{g'i}) = h_{g'i}(g)h_i(g')v \otimes r_{gg'i} = \\ &= h_i(gg')v \otimes r_{gg'i} = (gg')(v \otimes r_i) \end{aligned}$$

So indeed we have a group action of G on the space $V \upharpoonright_H^G$, and this action is linear as

$$g(\alpha(v \otimes r_i)) = g(\alpha v \otimes r_i) = h_i(g)\alpha v \otimes r_{gi} = \alpha h_i(g)v \otimes r_{gi} = \alpha g(v \otimes r_i)$$

Thus $V \upharpoonright_H^G$ is indeed a G -space. Given $g \in G$ then by the above definition g acts on $V \upharpoonright_H^G$ by permuting the spaces $V \otimes r_i$ in the same way it permutes the cosets $r_i H$, by the action of G on $[n]$.

Definition 4.6.2. Under the condition of the previous definition, if ρ is the representation associated with ${}_H V$ then we call the representation associated with the induced G -space $V \upharpoonright_H^G$ the (aptly named) induced representation, we denote it by $\rho \upharpoonright_H^G$ but when the subgroup H is clear from context we denote it simply by ρ^G . The same goes for characters, so that χ_{ρ^G} is then the induced character which we can denote simply by χ_{ρ}^G .

Proposition 4.6.3. *Let ρ_V be a representation of $H \leq G$ and $\{r_i\}_{i \in [n]}$ representatives of the left cosets G/H , then for the induced representation ρ^G*

$$\ker(\rho^G) = \bigcap_{i \in [n]} r_i^{-1} \ker(\rho) r_i$$

Proof. If $g \in G$ acts trivially on $V|_H^G$, that is $g \in \ker(\rho^G)$, then it must be the case that $gr_i = r_i h_i(g)$ and $h_i(g) \in \ker(\rho)$ for all $i \in [n]$, hence for all $i \in [n]$, $g \in r_i^{-1} H r_i \cap r_i^{-1} \ker(\rho) r_i = r_i^{-1} \ker(\rho) r_i$ as $\ker(\rho) \subseteq H$, thus

$$g \in \bigcap_{i \in [n]} r_i^{-1} \ker(\rho) r_i$$

and of course if g is in the above set then clearly $g \in \ker(\rho^G)$, so in conclusion

$$\ker(\rho^G) = \bigcap_{i \in [n]} r_i^{-1} \ker(\rho) r_i$$

□

We note that as $\ker(\rho)$ is a normal subgroup of H we could equally write

$$\ker(\rho^G) = \bigcap_{g \in G} g^{-1} \ker(\rho) g$$

Considering we have a new representation we should now ask for its character.

Theorem 4.6.4. *Let ρ_V be a representation of $H \leq G$ and $\{r_i\}_{i \in [n]}$ representatives of the left cosets G/H , Let also χ_ρ be the character afforded by ρ and define the function*

$$\check{\chi}(g) = \begin{cases} \chi_\rho(g) & \text{if } g \in H \\ 0 & \text{otherwise} \end{cases}$$

for $g \in G$. Then the character afforded by ρ^G is

$$\chi_\rho^G(g) = \sum_{i \in [n]} \check{\chi}(r_i^{-1} g r_i) = \frac{1}{|H|} \sum_{k \in G} \check{\chi}(k^{-1} g k)$$

In particular the induced representation ρ^G does not depend on the choice of representatives of the left cosets G/H .

Proof. By definition

$$V|_H^G = \bigoplus_{i \in [n]} V \otimes r_i$$

and as we've seen for $g \in G$ with $gr_i = r_{gi}u_i(g)$, g acts on this space as $g(v \otimes r_i) = u_i(g)v \otimes r_{gi}$ thus we can consider the matrix representation as being a block diagonal matrix with $\rho(u_i(g))$, $i \in [n]$, in its diagonal composed with a (block) permutation matrix given by the action of G on $[n]$, hence the contribution to the trace of $\rho^G(g)$ comes from the blocks $\rho(u_i(g))$ that remain in the diagonal, that is $r_{gi} = r_i \Leftrightarrow r_i^{-1}gr_i \in H$, thus

$$\chi_\rho^G(g) = \sum_{\substack{i \in [n], \\ r_i^{-1}gr_i \in H}} \text{tr}(\rho(r_i^{-1}gr_i)) = \sum_{i \in [n]} \dot{\chi}(r_i^{-1}gr_i)$$

Now note that $r_i^{-1}gr_i \in H \Leftrightarrow h^{-1}r_i^{-1}gr_i h \in H$ for all $h \in H$ and so we have

$$\sum_{i \in [n]} \dot{\chi}(r_i^{-1}gr_i) = \frac{1}{|H|} \sum_{k \in G} \dot{\chi}(k^{-1}gk)$$

Considering the above any choice of representatives will produce the same induced character hence the conclusion follows immediately by theorem 4.3.22. $\square$

As it sits the definition of induced representation seems a bit esoteric, we would like to have it in a more tractable way. That can be achieved by constructing the *induced matrix representation*, as alluded in the proof of the theorem above, with some help from the character of the induced representation.

Theorem 4.6.5. *Let ρ be a matrix representation of $H \leq G$ and $\{r_i\}_{i \in [n]}$ representatives of the left cosets G/H , then the matrix of the induced representation ρ^G is given by the following block matrix*

$$\rho^G(g) \equiv [\rho_0(r_i^{-1}gr_j)]_{i,j \in [n]}$$

where $\rho_0(r_i^{-1}gr_j) = \rho(r_i^{-1}gr_j)$ if $r_i^{-1}gr_j \in H$ otherwise it's the zero matrix.

Proof. Let $g \in G$, as mentioned in the proof of the previous theorem the matrix representation of ρ^G can be seen as a block diagonal matrix with $\rho(u_i(g))$, $i \in [n]$, in its diagonal composed with a block permutation matrix given by the action of G on $[n]$.

If $gr_j = r_{gj}u_j(g)$ then the block permutation matrix sends the block $\rho(u_j(g))$ from position (j, j) to position (gj, j) . Now all we have to do is note that considering the block matrix

$$[\rho_0(r_i^{-1}gr_j)]_{i,j \in [n]}$$

then because there can only be one r_i such that $r_i^{-1}gr_i \in H$ then in each column $j \in [n]$ there is only one block matrix $\rho_0(r_i^{-1}gr_j)$ different than 0, moreover that block is in position

(gj, j) considering that $r_{gj}^{-1}gr_j \in H$, therefore

$$\rho^G(g) \equiv [\rho_0(r_i^{-1}gr_j)]_{i,j \in [n]}$$

□

Note that the matrix representation above has the expected character. Considering the induced character the next proposition comes easily.

Proposition 4.6.6. *Let ρ, σ be representations of $H \leq G$ then*

$$(\rho \oplus \sigma)^G = \rho^G \oplus \sigma^G$$

Proof. If we let χ_ρ, χ_σ be the respective characters of the aforementioned representation then by theorem 4.6.4

$$\chi_{\rho \oplus \sigma}^G = \frac{1}{|H|} \sum_{k \in G} \dot{\chi}(k^{-1}gk)$$

where

$$\dot{\chi}(g) = \begin{cases} \chi_{\rho \oplus \sigma}(g) & \text{if } g \in H \\ 0 & \text{otherwise} \end{cases} = \begin{cases} \chi_\rho(g) + \chi_\sigma(g) & \text{if } g \in H \\ 0 & \text{otherwise} \end{cases}$$

Thus we can decompose $\dot{\chi}(g)$ as

$$\dot{\chi}_\rho(g) + \dot{\chi}_\sigma(g) = \begin{cases} \chi_\rho(g) & \text{if } g \in H \\ 0 & \text{otherwise} \end{cases} + \begin{cases} \chi_\sigma(g) & \text{if } g \in H \\ 0 & \text{otherwise} \end{cases}$$

Now note that

$$\frac{1}{|H|} \sum_{k \in G} \dot{\chi}(k^{-1}gk) = \frac{1}{|H|} \sum_{k \in G} \dot{\chi}_\rho(k^{-1}gk) + \frac{1}{|H|} \sum_{k \in G} \dot{\chi}_\sigma(k^{-1}gk) = \chi_\rho^G + \chi_\sigma^G$$

Thus the conclusion follows immediately by theorem 4.3.22. □

The next theorem is known as *Frobenius' reciprocity law* and as we'll see it's very useful when dealing with induced representations. Recall the definition of restricted representation given in 4.1.14.

Definition 4.6.7. *When talking about a G -space V we denote its restriction to $H \leq G$ by $V \downarrow_H$, and when dealing with characters we denote a character χ of G restricted to H by χ_H , clearly for $h \in H$, $\chi_H(h) = \chi(h)$.*

Theorem 4.6.8 (Frobenius). *Let ξ be a character of $H \leq G$, that is ξ is a character afforded by some representation of H , and let χ be a character of G , then*

$$\langle \xi^G, \chi \rangle_G = \langle \xi, \chi_H \rangle_H$$

Proof. Using the definitions

$$\begin{aligned} \langle \xi^G, \chi \rangle_G &= \frac{1}{|G|} \sum_{g \in G} \xi^G(g) \overline{\chi(g)} = \frac{1}{|H||G|} \sum_{g \in G} \sum_{k \in G} \xi(k^{-1}gk) \overline{\chi(g)} = \\ &= \frac{1}{|H||G|} \sum_{g \in G} \sum_{k \in G} \xi(g) \overline{\chi(kgk^{-1})} \end{aligned}$$

As χ is a class function on G then $\chi(k^{-1}gk) = \chi(g)$ for all $k \in G$ hence

$$\begin{aligned} \langle \xi^G, \chi \rangle_G &= \frac{1}{|H||G|} \sum_{g \in G} \sum_{k \in G} \xi(g) \overline{\chi(g)} = \\ &= \frac{1}{|H|} \sum_{g \in G} \xi(g) \overline{\chi(g)} \end{aligned}$$

Now as $\xi(g) = 0$ if $g \notin H$ and $\xi(g)$ otherwise then

$$\langle \xi^G, \chi \rangle_G = \frac{1}{|H|} \sum_{g \in H} \xi(g) \overline{\chi(g)} = \langle \xi, \chi_H \rangle_H$$

□

By corollary 4.3.21 the following result follows immediately from the previous theorem. It is the counterpart of the Frobenius' reciprocity law and it's usually called *Nakayama's reciprocity theorem*.

Corollary 4.6.9. *Let W be an H -space and V a G -space with $H \leq G$ then*

$$\dim_{\mathbb{C}}(\text{Hom}_G(W \upharpoonright_H^G, V)) = \dim_{\mathbb{C}}(\text{Hom}_H(W, V \downarrow_H))$$

Using Frobenius' reciprocity it's easy to prove that induction is a transitive operation.

Proposition 4.6.10. *Let ρ be a representation of T with $T \leq H \leq G$, then*

$$(\rho^H)^G = \rho^G$$

Proof. Indeed letting $\xi \in \widehat{G}$ be an irreducible character then by Frobenius' reciprocity

$$\langle (\chi_\rho^H)^G, \xi \rangle_G = \langle \chi_\rho^H, \xi_H \rangle_H = \langle \chi_\rho, (\xi_H)_T \rangle_T =$$

$$= \langle \chi_\rho, \xi_T \rangle_T = \langle \chi_\rho^G, \xi \rangle_G$$

Therefore the decomposition into irreducible representation of $(\rho^H)^G$ must be the same as for ρ^G hence they must be the same representation. $\square$

Example 4.6.11. An interesting, and quite useful example of an induced representation comes from the trivial representation ι_H with $H \leq G$. Note that for $g \in G$

$$\begin{aligned} \chi_i^G(g) &= \frac{1}{|H|} \sum_{k \in G} \check{\chi}_i(k^{-1}gk) = \frac{1}{|H|} |\{k \in G \mid k^{-1}gk \in H\}| = \\ &= \frac{1}{|H|} |\{k \in G \mid gkH = kH\}| = |\{kH \in G/H \mid gkH = kH\}| \end{aligned}$$

But this is precisely the character of the permutation representation determined by the action of G on the left cosets G/H by left multiplication, hence ι_H^G is precisely that permutation representation.

Proposition 4.6.12. Let ξ be a character of $H \leq G$ and χ a character of G , then

$$\chi \xi^G = (\chi_H \xi)^G$$

In particular

$$\chi_H^G = \chi \chi_\lambda$$

Where χ_λ is the character of the permutation representation determined by the action of G on G/H by left multiplication.

Proof. Indeed for $g \in G$

$$(\chi_H \xi)^G(g) = \frac{1}{|H|} \sum_{k \in G} \chi(k^{-1}gk) \check{\xi}(k^{-1}gk)$$

As χ is a class function on G

$$(\chi_H \xi)^G(g) = \frac{1}{|H|} \sum_{k \in G} \chi(g) \check{\xi}(k^{-1}gk) = \chi(g) \frac{1}{|H|} \sum_{k \in G} \check{\xi}(k^{-1}gk) = \chi(g) \xi^G(g)$$

The particular case follows from the fact that $\chi_H = \chi_H \chi_\lambda$ and the previous example. $\square$

4.7. Permutation representations

In this section we'll focus on permutation representations, as defined in 4.1.17. By proposition 4.3.17 we know their characters explicitly and so we can exploit that to our advantage in order to understand their structure. We've done significant work in understanding the structure of the regular representation so it's only natural now to consider its general counterpart.

We start with some preliminary result with the objective of producing a generalization of proposition 4.4.1. Recall that in definition 4.1.17 given a set X we defined the group ring $\mathbb{F}X$, here we take $\mathbb{F} = \mathbb{C}$ and so we get the group ring $\mathbb{C}X$, if we consider the set $X \times X$ then we can analogously construct the group ring $\mathbb{C}(X \times X)$, recall these are in particular vector spaces. Also note we'll always assume the set X is finite, as per the definition (otherwise we'd be working with an infinite dimensional vector space). When a group action of a group G is defined on a set X we usually say X is a G -set.

Theorem 4.7.1. *Let X be a finite set and consider the vector spaces $\mathbb{C}X$, $\mathbb{C}(X \times X)$ and $\text{End}(\mathbb{C}X)$, then $\mathbb{C}(X \times X) \cong \text{End}(\mathbb{C}X)$*

Proof. As $\text{End}(\mathbb{C}X)$ is composed of linear maps on the space $\mathbb{C}X$ it is isomorphic to the space of square matrices $M_{|X|}(\mathbb{C})$, as such consider the map

$$\begin{aligned}\Phi : \mathbb{C}(X \times X) &\rightarrow M_{|X|}(\mathbb{C}) \\ f &\mapsto T_f\end{aligned}$$

Such that $T_f = (\alpha_{xy})_{x,y \in X}$ is a square matrix with the coefficients in f of the basis elements $(x, y) \in X \times X$ as its entries. It is clear that this function is well defined in the sense that $\Phi(f) \in M_{|X|}(\mathbb{C})$, moreover it's straightforward to verify that it's a bijective linear map hence the conclusion follows. $\square$

If we define a product on $\mathbb{C}(X \times X)$ as, for $f_1 = \sum_{x,y \in X} \alpha_{x,y}(x, y)$, $f_2 = \sum_{x,y \in X} \beta_{x,y}(x, y) \in \mathbb{C}(X \times X)$

$$f_1 f_2 = \sum_{x,y \in X} \sum_{z \in X} \alpha_{x,z} \beta_{z,y}(x, y) \quad (4.11)$$

then $\mathbb{C}(X \times X)$ becomes an algebra. Note the above product is essentially the same as matrix multiplication. Considering additionally the natural product on $\text{End}(\mathbb{C}X)$ as composition of maps we have the following.

Corollary 4.7.2. *Considering the previous theorem and $\mathbb{C}(X \times X)$ and $\text{End}(\mathbb{C}X)$ as algebras, with the above products, then $\mathbb{C}(X \times X) \cong \text{End}(\mathbb{C}X)$ as algebras.*

Proof. Indeed if $f_1, f_2 \in \mathbb{C}(X \times X)$ then it follows immediately from the definition of the product on $\mathbb{C}(X \times X)$ that $T_{f_1}T_{f_2} = T_{f_1f_2}$. $\square$

The above corollary can be restricted to the sub-algebra of G -homomorphisms of $\text{End}(\mathbb{C}X)$, that is $\text{End}_G(\mathbb{C}X)$. To see it we first define the G -fixed subspace of a G -space.

Definition 4.7.3. *Let V be a G -space, the G -fixed subspace of V is defined as*

$$V^G = \{v \in V \mid gv = v \text{ for all } g \in G\}$$

If we consider an action of G on the set X then we can easily extend that action to $X \times X$ by letting $g(x, y) = (gx, gy)$, with it we can define the respective permutation representations of G on X and on $X \times X$. Under this setting the G -fixed subspace of $\mathbb{C}(X \times X)$ is composed of elements $f \in \mathbb{C}(X \times X)$ such that for all $g \in G$, $gf = f$. To understand what f must be we define the following.

Definition 4.7.4. *Let X be a G -set, we define the G -orbits of G on X to be the sets*

$$Gx = \{gx \mid g \in G\}$$

for $x \in X$. It is clear that these sets form a partition of X , indeed if $hy \in Gx$ for some $h \in G$ and $x, y \in X$ then $hy = gx$ for some $g \in G$ and so $y = h^{-1}hy = h^{-1}gx \in Gx$ thus $Gy \subseteq Gx$ and in the same way one proves that $Gx \subseteq Gy$ therefore $Gx = Gy$. Considering that $x = 1_Gx \in Gx$ it follows that X is a disjoint union of its G -orbits.

Going back to our $f \in \mathbb{C}(X \times X)$ it is then clear that if $gf = f$ for all $g \in G$ then it must be the case that the coefficients of f are equal in the G -orbits of $X \times X$, that is if we let $\{z_i\}_{i \in [k]}$ be representatives of each G -orbit then

$$f = \sum_{i \in [k]} \sum_{(x,y) \in Gz_i} \alpha_i(x, y)$$

Hence $\mathbb{C}(X \times X)^G$ is composed of elements of this form. We note that $\mathbb{C}(X \times X)^G$ is a subalgebra of $\mathbb{C}(X \times X)$, indeed it contains the identity and zero of the algebra $\mathbb{C}(X \times X)$, it's clearly closed for sums and if we consider 4.11 with $f_1, f_2 \in \mathbb{C}(X \times X)^G$ then for $g \in G$

$$g(f_1f_2) = \sum_{x,y \in X} \sum_{z \in X} \alpha_{x,z} \beta_{z,y}(gx, gy) = \sum_{x,y \in X} \sum_{z \in X} \alpha_{g^{-1}x,z} \beta_{z,g^{-1}y}(x, y) =$$

$$= \sum_{x,y \in X} \sum_{z \in X} \alpha_{g^{-1}x, g^{-1}z} \beta_{g^{-1}z, g^{-1}y}(x, y) = \sum_{x,y \in X} \sum_{z \in X} \alpha_{x,z} \beta_{z,y}(x, y)$$

considering $\alpha_{x,z}$ and $\beta_{z,y}$ are constant on the G -orbits.

Proposition 4.7.5. *Given an action of a group G on a set X and the aforementioned permutation representation of G on $\mathbb{C}(X \times X)$ it is the case that*

$$\text{End}_G(\mathbb{C}X) \cong \mathbb{C}(X \times X)^G$$

as algebras.

Proof. Considering corollary 4.7.2 all we have to show is a bijection between the elements of the algebras, specifically that for an $f \in \mathbb{C}(X \times X)^G$, $T_f = (\alpha_{xy})_{x,y \in X}$ as in theorem 4.7.1 corresponds to a G -homomorphism and vice versa. As such let θ_f be the linear map on $\mathbb{C}X$ associated with T_f and $g \in G$, if T_f is a G -homomorphism then considering the G -invariant inner product on $\mathbb{C}X$ as in theorem 4.1.12 we have that

$$\langle \theta_f(gx), gy \rangle_{\mathbb{C}X} = \langle g\theta_f(x), gy \rangle_{\mathbb{C}X} = \langle \theta_f(x), y \rangle_{\mathbb{C}X}$$

Hence f must have coefficients constant on the G -orbits. On the other hand for $x \in X$ as

$$\theta_f(gx) = \sum_{y \in X} \alpha_{gx,y} gx = \sum_{y \in X} \alpha_{gx,gy} gx$$

where the last equality follows from the fact that g acts bijectively on X so it simply permutes the elements of X (by the properties of a group action), we have that if f has coefficients constant on the G -orbits then

$$\theta_f(gx) = \sum_{y \in X} \alpha_{gx,gy} gx = \sum_{y \in X} \alpha_{x,y} gx = g \sum_{y \in X} \alpha_{x,y} x = g\theta_f(x)$$

Thus T_f is a G -homomorphism. □

The above result is the generalization of proposition 4.4.1 we sought. Following the proofs above one can analogously show the following.

Proposition 4.7.6. *Let X and Y be G -sets and let G act on $X \times Y$ as $g(x, y) = (gx, gy)$ for $(x, y) \in X \times Y$, then considering the associated permutation representations*

$$\text{Hom}_G(\mathbb{C}X, \mathbb{C}Y) \cong \mathbb{C}(X \times Y)^G$$

as algebras.

In the setting of the above proposition, by the analogous proposition 4.7.5 to this case the characteristic functions of the G -orbits, that is letting $G(x_i, y_i)$, $i \in [k]$, be the G -orbits of $X \times Y$ the functions $\Delta_i : X \times Y \rightarrow \mathbb{C}$ such that $\Delta_i((z, t)) = 1$ if $(z, t) \in G(x_i, y_i)$ and zero otherwise, are a basis of $C(X \times Y)^G$ considering that functions in $C(X \times Y)^G$ are constant in the G -orbits, and so are mapped to a basis of $\text{Hom}_G(\mathbb{C}X, \mathbb{C}Y)$, it's then easy to see by the result analogous to theorem 4.7.1 that those basis elements/matrices are defined as

$$(T_{\Delta_i})_{xy} = \begin{cases} 1 & \text{if } (x, y) \in G(x_i, y_i) \\ 0 & \text{otherwise} \end{cases}$$

for $i \in [k]$, $(x, y) \in X \times Y$. The map above is called the *Radon Transform* associated with $G(x_i, y_i)$.

Corollary 4.7.7. *Let X be a G -set and consider the canonical decomposition of the respective permutation representation, $\oplus_{V_i \in \widehat{G}} n_i V_i$, moreover let k be the number of G -orbits on $X \times X$, then*

$$\sum_{V_i \in \widehat{G}} n_i^2 = k$$

Proof. By proposition 4.7.5 $\dim_{\mathbb{C}}(\text{End}_G(\mathbb{C}X)) = \dim_{\mathbb{C}}(\mathbb{C}(X \times X)^G) = k$, on the other hand by corollary 4.3.21 $\dim_{\mathbb{C}}(\text{End}_G(\mathbb{C}X)) = \sum_{V_i \in \widehat{G}} n_i^2$. □

If we consider the above corollary in the setting of proposition 4.7.6 then by our discussion above the following comes just as easily.

Proposition 4.7.8. *Let X, Y be G -sets with respective permutation representations λ_1, λ_2 affording characters $\chi_{\lambda_1}, \chi_{\lambda_2}$ and let k be the number of G -orbits on $X \times Y$ then*

$$\dim_{\mathbb{C}}(\text{Hom}_G(\mathbb{C}X, \mathbb{C}Y)) = \langle \chi_{\lambda_1}, \chi_{\lambda_2} \rangle_G = k$$

Hence the multiplicity of any permutation representation $\mathbb{C}Y$ contained in a permutation representation $\mathbb{C}X$ is given by the orbits of G on $X \times Y$ and in particular we have the following.

Corollary 4.7.9. *The multiplicity of the trivial representation in a permutation representation λ is equal to the number of orbits of G on X .*

Proof. Note that if we let Y be a G -set such that $|Y| = 1$ then $\mathbb{C}Y$ is a one dimensional space with trivial G action and so it's the trivial representation, now using proposition 4.7.8

and noting that the orbits of G on $X \times Y$ coincide with the orbits of G on X we get that $\langle \chi_i, \chi_X \rangle_G = \langle \chi_Y, \chi_X \rangle_G$ is equal to the number of orbits of G on X . $\square$

Example 4.7.10. Consider the natural action of S_n on $[n]$ and the corresponding permutation representation λ , as we've seen in example 4.1.32 this representation decomposes as

$$\mathbb{C}X = W \oplus U$$

Where W and U are irreducible G -spaces, thus $n_1 = n_2 = 1$. As the action of S_n in $[n]$ is transitive, that is has only one orbit as for any $m, p \in [n]$ there exists $\pi \in S_n$ such that $\pi(m) = p$, then $[n] \times [n]$ decomposes into only two orbits, the orbit generated by $(1, 1)$ and the orbit generated by say $(1, 2)$. Hence indeed this is the canonical decomposition of λ as it has two non-equivalent irreducible sub-representations and $n_1^2 + n_2^2 = 2 = k$.

We now focus on the characters of permutation representations, recall that by proposition 4.3.17 these characters are real functions thus by proposition 4.3.29 the corollary below follows immediately.

Corollary 4.7.11. *A permutation representation is self-dual.*

Hence the fact that a regular representation is self-dual follows more generally from the fact that a permutation representation is self-dual.

In what follows let X be a G -set and consider the respective permutation representation λ affording character χ_X .

Proposition 4.7.12. *The space $I = \langle \sum_{x \in X} x \rangle$ is a G -subspace of $\mathbb{C}X$ trivial for the action of G and has complementary G -space $W = \{ \sum_{x \in X} \alpha_x x \mid \sum_{x \in X} \alpha_x = 0 \}$. In particular the permutation representation always contains the trivial representation.*

Proof. By proposition 4.3.17

$$\langle \chi_i, \chi_X \rangle_G = \frac{1}{|G|} \sum_{g \in G} \chi_X(g)$$

and as $\chi_X(g)$ is a non-negative integer for all $g \in G$ then

$$\langle \chi_i, \chi_X \rangle_G \geq \frac{1}{|G|} \chi_X(1_G) = \frac{|X|}{|G|} > 0$$

So we know right from the get go that λ must contain the trivial representation, yet we can give a constructive proof of this by noting that the subspace I is not only G -invariant but

also G -fixed and so G acts trivially on this space, as this space has dimension 1 it must be the trivial representation. The complement of this space is, as one can easily verify W and it is clearly G -invariant. $\square$

Note that for the G -space W above we can no longer straightforwardly say as in example 4.1.32) that it's irreducible, in fact it might well be the case that it too contains the trivial representation, just imagine an element $f \in W$ that is constant on the G -orbits, yet there is a clear way to determine when this space is irreducible. In fact we already used it in example 4.7.10.

Definition 4.7.13. Let X be a G -set and consider the action of G on $X \times X$ as $g(x, y) = (gx, gy)$. We say the action of G is 2-transitive if $X \times X$ has only two G -orbits, those forcibly being $\{(x, x) \mid x \in X\}$ and $\{(x, y) \mid x, y \in X, x \neq y\}$.

Theorem 4.7.14. Let X be a G -set with $|X| > 1$, then

$$\chi_X = \iota_G \oplus \chi_W$$

Where W is the space as in proposition 4.7.12. Moreover W is irreducible if and only if the action of G is 2-transitive.

Proof. Indeed by proposition 4.7.12, with the same notation

$$\mathbb{C}X = I \oplus W$$

and as we've seen I is the trivial representation. If the space W is irreducible then by corollary 4.7.7 the action of G on $X \times X$ has only two orbits and in the same way if it has only two orbits then we must have W irreducible. $\square$

Proposition 4.7.15. Let X, Y be G -sets and consider the action of G on $X \times Y$ as $g(x, y) = (gx, gy)$, then

$$\chi_{X \times Y}(g) = \chi_X(g)\chi_Y(g)$$

for $g \in G$.

Proof. This follows from part c) of proposition 4.3.12 by noting that $\mathbb{C}(X \times Y) = \mathbb{C}X \otimes \mathbb{C}Y$, recalling definition 4.1.43 of tensor product space, and that these spaces have the same action of G by definition 4.1.44. $\square$

4.7.1 Gelfand pairs

Put simply theorem 4.7.14 shows the following.

Corollary 4.7.16. *If the action of G on a set X is 2-transitive then the respective permutation representation is multiplicity free as in definition 4.2.15, hence $\text{End}_G(\mathbb{C}X)$ is commutative.*

Proof. This follows immediately from corollary 4.2.17 and theorem 4.7.14. □

Note that if $X = G$ then by proposition 4.4.1 if the action we define of G on itself is 2-transitive then $\mathbb{C}G$ is commutative and so G is commutative.

Proposition 4.7.17. *Let G be a group, if there exists a 2-transitive action of G on itself then G is commutative.*

When the action of a group G on set X is transitive then recalling remark 2 we can equivalently work with the action of G on the set of left cosets G/K by left multiplication, where $K = \{g \in G \mid gx = x\}$ is the stabilizer of an element $x \in X$. Hence for any G -set X with transitive action there's an equivalent G -set G/H with H a subgroup of G , and going the other way around for any H subgroup of G , G/H is a G -set with transitive action by left multiplication*, thus when working with permutation representations given by a transitive action we can focus solely on G , its subgroups and the action of G by left multiplication. We note that transitivity does not imply 2-transitivity; we already know that 2-transitivity results in a multiplicity free decomposition so now we ask what happens if we relax 2-transitivity to just transitivity.

By example 4.3.23 the regular representation of S_3 does not have a multiplicity free decomposition into irreducibles, yet the action of S_3 on itself is transitive as for any $g, h \in S_3$, $(hg^{-1})g = h$, hence transitivity does not imply multiplicity free decomposition.

Definition 4.7.18. *Let G be a group, H a subgroup and consider the action of G on G/H by left multiplication. Then if the corresponding permutation representation is multiplicity free we say (G, H) is a Gelfand pair.*

If the action of G is 2-transitive then by corollary 4.7.16, (G, H) is a Gelfand pair, in this case we say we have a *symmetric* Gelfand pair.

Gelfand pairs can be easily characterized in terms of H -fixed subspaces of irreducible representations of G . For that we first need a generalization of corollary 4.3.20.

*The action of G on G/H by left multiplication is always transitive as for any $k_1H, k_2H \in G/H$, $k_2k_1^{-1}k_1H = k_2H$.

Theorem 4.7.19. *Let ρ_V be a representation of G affording the character χ_ρ , then*

$$\frac{1}{|G|} \sum_{g \in G} \chi_\rho(g) = \dim_C(V^G)$$

Proof. First note that, letting $e = \frac{1}{|G|} \sum_{g \in G} g$

$$\frac{1}{|G|} \sum_{g \in G} \chi_\rho(g) = \frac{1}{|G|} \sum_{g \in G} \text{tr}(\rho(g)) = \text{tr} \left(\frac{1}{|G|} \sum_{g \in G} \rho(g) \right) = \text{tr}(\hat{\rho}(e))$$

where $\hat{\rho}$ is the Fourier transform of G at ρ as in 4.9. Because $e^2 = e$ and $\hat{\rho}$ is a homomorphism then $\hat{\rho}(e)\hat{\rho}(e) = \hat{\rho}(e)$ hence by proposition C.0.12 and C.0.13 it is diagonalizable and can have eigenvalues 0 and 1. Now the eigenspace of 1 is

$$S = \{v \in V \mid \hat{\rho}(e)v = v\}$$

and $\hat{\rho}(e)v = \frac{1}{|G|} \sum_{g \in G} \rho_g v$ hence if $g \in G$ and $v \in S$ then

$$\rho(h)v = \rho(h)\hat{\rho}(e)v = \hat{\rho}(e)v = v$$

Thus $v \in V^G$, on the other hand it's clear that for any $v \in V^G$, $\hat{\rho}(e)v = v$. Therefore $S = V^G$ and so $\text{tr}(\hat{\rho}(e)) = \dim_C(V^G)$ completing the proof. $\square$

Corollary 4.7.20. *Let V, W be G -spaces, then*

$$\text{Hom}_G(W, V) \cong (W \otimes V^*)^G$$

as vector spaces over $\mathbb{C}$.

Proof. Indeed by part b) and c) of proposition 4.3.12, corollary 4.3.21 and the previous theorem

$$\begin{aligned} \dim_{\mathbb{C}}(\text{Hom}_G(W, V)) &= \langle \chi_W, \chi_V \rangle_G = \frac{1}{|G|} \sum_{g \in G} \chi_W(g) \overline{\chi_V(g)} = \\ &= \frac{1}{|G|} \sum_{g \in G} \chi_{W \otimes V^*}(g) = \langle \chi_{W \otimes V^*}, \chi_1 \rangle_G = \dim_C((W \otimes V^*)^G) \end{aligned}$$

$\square$

The above corollary is a particular case of a more general result establishing $\text{Hom}(W, V) \cong W \otimes V^*$ which follows easily by noting that both vector spaces over $\mathbb{C}$ have the same dimension, this can be made precise by simply organizing the coefficients of elements of $W \otimes V^*$ in a matrix of size $\dim_{\mathbb{C}}(V) \times \dim_{\mathbb{C}}(W)$ thus producing a map $W \otimes V^* \rightarrow$

$\text{Hom}(W, V)$.

In our case letting V be a permutation representation then by corollary 4.7.11, V is self-dual so $V^* \cong V$ as G -spaces, hence $\dim_{\mathbb{C}}((W \otimes V^*)^G) = \dim_{\mathbb{C}}((W \otimes V)^G)$. Considering this we have the following.

Proposition 4.7.21. *Let G be a group and H a subgroup, then (G, H) is a Gelfand pair if and only if for all $\rho_W \in \widehat{G}$, $\dim_{\mathbb{C}}((W \otimes V)^G) \leq 1$, where $V = \mathbb{C}(G/H)$.*

Proof. This follows immediately from the previous corollary and the fact that V is self-dual. □

Theorem 4.7.22. *Let G be a group and H a subgroup such that G acts on $X = G/H$ by left multiplication, then for the corresponding permutation representation and $\rho_W \in \widehat{G}$*

$$\text{Hom}_H(\iota_H, W) \cong \text{Hom}_G(W, \mathbb{C}X)$$

as vector spaces over $\mathbb{C}$, where ι_H is the trivial representation of H . In particular $W^H \cong \text{Hom}_G(W, \mathbb{C}X)$.

Proof. By Nakayama's reciprocity 4.6.9 and example 4.6.11

$$\text{Hom}_H(W|_H, \iota_H) \cong \text{Hom}_G(W, \mathbb{C}X)$$

By theorem 4.3.31, $\text{Hom}_H(W|_H, \iota_H) \cong \text{Hom}_H(\iota_H, W|_H)$ thus

$$\text{Hom}_H(\iota_H, W|_H) \cong \text{Hom}_G(W, \mathbb{C}X)$$

The result then follows from theorem 4.7.19 by noting that $\dim_{\mathbb{C}}(\text{Hom}_H(\iota_H, W|_H)) = \langle \chi_{\iota}, \chi_W \rangle_H = \dim_{\mathbb{C}}(W^H)$. □

The result we seek is now an immediate consequence of the previous theorem.

Corollary 4.7.23. *In the condition of the previous theorem*

$$\dim_{\mathbb{C}}(\text{Hom}_G(W, \mathbb{C}X)) = \dim_{\mathbb{C}}(W^H)$$

In particular (G, H) is a Gelfand pair if and only if for all $\rho_W \in \widehat{G}$, $\dim_{\mathbb{C}}(W^H) \leq 1$.

Although an interesting result, one must admit it doesn't seem to be very practical, we must strive to achieve better. As it turns out the *double cosets* hold the answer, we start with a couple definitions.

Definition 4.7.24. Let G be a group and H a subgroup, we define the double cosets of H in G as $H \backslash G / H = \{HgH \mid g \in G\}$. The double cosets HgH are determined by the equivalence relation $g \sim k \Leftrightarrow \exists_{h_1, h_2 \in H} g = h_1 k h_2$, for $g, k \in G$, thus they determine a partition of G , that is G equals the disjoint union of the double cosets.

Definition 4.7.25. Let G be a group and H a subgroup, we define $L(H \backslash G / H)$ to be the subspace of the algebra $\mathbb{C}G$ whose elements have coefficients constant on the double cosets, that is

$$L(H \backslash G / H) = \left\{ \sum_{g \in G} \alpha_g g \in \mathbb{C}G \mid \text{for all } h_1, h_2 \in H, g \in G; \alpha_{h_1 g h_2} = \alpha_g \right\}$$

The space $L(H \backslash G / H)$ is in fact a subalgebra of $\mathbb{C}G$, indeed for $f_1, f_2 \in L(H \backslash G / H)$ let $[f_1 f_2](g)$ be the coefficient of $g \in G$ in $f_1 f_2$, then for $h_1, h_2 \in H$ and $k \in G$

$$[f_1 f_2](h_1 k h_2) = \sum_{gh = h_1 k h_2} \alpha_g \beta_h = \sum_{gh = k} \alpha_{h_1 g} \beta_{h h_2}$$

Because $f_1, f_2 \in L(H \backslash G / H)$ then by the definition of this space $\alpha_{h_1 g} = \alpha_g$ and $\beta_{h h_2} = \beta_h$, hence $[f_1 f_2](h_1 k h_2) = [f_1 f_2](k)$ and so $f_1 f_2 \in L(H \backslash G / H)$.

Theorem 4.7.26. Let G be a group and H a subgroup and $X = G/H$ the respective G -set, then

$$\mathbb{C}(X \times X)^G \cong L(H \backslash G / H)$$

as algebras.

Proof. Consider the map

$$\begin{aligned} \Phi : \mathbb{C}(X \times X)^G &\rightarrow L(H \backslash G / H) \\ f &\mapsto \hat{f} \end{aligned}$$

such that $\hat{f}(g) = \frac{1}{|H|} f(H, gH)$. We first note this map is well defined, indeed for $h_1, h_2 \in H$

$$\hat{f}(h_1 g h_2) = \frac{1}{|H|} f(H, h_1 g h_2 H) = \frac{1}{|H|} f(h_1^{-1} H, gH) = \frac{1}{|H|} f(H, gH) = \hat{f}(g)$$

as $f(h_1^{-1}(H, h_1 g h_2 H)) = f(H, h_1 g h_2 H)$ and $h_2 H = H = h_1^{-1} H$. This map is clearly linear so we need only to prove that it's a bijective ring homomorphism. We first prove it's bijective, for that let $\{G(x_i, y_i)\}_{i \in [n]}$ be the orbits of G on $X \times X$ and consider the elements $f_i \in \mathbb{C}(X \times X)$ that have coefficient equal to 1 in the elements of $G(x_i, y_i)$ and 0 elsewhere. Clearly these form a basis of $\mathbb{C}(X \times X)^G$.

It's straightforward to see that any orbit $G(x_i, y_i)$ has a representative of the form (H, gH)

for some $g \in G$, indeed $k_1^{-1}(k_1H, k_2H) = (H, h_1^{-1}h_2H)$ for any $k_1, k_2 \in G$, moreover

$$\begin{aligned} (H, gH) = (H, kH) &\Leftrightarrow \exists_{h_1 \in G} h_1(H, gH) = (H, kH) \Leftrightarrow \exists_{h_1 \in H} h_1gH = kH \Leftrightarrow \\ &\exists_{h_1, h_2 \in H} h_1g = kh_2 \Leftrightarrow \exists_{h_1, h_2 \in H} h_1gh_2^{-1} = k \end{aligned}$$

This is precisely the equivalence relation that determined the double cosets, hence

$$(H, gH) = (H, kH) \Leftrightarrow HgH = HkH$$

thus if $\{(H, g_iH)\}_{i \in [n]}$ are representative of the G orbits on $X \times X$ then $\{g_i\}_{i \in [n]}$ are representatives of the double cosets and

$$\hat{f}_i(g) = \begin{cases} \frac{1}{|H|} & \text{if } g \in Hg_iH \\ 0 & \text{otherwise} \end{cases}$$

Therefore $\{\hat{f}_i\}_{i \in [n]}$ is a basis of $L(H \backslash G / H)$, thus Φ is bijective. Finally note that

$$\begin{aligned} \widehat{f_1 f_2}(g) &= \frac{1}{|H|} [f_1 f_2](H, gH) = \frac{1}{|H|} \sum_{kH \in G/H} f_1(H, kH) f_2(kH, gH) = \\ &= \frac{1}{|H|} \sum_{kH \in G/H} f_1(H, kH) f_2(H, k^{-1}gH) = \frac{1}{|H|} \sum_{kH \in G/H} \hat{f}_1(k) \hat{f}_2(k^{-1}g) = \\ &= \frac{1}{|H||H|} \sum_{kh=g} \hat{f}_1(k) \hat{f}_2(h) = [\hat{f}_1 \hat{f}_2](g) \end{aligned}$$

Hence we have a ring homomorphism and so in conclusion an algebra isomorphism. $\square$

By proposition 4.7.5 $C(X \times X)^G \cong \text{End}_G(\mathbb{C}X)$ hence $L(H \backslash G / H) \cong \text{End}_G(\mathbb{C}X)$, applying corollary 4.2.18 and noting that $L(H \backslash G / H)$ is a commutative algebra if and only if the double cosets $H \backslash G / H$ commute with each other, we have the following.

Corollary 4.7.27. *Let G be a group and H a subgroup, then (G, H) is a Gelfand pair if and only if the double cosets $H \backslash G / H$ commute with each other.*

The following proposition is known as *Gelfand's lemma* and it gives a more useful criterion for Gelfand pairs.

Proposition 4.7.28. *Let G be a group and H a subgroup. If there exists a group automorphism $\psi : G \rightarrow G$ such that for all $g \in G$, $g \in H\psi(g^{-1})H$ then (G, H) is a Gelfand pair.*

Proof. Let ψ be a homomorphism as described above, then for $g, k \in G$ as by our hypothesis we have that $HgH = H\psi(g^{-1})H$

$$\begin{aligned} HgHHkH &= H\psi(g^{-1})H\psi(k^{-1})H = H\psi(g^{-1}Hk^{-1})H = H\psi((kHg)^{-1})H = \\ &= HkHgH = HkHHgH \end{aligned}$$

Hence the double cosets commute with each other and the conclusion follows by the previous corollary. □

We finish this section with a simple example.

Example 4.7.29. Let G be a group and consider the action of $G \times G$ on G by $(h_1, h_2)g = h_1gh_2^{-1}$, it's easy to verify this action is transitive. The stabilizer of 1_G is $\tilde{G} = \{(g, g) \mid g \in G\}$ hence we can take $G = (G \times G)/\tilde{G}$ and consider the action of $G \times G$ on $(G \times G)/\tilde{G}$ by left multiplication.

If we consider the automorphism $\psi(g_1, g_2) = (g_2, g_1)$ for $g_1, g_2 \in G$ then

$$(g_1, g_2) = (g_1, g_1)(g_2^{-1}, g_1^{-1})(g_2, g_2) \in \tilde{G}\psi((g_1, g_2)^{-1})\tilde{G}$$

As this is true for any $g_1, g_2 \in G$ we have by the previous proposition that $(G \times G, \tilde{G})$ is a Gelfand pair.

5. The Okounkov–Vershik approach

The representation theory of Symmetric groups was first studied by *F.G. Frobenius* and *A. Young*, and then further developed by Young, in a long series of papers, since then many author have added to his great volume of work. In this section we present a new development in this theory.

The theory developed by Young and subsequent authors, here called the "Classical approach", resorts to objects called *Young tableaux* to build the foundations of the theory. Albeit giving a quite natural approach to the representation theory of symmetric groups, it's done in a way specifically oriented towards the symmetric group, and so the door was left open for a more general foundation to be constructed. That finally came in 2005, with the paper "A New Approach to the Representation Theory of the Symmetric Groups. 2" [AA05] by *A.M. Vershik* and *A.Yu. Okounkov*, this section is devoted to their work, seeking to fully explain their new approach to the representation theory of Symmetric groups, here called the "Modern approach".

5.1. Multiplicity free subgroups

We've seen restricted representations in definition 4.1.14 and section 4.6, specifically restricting a representation ρ of a group G to a subgroup H of G . It's easy to find examples of irreducible representations for which a restriction to a subgroup does not produce an irreducible representation, for instance the representation σ of S_3 in example 4.3.36 restricted to $H = \{1_{S_3}, (1\ 2)\}$.

Although restrictions to subgroups rarely preserve irreducibility we can ask a more relaxed question: *When does restricting an irreducible representation produce a multiplicity free representation?*

This question may seem quite esoteric now, but in fact it has proven to be very useful in characterizing irreducible representations of groups, in particular of the symmetric groups S_n . When for an irreducible representation ρ_V of G restricted to a subgroup H we get a multiplicity free representation we can decompose the space V uniquely into a direct sum of irreducible H -spaces, which under ideal circumstances can be of use to gain insight into the irreducible representation ρ_V .

Here we'll answer the above question using Gelfand pairs and later in Subsection 5.3.3

we'll take a more module theoretic approach to the question, but first a couple interesting results about restricted representations.

The first result shows that the restriction of an irreducible character of a group G to a subgroup H can contain any (irreducible) character of H .

Proposition 5.1.1. *Let ξ be a non-zero character of a group $H \leq G$, then there exists an irreducible character ψ of G such that*

$$\langle \xi, \psi_H \rangle_H > 0$$

That is, ξ is a constituent of the decomposition of ψ_H into irreducibles.

Proof. Consider the regular representation λ of G affording character χ , as we know this representation contain every irreducible representation of G , using example 4.3.16 and noting that $1_H = 1_G$ we have that

$$\langle \xi, \chi_H \rangle_H = \frac{1}{|H|} \xi(1_H) \chi(1_H) = \frac{|G|}{|H|} \xi(1_H)$$

As ξ is a non-zero character it must be the case that $\xi(1_H) > 0$ hence

$$\langle \xi, \chi_H \rangle_H > 0$$

Thus there exists some irreducible character ψ of G such that

$$\langle \xi, \psi_H \rangle_H > 0$$

□

The following proposition can be of assistance in determining the decomposition of a restricted representation.

Proposition 5.1.2. *Let χ be an irreducible character of a group G and $H \leq G$, additionally assume χ_H decomposes into irreducible characters of H as*

$$\chi_H = \sum_{i \in [n]} n_i \chi_i$$

So χ_i are irreducible characters of H and $n_i \in \mathbb{N}$, for $i \in [n]$. Then

$$\sum_{i \in [n]} n_i^2 \leq |G/H|$$

With equality if and only if $\chi(g) = 0$ for $g \in G \setminus H$.

Proof. Indeed as χ is irreducible then by theorem 4.3.28

$$\begin{aligned} 1 &= \langle \chi, \chi \rangle_G = \frac{1}{|G|} \sum_{g \in G} \chi(g) \overline{\chi(g)} = \\ &= \frac{1}{|G|} \left(\sum_{h \in H} \chi(h) \overline{\chi(h)} + \sum_{g \in G \setminus H} \chi(g) \overline{\chi(g)} \right) \end{aligned}$$

Note that $\frac{1}{|G|} \sum_{h \in H} \chi(h) \overline{\chi(h)} = \frac{|H|}{|G|} \langle \chi_H, \chi_H \rangle_H$ hence by 4.3.21 the above is equal to

$$\frac{|H|}{|G|} \sum_{i \in [n]} n_i^2 + \frac{1}{|G|} \sum_{g \in G \setminus H} \chi(g) \overline{\chi(g)}$$

As $\chi(g) \overline{\chi(g)} = |\chi(g)|^2 \geq 0$ then

$$\frac{|H|}{|G|} \sum_{i \in [n]} n_i^2 + \frac{1}{|G|} \sum_{g \in G \setminus H} \chi(g) \overline{\chi(g)} \geq \frac{|H|}{|G|} \sum_{i \in [n]} n_i^2$$

Therefore

$$1 \geq \frac{|H|}{|G|} \sum_{i \in [n]} n_i^2$$

and the conclusion follows immediately. □

Note the above result is only useful if $|G/H|$ is small or we have significant information beforehand of the decomposition of χ_H , not unlike corollary 4.7.7.

We now move to deeper matters.

Definition 5.1.3. Let H be a subgroup of a group G , we say H is a multiplicity free subgroup of G if for every $\rho \in \widehat{G}$ the restriction $\rho|_H^G$ is a multiplicity free representation, that is for every $\sigma \in \widehat{H}$ and $\rho \in \widehat{G}$, $\dim_{\mathbb{C}}(\text{Hom}_H(\sigma, \rho|_H^G)) \leq 1$.

As stated initially we'll use Gelfand pairs to characterize multiplicity free subgroups, in particular we'll see that there's a relation between restricted representations and specific permutation representations involving both the group G and the subgroup H .

Let $H \leq G$ and consider the action of $G \times H$ on G as

$$(g, h) \cdot k = gkh^{-1}$$

for $(g, h) \in G \times H$ and $k \in G$. We denote the associated permutation representation by η . Note that the stabilizer of 1_G is

$$\tilde{H} = \{(h, h) \mid h \in H\}$$

Thus by remark 2 we can equivalently consider the action of $G \times H$ on $(G \times H)/\tilde{H}$ by left multiplication. Recall that by theorem 4.3.34 the irreducible representations of $G \times H$ are of the form $\rho \otimes \sigma$ for $\rho \in \hat{G}$ and $\sigma \in \hat{H}$.

Lemma 5.1.4. *Consider $H \leq G$, the representation η above and let $\rho \in \hat{G}$ and $\sigma \in \hat{H}$, then*

$$\text{Hom}_{G \times H}(\rho \otimes \sigma, \eta) \cong \text{Hom}_H(\sigma^\circ, \rho|_H)$$

as vector spaces (over $\mathbb{C}$).

Proof. Indeed by corollary 4.3.21 and part d) of proposition 4.3.12

$$\begin{aligned} \dim_{\mathbb{C}}(\text{Hom}_{G \times H}(\rho \otimes \sigma, \eta)) &= \langle \chi_{\rho \otimes \sigma}, \chi_\eta \rangle_{G \times H} = \\ &= \frac{1}{|G||H|} \sum_{\substack{g \in G, \\ h \in H}} \chi_\rho(g) \chi_\sigma(h) \overline{\chi_\eta(g, h)} \end{aligned}$$

We note that by proposition 4.3.17, $\chi_\eta(g, h) = |\{k \in G \mid gkh^{-1} = k\}|$, but $gkh^{-1} = k \Leftrightarrow g = khk^{-1}$ hence g must be a conjugate of h on G for $\chi_\eta(g, h)$ to be non-zero. As characters are class functions then letting J_h be the conjugacy class of h the above reduces to

$$\frac{1}{|G||H|} \sum_{h \in H} |J_h| \chi_\rho(h) \chi_\sigma(h) \overline{\chi_\eta(h, h)}$$

Now note that $\chi_\eta(h, h) = |\{k \in G \mid hkh^{-1} = k\}| = |C_G(h)|$ where $C_G(h)$ is the centralizer of h on G , thus we get

$$\dim_{\mathbb{C}}(\text{Hom}_{G \times H}(\rho \otimes \sigma, \eta)) = \frac{1}{|G||H|} \sum_{h \in H} |J_h| \chi_\rho(h) \chi_\sigma(h) |C_G(h)|$$

As is know from group theory [Rot95, 3.2]

$$\frac{|G|}{|C_G(h)|} = |J_h| \Leftrightarrow \frac{|J_h|}{|G|} = \frac{1}{|C_G(h)|}$$

therefore

$$\begin{aligned} \dim_{\mathbb{C}}(\text{Hom}_{G \times H}(\rho \otimes \sigma, \eta)) &= \frac{1}{|H|} \sum_{h \in H} \chi_\rho(h) \chi_\sigma(h) = \\ &= \langle \chi_\rho, \chi_{\sigma^\circ} \rangle_H = \dim_{\mathbb{C}}(\text{Hom}_H(\rho|_H, \sigma^\circ)) \end{aligned}$$

The conclusion now follows by theorem 4.3.31. □

By the above result the multiplicity of an irreducible representation $\rho \otimes \sigma$ in η , with $\rho \in \hat{G}$ and $\sigma \in \hat{H}$ is equal to the multiplicity of the dual representation σ° in $\rho|_H$. This is almost what we're looking for, all we need is a final remark.

First by our discussion preceding definition 4.1.40 as σ is irreducible then σ° is also irreducible, moreover we can produce a bijection from $\widehat{H}$ to $\widehat{H}$ by $\sigma \mapsto \sigma^\circ$ for $\sigma \in \widehat{H}$, indeed as $\chi_{(\sigma^\circ)^\circ} = \overline{\chi_\sigma} = \chi_\sigma$ the aforementioned map must be injective and as $\widehat{H}$ is a finite set then it must also be surjective. Considering this we have the following theorem.

Theorem 5.1.5. *Let $H \leq G$ and consider the representation η , then H is a multiplicity free subgroup of G if and only if $(G \times H, \tilde{H})$ is a Gelfand pair.*

Proof. Indeed by the previous lemma for $\rho \in \widehat{G}$ and $\sigma \in \widehat{H}$ if $(G \times H, \tilde{H})$ is a Gelfand pair then $\dim_{\mathbb{C}}(\text{Hom}_H(\sigma^\circ, \rho|_H)) = \dim_{\mathbb{C}}(\text{Hom}_{G \times H}(\rho \otimes \sigma, \eta)) \leq 1$, and because σ is any in $\widehat{H}$ then by the bijection mentioned above we get that for any $\sigma \in \widehat{H}$ and any $\rho \in \widehat{G}$, $\dim_{\mathbb{C}}(\text{Hom}_H(\sigma, \rho|_H)) \leq 1$, thus by definition 5.1.3 H is a multiplicity free subgroup of G .

On the other hand if H is a multiplicity free subgroup of G then for any $\sigma \in \widehat{H}$ and any $\rho \in \widehat{G}$, $\dim_{\mathbb{C}}(\text{Hom}_H(\sigma, \rho|_H)) \leq 1$, and again by the previous theorem and the aforementioned bijection on $\widehat{H}$ it must be the case that for any $\sigma \in \widehat{H}$ and any $\rho \in \widehat{G}$, $\dim_{\mathbb{C}}(\text{Hom}_{G \times H}(\rho \otimes \sigma, \eta)) \leq 1$, thus $(G \times H, \tilde{H})$ is a Gelfand pair. $\square$

We can now use proposition 4.7.28 to get the following result, recall $\tilde{H} = \{(h, h) \mid h \in H\}$.

Corollary 5.1.6. *Let $H \leq G$. If there exists a group automorphism $\psi : G \times H \rightarrow G \times H$ such that for all $(g, h) \in G \times H$, $(g, h) \in \tilde{H}\psi((g^{-1}, h^{-1}))\tilde{H}$ then H is a multiplicity free subgroup of G .*

5.2. Classical representation theory of S_n

Characterizing the irreducible representations of the symmetric groups S_n following the classic approach as developed by *F.G. Frobenius*, *I. Schur* and *A. Young* requires a non trivial amount of work even resorting to more modern simplification, here we'll do a short exposition of the theory not focusing on proofs, for a more extensive treatment refer to [Sag01] and [Ful96].

To understand the modern approach by *Okounkov* and *Vershik* one must first understand where it comes from, as such this section introduces the necessary concepts that we'll use later.

5.2.1 Young tableaux

Considering theorem 4.3.32 the group S_n must have a number of (non-equivalent) irreducible representations equal to the number of conjugacy classes of S_n which as is known is equal to the number of possible cycle structures of a permutation of $[n]$, moreover these cycle structures are in bijection with partitions of n , as such one could expect the existence of some relation between partitions of n and irreducibles of S_n . We first properly define partitions of $n \in \mathbb{N}$.

Definition 5.2.1. For $n \in \mathbb{N}$ a partition of n is defined as a tuple

$$\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k)$$

such that $k, \lambda_i \in \mathbb{N}$, $\sum_{i \in [k]} \lambda_i = n$ with the elements of λ written in decreasing order, that is $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$. We write $\lambda \vdash n$ to indicate λ is a partition of n , the set of all partition of n is denoted by $P(n)$.

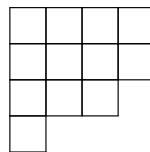
The above notation can be extended in the following way, for $\pi \in S_n$ with decomposition into disjoint cycles

$$\pi = (a_1 a_2 \dots a_{\lambda_1})(b_1 b_2 \dots b_{\lambda_2}) \dots (x_1 x_2 \dots x_{\lambda_k})$$

where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$ let $\lambda(\pi) = (\lambda_1, \lambda_2, \dots, \lambda_k)$, for instance if $\pi = (1\ 3\ 2)(4\ 5) \in S_5$ then $\lambda(\pi) = (3, 2)$, or in other words π has cycle structure $(3, 2)$. Under this definition λ is a function from S_n to $P(n)$ and with it, noting that, for $\sigma, \lambda \in S_n$, $\lambda(\sigma) = \lambda(\pi)$ if and only if σ and π are conjugate, it becomes then clear that indeed λ produces a bijection from the conjugacy classes of S_n to $P(n)$.

The first step towards defining Young tableaux are *Ferrers diagrams*, these are essentially a visual representation of partitions.

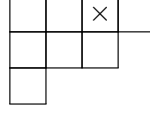
Definition 5.2.2. Given a partition $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k) \vdash n$ we represent it diagrammatically with a pattern of boxes/cells arranged in rows, one row for each entry in λ with each row having λ_i boxes respectively, for instance if we let $\lambda = (4, 4, 3, 1)$ then we represent it with the diagram



This is then called the Ferrers diagram of λ .*

*Traditionally Ferrers diagrams use dots instead of boxes, so what we defined above would more correctly be called a Young diagram, yet to avoid confusion we'll continue to call them Ferrers diagrams.

Each box in a Ferrers diagram has associated coordinates (i, j) , for instance considering $\lambda = (4, 3, 1)$ then $i \in [3], j \in [4]$ and its following Ferrers diagram has an $\times$ in position $(1, 3)$



Given a Ferrers diagram of a partition $\lambda \vdash n$ we say a box is removable/addable if upon removing/adding that box the resulting diagram is a Ferrers diagram of some partition $\mu \vdash n \pm 1$. In the example above the boxes with coordinates $(1, 4), (2, 3), (3, 1)$ are removable, and we can add boxes in position $(1, 5), (2, 4), (3, 2), (4, 1)$.

Definition 5.2.3. Given $\lambda \vdash n$, a Young tableau of shape λ is a Ferrers diagram of λ where each box is populated with an element of $[n]$ bijectively, that is each box has an unique element of $[n]$ in it and each element of $[n]$ is uniquely in a box of the Ferrers diagram of λ . To simplify notation we refer to a Young tableau of shape λ as a λ -tableau and denote such a tableau by t^λ or simply t if λ is clear from context.

As an example letting $\lambda = (4, 3, 1, 1)$ the following is a Young tableau of shape λ .

4	6	3	1
2	9	8	
7			
5			

Considering the above it's immediate to note that for a given $\lambda \vdash n$ there are $n!$ λ -tableaux. If we call the above λ -tableau t then we can refer to elements in a particular box by subscripting its coordinates, so that for example $t_{2,3} = 8$ and $t_{3,1} = 7$, this also allows us to define a λ -tableau by specifying the elements in each coordinate.

For $\lambda \vdash n$, on the set of all Young tableau of shape λ we define an equivalence relation.

Definition 5.2.4. Given $\lambda \vdash n$, we say two λ -tableaux t_1, t_2 are row equivalent, denoted $t_1 \sim t_2$ if they contain the same elements in each row. For a given λ -tableau t we call its respective equivalence class a λ -tabloid and denote it by $\{t\}$, so that

$$\{t\} = \{t' \mid t' \text{ is a } \lambda\text{-tableau, } t' \sim t\}$$

It's straightforward to count the number of Young tableaux in a tabloid, if we let $\lambda =$

$(\lambda_1, \lambda_2, \dots, \lambda_k) \vdash n$ then a λ -tabloid $\{t\}$ has $\lambda_1! \lambda_2! \dots \lambda_k!$ tableaux in it, obtained by permuting the elements in each row of t . As such the total number of λ -tabloids is then

$$\frac{n!}{\lambda_1! \lambda_2! \dots \lambda_k!} \quad (5.1)$$

We can easily define an action of S_n on λ -tableaux, for $\lambda \vdash n$. Let $\pi \in S_n$ and t a λ -tableau such that its coordinates are valid for (i, j) with $i \in [k]$ and $j \in [m]$, then

$$\pi t = (\pi(t_{ij}))_{\substack{i \in [k], \\ j \in [m]}} \quad (5.2)$$

As an example let $n = 3$, $\pi = (1\ 2\ 3)$, $\lambda = (2, 1)$ and t the λ -tableau

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

then

$$\pi t = (1\ 2\ 3) \begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & \\ \hline \end{array} = \begin{array}{|c|c|} \hline 2 & 3 \\ \hline 1 & \\ \hline \end{array}$$

This naturally induces an action on λ -tabloids as $\pi\{t\} = \{\pi t\}$. It's straightforward to verify that this action is well defined, in the sense that it does not depend on the representative t of the class/tabloid, indeed if $\{t_1\} = \{t_2\}$ then t_1 and t_2 have the same elements in each row and so it's clear that any π will preserve this property, thus $\{\pi t_1\} = \{\pi t_2\}$.

For a given $\lambda \vdash n$ let $\mathcal{T}_\lambda$ be the set of all λ -tabloids.

Definition 5.2.5. Let $\lambda \vdash n$ and consider $\mathcal{T}_\lambda$, additionally let $\mathbb{C}\mathcal{T}_\lambda$ be the vector space over $\mathbb{C}$ with basis the elements of $\mathcal{T}_\lambda$. We can extend the action of S_n to $\mathbb{C}\mathcal{T}_\lambda$ by linearity thus producing an S_n -space and so a representation of S_n , we denote this S_n -space by M^λ .

The dimension of M^λ is equal to the number of λ -tabloids, thus given by 5.1. As it turns out the irreducible representations of S_n are obtained by considering an irreducible S_n -subspace of M^λ for each $\lambda \vdash n$.

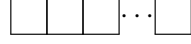
Example 5.2.6. Let $\lambda = (1, 1, \dots, 1) \vdash n$, this partition has Ferrers diagram

$$\begin{array}{c} \square \\ \square \\ \square \\ \vdots \\ \square \end{array}$$

Each λ -tabloid contains a single Young tableau obtained by populating each box with a number $i \in [n]$, hence we have $n!$ tabloids and we can associate each one with a permutation in S_n . From this it's clear that the action of S_n on $\mathcal{T}_\lambda$ is the same as the

action of S_n on itself by left multiplication, hence the S_n -space M^λ is the same as the one obtained from the regular representation of S_n .

On the other hand if we were to consider the partition $\lambda = (n) \vdash n$, which has Ferrers diagram



It's easy to see we have a single λ -tabloid $\{t\}$ and so S_n acts trivially on this tabloid, thus M^λ is the trivial representation of S_n .

5.2.2 Specht modules

As stated the irreducible representations of S_n are obtained by considering an irreducible S_n -subspace of M^λ for each $\lambda \vdash n$, these S_n -subspaces are called *Specht modules*, denoted S^λ .

Recall that given a finite set X we can define the symmetric group S_X as the group of permutation on the elements of X and if $|X| = n$ then $S_X \cong S_n$. Given a λ -tableau t with $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k)$ we denote its rows by $R_i, i \in [k]$ and its columns by $C_i, i \in [\lambda_1]$, so if $\lambda = (3, 2) \vdash 5$ and

$$t = \begin{array}{|c|c|c|} \hline 4 & 5 & 3 \\ \hline 2 & 1 & \\ \hline \end{array} \quad (5.3)$$

then $R_1 = \{4, 5, 3\}$, $R_2 = \{2, 1\}$ and $C_1 = \{4, 2\}$, $C_2 = \{5, 1\}$, $C_3 = \{3\}$.

Definition 5.2.7. Let t be a λ -tableau with rows $\{R_i\}_{i \in [k]}$ and columns $\{C_i\}_{i \in \lambda_1}$. We define the groups

$$R_t = S_{R_1} \times S_{R_2} \times \dots \times S_{R_k}$$

$$C_t = S_{C_1} \times S_{C_2} \times \dots \times S_{C_{\lambda_1}}$$

These are called the row-stabilizer and column-stabilizer groups of t , respectively. We see these groups as acting on t by each component acting on the respective row, or column.*

Considering $\lambda \vdash n$ the above groups have a natural imbedding in S_n . Note that given a Young tableau t , the tabloid $\{t\}$ equals $R_t t = \{\pi t \mid \pi \in R_t\}$. The above groups can be used to construct special elements of the group algebra $\mathbb{C}S_n$, namely

$$r_t^+ = \sum_{\pi \in R_t} \pi$$

*Note that, letting $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k) \vdash n$, then $R_t \cong S_{\lambda_1} \times S_{\lambda_2} \times \dots \times S_{\lambda_k}$, this group is called the *Young subgroup* of S_n associated with λ , usually denoted S_λ (not to be confused with S^λ , the Specht module of λ).

$$c_t^- = \sum_{\pi \in C_t} \text{sgn}(\pi) \pi$$

where the function $\text{sgn} : S_n \rightarrow \mathbb{C}$ equals 1 if π is composition of an even number of transpositions and -1 if it's a composition of an odd number of transpositions (recall every permutation can be decomposed as a composition of transpositions, that is cycles of length 2).

Definition 5.2.8. Let t be a Young tableau and consider the following element of M^λ

$$e_t = c_t^- \{t\}$$

This is called the polytabloid of t .

As an example consider the tableau in 5.3, then

$$c_t^- = id_{S_6} - (2\ 4) - (5\ 1) + (2\ 4)(5\ 1)$$

Hence

$$e_t = c_t^- \{t\} = id_{S_6} \{t\} - (2\ 4) \{t\} - (5\ 1) \{t\} + (2\ 4)(5\ 1) \{t\}$$

Lemma 5.2.9. Let e_t be a polytabloid of a λ -tableau with $\lambda \vdash n$ and $\pi \in S_n$, then $\pi e_t = e_{\pi t}$.

Proof. We first show that $C_{\pi t} = \pi C_t \pi^{-1}$, indeed we may define an equivalence relation on Young tableaux similar to the one in definition 5.2.4 but instead of row equivalence we define it for column equivalence, denote such a class by $[t]$, then

$$\begin{aligned} \sigma \in C_{\pi t} &\Leftrightarrow \sigma[\pi t] = [\pi t] \Leftrightarrow \pi^{-1} \sigma \pi [t] = [t] \Leftrightarrow \pi^{-1} \sigma \pi \in C_t \Leftrightarrow \\ &\Leftrightarrow \sigma \in \pi C_t \pi^{-1} \end{aligned}$$

With this we prove that $c_{\pi t}^- = \pi c_t^- \pi^{-1}$, for this simply note that

$$c_{\pi t}^- = \sum_{\sigma \in C_{\pi t}} \text{sgn}(\sigma) \sigma = \sum_{\pi^{-1} \sigma \pi \in C_t} \text{sgn}(\sigma) \sigma = \sum_{\sigma \in C_t} \text{sgn}(\pi \sigma \pi^{-1}) \pi \sigma \pi^{-1}$$

As $\text{sgn}(\pi \sigma \pi^{-1}) = \text{sgn}(\sigma)$ the conclusion follows. Now with this we have that

$$e_{\pi t} = c_{\pi t}^- \{\pi t\} = \pi c_t^- \pi^{-1} \{\pi t\} = \pi c_t^- \{t\} = \pi e_t$$

□

We are now ready to define the Specht modules.

Definition 5.2.10. Let $\lambda \vdash n$, we define the Specht module S^λ to be the S_n -subspace of M^λ generated by the elements $\{e_t \mid t \text{ is a } \lambda\text{-tableau}\}$.

By the previous lemma the subspace of M^λ generated by $\{e_t \mid t \text{ is a } \lambda\text{-tableau}\}$ is an S_n -subspace of M^λ .

Theorem 5.2.11. The Specht modules S^λ , for $\lambda \vdash n$ are irreducible S_n -spaces, moreover they constitute a full set of non-equivalent irreducible S_n -spaces, that is they pertain to all the non-equivalent irreducible representations of S_n .

Refer to [Ful96, sec 7.2] for a proof of the above theorem.

It is well known that $\{e_t \mid t \text{ is a } \lambda\text{-tableau}\}$ is not a linearly independent set, yet we can find a linearly independent subset of it by choosing appropriate tableaux t .

Definition 5.2.12. We say a Young tableau t is standard if the numbers are in increasing order in each row and column. For $\lambda \vdash n$ we denote the set of all standard Young λ -tableaux by $\text{SYT}(\lambda)$ and let

$$\text{SYT}(n) = \bigcup_{\lambda \vdash n} \text{SYT}(\lambda)$$

For instance the following is a standard Young tableau.

1	3	5
2	4	
6		

A basis for a Specht module S^λ is then $\{e_t \mid t \in \text{SYT}(\lambda)\}$. Hence the dimension of each Specht module S^λ is given by the number of standard λ -tableaux, a well studied combinatorial quantity we'll later come back to.

Note that in a standard Young tableau the number 1 is always in position $(1, 1)$ while n is always in a removable box.

5.2.3 Young graph

It's possible to construct standard Young tableaux for any Ferrers diagram by imposing a partial order on the set of partition $\lambda \vdash n$ for all $n \in \mathbb{N}$ and exploit the graph this partial order naturally produces. Not only do we know partitions are related to irreducible representations but we've also remarked that standard Young tableaux are related to bases of the vector spaces of said irreducible representations, so one can expect the mentioned

graph to be remarkably related to the representation theory of symmetric groups.

We recall the definition of a partial ordered set.

Definition 5.2.13. A partial order on a set X is a binary relation $\leq$, called an order relation, such that

- a) For all $a \in X$, $a \leq a$.
- b) For all $a, b \in X$, if $a \leq b$ and $b \leq a$ then $a = b$.
- c) For all $a, b, c \in X$, if $a \leq b$ and $b \leq c$ then $a \leq c$.

These properties are called reflexivity, antisymmetry and transitivity respectively from first to last. We denote the set X equipped with such relation by $(X, \leq)$ and say X is a partially ordered set, or poset for short. If additionally for every $a, b \in X$ either $a \leq b$ or $b \leq a$ then we say X is a totally ordered set, or a chain.

The choice of the symbol $\leq$ (or other similar) for a order relation is quite flexible as we can also write $b \geq a$ equivalently to $a \leq b$ and write $a < b$ equivalently to $a \leq b$ and $a \neq b$. We now define the mentioned order relation on $\{\lambda \vdash n \mid n \in \mathbb{N}\}$.

Definition 5.2.14. Let $\Lambda = \{\lambda \vdash n \mid n \in \mathbb{N}\}$, the set of all partition of n for all $n \in \mathbb{N}$, we define on it the order relation $\preceq$ defined as, for $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k) \vdash n$ and $\mu = (\mu_1, \mu_2, \dots, \mu_l) \vdash m$

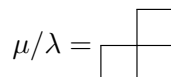
$$\lambda \preceq \mu \Leftrightarrow \forall_{i \in [\max(k, l)]} \lambda_i \leq \mu_i$$

where in the last condition if i is not a valid index for λ or μ we set $\lambda_i = 0$ or $\mu_i = 0$ respectively. Additionally we say μ covers λ if $\lambda \prec \mu$ and if $\lambda \preceq \nu \preceq \mu$ for some $\nu \in \Lambda$ then $\nu = \lambda$ or $\nu = \mu$.

Intuitively the above definition simply means that $\lambda \preceq \mu$ if and only if the Ferrers diagram of λ is contained in the Ferrers diagram of μ . For instance if $\lambda = (3, 2, 1)$ and $\mu = (4, 3, 1)$ then they have respective Ferrers diagrams



With this it becomes clear that $\lambda \preceq \mu$. We can also define a sense of subtraction between Ferrers diagrams λ, μ such that $\lambda \preceq \mu$, denoted μ/λ (and usually called a skew diagram) where we simply remove from μ the respective boxes in λ . For the examples above we have that



Again it becomes clear that μ covers λ if and only if μ/λ is a single box.

As stated a poset has a graph naturally associated with it, called the *Hasse diagram*, which is a directed graph. Given a poset $(X, \leq)$ and $a, b \in X$ we say b covers a if $a < b$ and if there exists $c \in X$ such that $a \leq c \leq b$ then $c = a$ or $c = b$. With this in mind we construct the graph Γ whose vertices are the elements of X and the edges are $E = \{(a, b) \in X \times X \mid a \text{ covers } b\}$

Definition 5.2.15. *The Young (branching) graph is the Hasse diagram of the poset $(\Lambda, \preceq)$ as defined in 5.2.14.*

In figure 5.1 we present the bottom part of the Young graph starting at level $n = 4$ and going down to $n = 1$. Note that because for $n = 1$ there's only one possible partition the whole Young graph "sinks" to the partition $\lambda = (1) \vdash 1$.

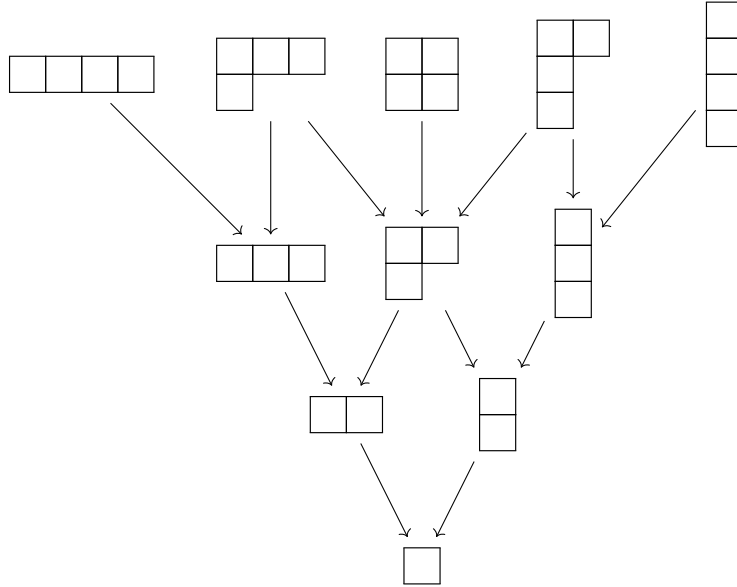


FIGURE 5.1: Bottom part of the Young graph, starting at level $n = 4$ going down to $n = 1$.

Considering the graph in 5.1 the notion of a path in it is immediate.

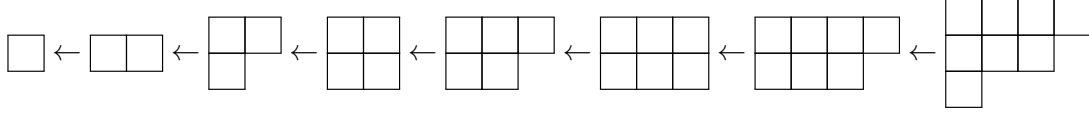
Definition 5.2.16. *A path in the Young graph is a sequence*

$$p = (\lambda^n \rightarrow \lambda^{n-1} \rightarrow \dots \rightarrow \lambda^2 \rightarrow \lambda^1)$$

Where $\lambda^i \in \Lambda$ for all $i \in [n]$, in particular $\lambda^i \vdash i$ and $(\lambda^i, \lambda^{i-1})$ is an edge in the Young graph.

Note that by definition a path always ends in the partition $\lambda = (1) \vdash 1$, for our purposes here there's no need to consider more general paths ending prior to this partition in the graph.

Recall that for an edge in the Young graph $(\lambda^i, \lambda^{i-1})$ it must be the case that λ^i covers λ^{i-1} hence for the Ferrers diagrams $\lambda^i / \lambda^{i-1}$ is composed of a single box, so given a path in the Young graph we can exploit this to inductively construct a standard Young tableau. This is best seen by example. Consider the following path (seen now from bottom to top on the graph)



We start by populating the Ferrers diagram of $\lambda = (1)$ with a 1, and at each step populate the added box with the succeeding natural number, so in the first step we get

$$\boxed{1} \rightsquigarrow \boxed{1 \ 2}$$

In the next step we get

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

And following this procedure we obtain at the end the chain

$$\boxed{1} \rightsquigarrow \boxed{1 \ 2} \rightsquigarrow \begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & \\ \hline \end{array} \rightsquigarrow \begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & 4 \\ \hline \end{array} \rightsquigarrow \begin{array}{|c|c|c|} \hline 1 & 2 & 5 \\ \hline 3 & 4 & \\ \hline \end{array} \rightsquigarrow \begin{array}{|c|c|c|} \hline 1 & 2 & 5 \\ \hline 3 & 4 & 6 \\ \hline \end{array} \rightsquigarrow \begin{array}{|c|c|c|c|} \hline 1 & 2 & 5 & 7 \\ \hline 3 & 4 & 6 & \\ \hline \end{array} \rightsquigarrow \begin{array}{|c|c|c|c|} \hline 1 & 2 & 5 & 7 \\ \hline 3 & 4 & 6 & \\ \hline 8 & & & \\ \hline \end{array}$$

Note that the resulting tableau

1	2	5	7
3	4	6	
8			

is indeed a standard Young tableau, as promised. In fact every tableau in the above chain is standard. It's very easy to prove by induction that any such path in the Young graph produces a standard tableau, indeed the first step has only two options and both produce a standard Young tableau, now assuming by induction that at step k we have a standard Young tableau then at step $k + 1$ we populate a box with the highest number so far and this box is easily seen to always be at a fortuitous position in the diagram. On the other hand we can also see that two distinct paths result in standard tableaux that produce distinct tabloids, indeed if the paths are different then, if say they are distinct at level j of the Young graph, then we place j in different rows on the two tableaux, thus the tabloids they produce must be distinct.

So different paths produce different standard Young tableaux, the question in the back of our minds now is can we in the above way produce all standard Young tableaux for any given partition $\lambda \vdash n$ and $n \in \mathbb{N}$. As it turns out this is indeed the case and we'll be able to

prove this indirectly via the Okounkov-Vershik approach.

Theorem 5.2.17. *There's a bijection between paths in the Young graph and standard Young tableaux, in particular for any $\lambda \vdash n$ and $n \in \mathbb{N}$ there's a bijection between paths leading to the partition λ in the Young graph and standard Young λ -tableaux, that is $\text{SYT}(\lambda)$.*

A direct consequence of the above theorem is that the dimension of a Specht module V^λ is equal to the number of paths leading to λ in the Young graph.

The Young graph gives us a lot more than just standard tableaux, recalling the Specht modules given in the previous section consider an edge (λ, μ) from level n to level $n-1$ in the Young graph, then V^λ gives an irreducible representation of S_n and V^μ an irreducible of S_{n-1} . Now given a path

$$p = (\lambda^n \rightarrow \lambda^{n-1} \rightarrow \dots \rightarrow \lambda^2 \rightarrow \lambda^1)$$

in the Young graph we produce a standard Young tableau t , and the subpath of p

$$p' = (\lambda^{n-1} \rightarrow \dots \rightarrow \lambda^2 \rightarrow \lambda^1)$$

also produces a standard Young tableau t' , moreover t' is a "sub-tableau" of t . With this it's not hard to see that the polytabloid associated with t is very much related to the polytabloid of t' , for instance $c_{t'}^-$ is contained in the sum c_t^- . The remarkable result that follows is called the *Branching rule*.

Theorem 5.2.18. *For $\lambda \vdash n$ consider the Specht module S^λ and consider its restriction to S_{n-1} , $S^\lambda \downarrow_{S_{n-1}}$, then*

$$S^\lambda \downarrow_{S_{n-1}} = \bigoplus_{\lambda \rightarrow \mu} V^\mu$$

That is $S^\lambda \downarrow_{S_{n-1}}$ decomposes into a direct sum of irreducibles of S_{n-1} , V^μ , where (λ, μ) is a edge in the Young graph.

The implications of this theorem are immense and we are not going to discuss them here, we shall arrive at those conclusion more formally in the next section, in the context of the O-V approach. So one doesn't need to delve on it much for now, yet later on it will prove itself useful to properly link the classical approach and the modern approach to the representation theory of symmetric groups.

The Young graph can actually be constructed in a different way considering more directly the irreducible representations/Specht modules S^λ , then called the *Bratelli diagram*. This

gives rise in a more natural way to the Branching rule. Every level of the Young graph is associated to some $n \in \mathbb{N}$, where each vertex in that level is a partition $\lambda \vdash n$. Let's now consider for each $n \in \mathbb{N}$ the symmetric group S_n and its corresponding irreducible representations/Specht modules S^λ . For $n \in \mathbb{N}$ and $\mathbb{N} \ni m \leq n$ there's a natural imbedding of S_m in S_n , just consider the permutations in S_n that only permute the elements in $[m]$ and fix the remaining elements in $[n]$. With this we can consider S_m as a subgroup of S_n and build a chain of groups, using the subgroup (order) relation $\leq$:

$$S_1 \leq S_2 \leq S_3 \leq S_4 \leq \dots$$

A remarkable property of the symmetric groups is that, recalling definition 5.1.3, for each $n \in \mathbb{N}$, S_n is a multiplicity free subgroup of S_{n+1} .^{*} This is of course not at all trivial and will require us substantial work to later prove. But using this fact we can construct a graph in the following way, consider the set

$$\widehat{S} = \bigcup_{n \in \mathbb{N}} \widehat{S}_n$$

This will give the vertices of the graph. For $n \in \mathbb{N}$, $\rho \in \widehat{S}_{n+1}$ and $\sigma \in \widehat{S}_n$ we say ρ covers σ if $\dim_{\mathbb{C}}(\text{Hom}_{S_n}(\sigma, \rho|_{S_n})) = 1$ and then we add the edge (ρ, σ) to the graph. Due to the multiplicity free property mentioned we always know that $\dim_{\mathbb{C}}(\text{Hom}_{S_n}(\sigma, \rho|_{S_n})) \leq 1$ for any $n \in \mathbb{N}$ and $\rho \in \widehat{S}_{n+1}$, $\sigma \in \widehat{S}_n$ so note that ρ and σ are joined by an edge if and only if σ appears in the decomposition of $\rho|_{S_n}$. The Bratelli diagram is then the graph with vertices $\widehat{S}$ and edges

$$\{(\rho, \sigma) \in \widehat{S}_{n+1} \times \widehat{S}_n \mid \dim_{\mathbb{C}}(\text{Hom}_{S_n}(\sigma, \rho|_{S_n})) = 1, n \in \mathbb{N}\}$$

Now if we were to build this graph and compare it to the Young graph we'd conclude that interestingly enough they are the very same graph. The Bratelli diagram will be the starting point in developing the representation theory of symmetric groups via the Okounkov-Vershik approach, once of course we prove that indeed the symmetric groups enjoy the multiplicity free property previously mentioned, this way we'll bypass the need to introduce Young tableaux.

^{*}This can actually be seen as following from the Branching rule, in theorem 5.2.18 above.

5.2.4 Coxeter generators

Up to now we've seen a pretty good description of the vector spaces associated with the irreducible representations of S_n , yet we are still lacking in a good description of exactly how S_n acts on these spaces. To that end we present here some basic theory, that we'll later come back to.

To understand how S_n acts on its respective Specht modules it's best to simplify the problem by selecting a generating set for S_n . One such set can be obtained by simply considering the set of transpositions

$$\{(i \ i+1) \mid i \in [n-1]\}$$

Indeed it is clear that with these one can obtain any permutation.

Definition 5.2.19. *The set*

$$\{s_i = (i \ i+1) \mid i \in [n-1]\}$$

generates S_n , its elements are called the Coxeter generators of S_n . Moreover S_n is uniquely determined by the following relations between Coxeter generators, called the Coxeter relations

$$s_i^2 = 1; \quad s_i s_j s_i = s_j s_i s_j \text{ if } |i - j| = 1; \quad s_i s_j = s_j s_i \text{ if } |i - j| \geq 2$$

for $i, j \in [n-1]$.

Definition 5.2.20. *Given $\pi \in S_n \setminus \{1_{S_n}\}$ such that $\pi = s_{i_1} s_{i_2} \dots s_{i_k}$ we say $s_{i_1} s_{i_2} \dots s_{i_k}$ is reduced if there is no shorter product of Coxeter generator that produce π , and in this case we say that k is the Coxeter length of π . Additionally we convention the Coxeter length of 1_{S_n} to be 0.*

We've seen before, in 5.2, how S_n acts on tableaux, and not only that but we've also seen in lemma 5.2.9 how S_n acts on a particular basis of Specht modules. By that lemma these two actions of S_n are intrinsically related, and we may exploit it to our advantage. For instance one may wonder when a Coxeter generator simply maps a standard tableau into another standard tableau, thus meaning that it maps an element of the basis of a Specht module into another. This line of thought will be fundamental in understanding how S_n acts on its irreducible spaces, or in other words in fully characterizing the irreducible representations of S_n . For now we present a few useful results about Coxeter generators.

Definition 5.2.21. *Let t be a standard Young tableau and s_i a Coxeter generator. We say s_i is admissible for t if $s_i t$ is a standard Young tableau.*

Note that if s_i is admissible for t then i and $i + 1$ are neither in the same row nor the same column of t .

Definition 5.2.22. Let $\pi \in S_n$, an inversion in π is a pair (i, j) with $i, j \in [n]$ such that $i < j$ but $\pi(i) > \pi(j)$. We define $\mathcal{I}(\pi)$ to be the set of all inversions in π , that is

$$\mathcal{I}(\pi) = \{(i, j) \in [n] \times [n] \mid i < j \text{ and } \pi(i) > \pi(j)\}$$

Additionally we let $l(\pi) = |\mathcal{I}(\pi)|$, the number of inversions in π .

Remarkably we can characterize the Coxeter length of a permutation in terms of inversions.

Theorem 5.2.23. For $\pi \in S_n$, the Coxeter length of π equals $l(\pi)$.

Proof. Let $\pi \in S_n$ such that $\pi = s_{i_1} \cdots s_{i_k}$ with k the Coxeter length of π and consider $l(\pi)$. Let $i \in [n-1]$ and consider πs_i , then $l(\pi s_i) = l(\pi) + 1$ if $\pi(i) < \pi(i+1)$ and $l(\pi s_i) = l(\pi) - 1$ if $\pi(i) > \pi(i+1)$, hence $l(\pi) \leq l(\pi s_i) + 1$ and so, noting $s_i s_i = 1_{S_n}$ we have that

$$l(\pi) = l(s_{i_1} \cdots s_{i_k}) \leq l(s_{i_1} \cdots s_{i_{k-1}}) + 1 \leq l(s_{i_1} \cdots s_{i_{k-2}}) + 2 \leq \cdots \leq k$$

Therefore $l(\pi)$ is bounded above by the Coxeter length, we now prove that the Coxeter length is bounded above by $l(\pi)$, thus proving the Coxeter length of π equals $l(\pi)$.

We note that if $l(\pi) = 0$ then $\pi = 1_{S_n}$ thus the Coxeter length of π is 0, so equal to $l(\pi)$. With this using induction on $l(\pi)$, assuming then that for $l(\pi) \leq n$ the Coxeter length of π at most $l(\pi)$, let $l(\pi) = n + 1$ and k the Coxeter length of π . As then π is not the identity there must exist $i \in [n]$ such that $\pi(i) > \pi(i+1)$, indeed if for all $i \in [n]$, $\pi(i) < \pi(i+1)$ then $\pi(1) < \pi(2) < \pi(3) < \cdots < \pi(n)$ which is only possible if π is the identity. Hence by what we've seen above $l(\pi s_i) = l(\pi) - 1$, and so by the induction hypothesis the Coxeter length of πs_i is at most $l(\pi s_i)$, but again by the above we also have that the Coxeter length πs_i is at least $l(\pi s_i)$, hence the Coxeter length of πs_i is precisely $l(\pi s_i) = l(\pi) - 1$. Now simply note that $\pi = \pi s_i s_i$ hence π can be written as a product of $l(\pi)$ Coxeter generators, hence $k \leq l(\pi)$. □

With this we can now turn our attention back to standard tableaux, we start by defining a particular one of easy construction for all $\lambda \vdash n$.

Definition 5.2.24. For $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k) \vdash n$ we define the canonical λ -tableau, denoted t_s^λ or simply t_s when λ is clear from context, to be

$$\begin{array}{c}
 \mathbf{t}_s = \begin{array}{|c|c|c|} \hline 1 & 2 & 3 \\ \hline \lambda_1 + 1 & \lambda_1 + 2 & \lambda_1 + 2 \\ \hline \end{array} \quad \dots \quad \begin{array}{|c|c|} \hline \lambda_1 - 1 & \lambda_1 \\ \hline \lambda_1 + \lambda_2 & \\ \hline \end{array} \\
 \\
 \begin{array}{ccc} \vdots & \vdots & \vdots \end{array} \\
 \\
 \begin{array}{|c|} \hline \sum_{i=1}^{k-1} \lambda_i + 1 \\ \hline \end{array} \quad \dots \quad \begin{array}{|c|} \hline n \\ \hline \end{array}
 \end{array}$$

Proposition 5.2.25. *Given $t \in \text{SYT}(\lambda)$ with $\lambda \vdash n$ let σ_t be the (unique) permutation such that $\sigma_t t = \mathbf{t}_s$ and consider its Coxeter length, $l = l(\sigma_t)$. Then σ_t decomposes as a product of transpositions $s_{i_1} s_{i_2} \cdots s_{i_l}$, where s_{i_j} is admissible for $s_{i_{j+1}} s_{i_{j+2}} \cdots s_{i_l} t$.*

Proof. We prove this by induction on n . If $n = 1$ then $\lambda = (1)$ and the tableau t is trivially standard and canonical already, so $\sigma_t = 1_{S_1}$ and the proposition follows trivially. Assume now that the proposition is valid for all $n \leq k$, let then $n = k + 1$ and let j be the number in the rightmost box of the last row of t . There are two possibilities, either $j = n$ or $j \neq n$.

If $j = n$ then as its respective box is removable we may do so, producing then a standard tableau t' of $\mu \vdash n - 1$, by the induction hypothesis there exists a sequence of $l' = l(\sigma_{t'})$ admissible transposition mapping t' to $\mathbf{t}_s^\mu$. Well the same sequence will also map t to $\mathbf{t}_s$ and by the uniqueness of σ_t we have that that sequence equals σ_t , hence $l' = l$.

If $j \neq n$ then note that s_j is admissible for t . Similarly s_{j+1} is admissible for $s_j t$, and proceeding this way we get that s_{n-1} is admissible for $s_{n-2} s_{n-3} \cdots s_j t$. Now letting $t' = s_{n-2} s_{n-3} \cdots s_j t$ we get that for t' , $j = n$ and so we fall into the previous case. $\square$

Now if for any $t \in \text{SYT}(\lambda)$ we have a sequence of admissible transpositions mapping t to $\mathbf{t}_s$ then using the inverse of that sequence we have a sequence of admissible transpositions mapping $\mathbf{t}_s$ to t , thus we have the following result.

Corollary 5.2.26. *Given $t_1, t_2 \in \text{SYT}(\lambda)$ there exist a sequence of admissible transpositions mapping t_1 to t_2 , and vice versa.*

By the above corollary we may define an equivalence relation on $\text{SYT}(n)$ as follows.

Definition 5.2.27. We define the equivalence relation $\approx$ on $\text{SYT}(n)$ by, given $t_1, t_2 \in \text{SYT}(n)$

$$t_1 \approx t_2 \Leftrightarrow \text{there exist a sequence of admissible transpositions mapping } t_1 \text{ to } t_2$$

As the action of S_n on a λ -tableau, with $\lambda \vdash n$, can only map it to another λ -tableau we have by the above corollary that $|\text{SYT}(n)/\approx| = |\{\lambda \vdash n\}|$, and as we know $|\{\lambda \vdash n\}| = |\widehat{S}_n|$, hence $|\text{SYT}(n)/\approx| = |\widehat{S}_n|$.

5.3. The O-V approach

In this section we present the more general components of the Okounkov-Vershik approach, before restricting to the case of symmetric groups. As remarked at the end of Subsection 5.2.3 this is done by finding an alternative way of defining the Young graph, one that doesn't rely on tableaux but on irreducible representations of the groups in a chain of groups, that will be called the branching graph. This will allow us to define things in a more general way, as intended.

Using this graph we'll be able to find a basis for any irreducible representation of any group in the chain with very nice properties, called the GZ-basis. Next we'll focus on a group in the chain and wonder what elements in its respective group algebra act diagonally in the GZ-bases, the resulting subalgebra of the group algebra, called the GZ-algebra, will allow us to uniquely determine any element in the GZ-basis up to a scalar factor, and not only that but by finding a suitable set of generator for the GZ-algebra we'll be able to talk about irreducible representations in terms of eigenvalues of elements in that generating set. This will ultimately be used, in the case of symmetric group, to establish the relation between standard Young tableaux and bases of irreducible representations.

The last step in this section will be to find an alternative to theorem 5.1.5 of determining when a subgroup is multiplicity free, with hopes that ultimately we may use it to prove that any group in the chain of groups is a multiplicity free subgroup of the one succeeding it. This will be fundamental to the theory as it will give the branching graph a manageable structure.

5.3.1 The Gelfand-Tsetlin basis

As stated we start off by finding a basis for the irreducible representations of symmetric groups, or better said where that basis can come from provided ideal conditions. To that end we first formally define the Bratelli diagram of a chain of groups in greater generality.

Definition 5.3.1. Consider the (subgroup) order relation $\leq$ on the set of all finite groups and let

$$G_1 \leq G_2 \leq G_3 \leq G_4 \leq \dots$$

be a chain of finite groups. We say the above chain is multiplicity free if G_n is a multiplicity free subgroup of G_{n+1} for all indices $n \geq 1$ in the chain.

Definition 5.3.2. Consider a chain of finite groups

$$G_1 = \{1\} \leq G_2 \leq G_3 \leq G_4 \leq \dots$$

The branching multigraph* of this chain is the graph with vertices the set

$$\widehat{G} = \bigcup_{n \geq 1} \widehat{G}_n$$

where as always $\widehat{G}_n$ is the set of equivalence classes of irreducible representation of G_n . Two vertices $\rho \in \widehat{G}_{n+1}$, $\sigma \in \widehat{G}_n$ are joined by k directed edges, from ρ to σ , if

$$k = \dim_{\mathbb{C}}(\text{Hom}_{G_n}(\sigma, \rho|_{G_n}))$$

That is the multiplicity of σ in the decomposition of $\rho|_{G_n}$ is equal to k . If the chain is multiplicity free then we obtain a graph and not a multigraph, that is there are no multiple edges between vertices, in this case one usually refers to it as the branching graph/Bratelli diagram of the chain and says that the branching is simple.

Note that if $G_1 = \{1\}$, that is G_1 is the trivial group, then $\widehat{G}_1$ contains only the trivial representation ι , thus the branching multigraph "sinks" to a single irreducible representation, the trivial one.

Example 5.3.3. Consider the chain of symmetric groups mentioned in the previous section

$$S_1 \leq S_2 \leq S_3 \leq S_4 \leq \dots$$

We can easily obtain the character table of S_2 from example 4.1.36 and we've seen the character table of S_3 in example 4.3.36 and the character table of S_4 in example 4.5.7.

*A multigraph is a graph permitted to have multiple edges between any two vertices.

With this using character theory we can easily produce the branching graph of the chain $S_1 \leq S_2 \leq S_3 \leq S_4$, given in figure 5.2.

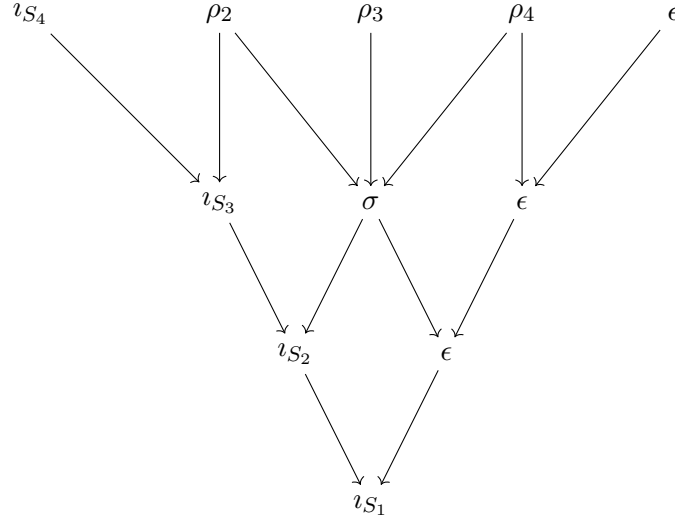


FIGURE 5.2: Branching graph of the chain $S_1 \leq S_2 \leq S_3 \leq S_4$.

Comparing figure 5.2 to figure 5.1 we can see that this branching graph coincides with the first 4 levels of the Young graph, which as we've stated before and shall later see, is no coincidence.

From now on we focus on the case where the chain considered is multiplicity free, or in other words the branching is simple. The theory next presented easily generalizes to the multigraph case.

Definition 5.3.4. Given a chain of groups $G_1 \leq G_2 \leq G_3 \leq \dots$ and its respective branching (multi)graph, if two representations $\rho \in \widehat{G}_{n+1}$, $\sigma \in \widehat{G}_n$, for some $n \in \mathbb{N}$, are joined by an edge, that is (ρ, σ) is an edge in the graph then we denote this relation between them by $\rho \rightarrow \sigma$. Additionally if $\rho \in \widehat{G}_n$, $\sigma \in \widehat{G}_k$ for some $k < n$ and there exists a path in the branching graph from ρ to σ we denote this relation between them by $\rho \searrow \sigma$.

Definition 5.3.5. Let $G_1 = \{1\} \leq G_2 \leq G_3 \leq \dots$ be a (multiplicity free) chain of groups and consider its branching graph, with vertices

$$\widehat{G} = \bigcup_{n \geq 1} \widehat{G}_n = \{\rho \in \widehat{G}_n \mid n \geq 1\}$$

We define $\mathfrak{P}$ to be the set of all possible paths $\mathfrak{p} = (\rho^n \rightarrow \rho^{n-1} \rightarrow \dots \rightarrow \rho^2 \rightarrow \iota)$ in the branching graph, for $n \geq 1$ (with $\rho^n \in \widehat{G}_n$). The set of all paths that start at a particular irreducible representation ρ

is denoted by $\mathfrak{P}^\rho$, and the set of all paths that start at a particular irreducible representation ρ and end at a particular irreducible representation σ below, also called truncated paths, is denoted by $\mathfrak{P}_\sigma^\rho$.

We now note the following, each level of the branching graph contains every irreducible representation of the respective group, thus the branching there, that is the edges, give the decomposition into irreducibles of each irreducible representation when restricted to the group on the level below. As restricting a representation from G_n to G_{n-1} gives a representation of G_{n-1} this means that the branching gives the decomposition of a restricted irreducible representation of G_n into irreducibles of G_{n-1} . Taking figure 5.2 as an example the representation $\rho_2 \downarrow_{S_3}$ decomposes as

$$\rho_2 \downarrow_{S_3} = \iota_{S_3} \oplus \sigma$$

Hence if V_{ρ_2} is the S_4 -space associated with ρ_2 and V_1, V_2 are the S_3 -spaces associated with ι_{S_3} and σ respectively then

$$V_{\rho_2} \downarrow_{S_3} \cong V_1 \oplus V_2$$

More generally if we consider a (multiplicity free) chain $G_1 = \{1\} \leq G_2 \leq G_3 \leq \dots$ then for $\rho_V \in \widehat{G}_{n+1}$

$$V \downarrow_{G_n} \cong \bigoplus_{\substack{\sigma_{V_i} \in \widehat{G}_n, \\ \rho \rightarrow \sigma_{V_i}}} V_i$$

In particular if we look at the spaces above as just vector spaces then given a basis for each V_i we get a basis for V .^{*} To that end let's not stop here, we can repeat the above process for each V_i , and repeatedly iterate until we can no longer. Doing that then produces

$$\begin{aligned} V \cong \bigoplus_{\substack{\sigma_{V_i} \in \widehat{G}_n, \\ \rho \rightarrow \sigma_{V_i}}} V_i &\cong \bigoplus_{\substack{\sigma_{V_i} \in \widehat{G}_n, \\ \rho \rightarrow \sigma_{V_i}}} \bigoplus_{\substack{\sigma_{W_j} \in \widehat{G}_{n-1}, \\ \sigma_{V_i} \rightarrow \sigma_{W_j}}} W_j \cong \bigoplus_{\substack{\sigma_{V_i} \in \widehat{G}_n, \\ \rho \rightarrow \sigma_{V_i}}} \bigoplus_{\substack{\sigma_{W_j} \in \widehat{G}_{n-1}, \\ \sigma_{V_i} \rightarrow \sigma_{W_j}}} \bigoplus_{\substack{\sigma_{U_l} \in \widehat{G}_{n-2}, \\ \sigma_{W_j} \rightarrow \sigma_{U_l}}} U_l \cong \dots \cong \\ &\cong \bigoplus_{p \in \mathfrak{P}^\rho} I \end{aligned} \tag{5.4}$$

Where I is the vector space associated with the trivial representation ι , that as we know is a one dimensional space. With this we immediately know the following.

^{*}Compare this to the Branching rule in theorem 5.2.18, we will recover the Branching rule from this once we prove the isomorphism between the branching graph (for symmetric groups) and the Young graph.

Corollary 5.3.6. *Let $\rho_V \in \widehat{G}_n$ with G_n a member of the multiplicity free chain $G_1 = \{1\} \leq G_2 \leq G_3 \leq \dots$, then considering the associated branching graph we have that*

$$\dim_{\mathbb{C}}(V) = |\mathfrak{P}^\rho|$$

Moreover for $k \in [n - 1]$ and $\sigma_W \in \widehat{G}_k$ the multiplicity of W in $V \downarrow_{S_k}$ is equal to $|\mathfrak{P}_\sigma^\rho|$.

The above procedure also gives a basis for the G_n -space V , indeed we can at each step in the above process force the direct sum to be an inner direct sum. By first finding the distinct subspaces of V that are isomorphic to the V_i for each i , and then the ones from the level below that are isomorphic to respective subspaces of V_i , for each i , and so on. This is done in a manner that preserves invariance for the action of the respective group in that level, giving this basis very nice properties. Thus for correctness we'd write the decomposition in 5.4 as

$$V = \bigoplus_{\mathfrak{p} \in \mathfrak{P}^\rho} I_{\mathfrak{p}} \tag{5.5}$$

with $I_{\mathfrak{p}}$, for $\mathfrak{p} \in \mathfrak{P}^\rho$, orthogonal one dimensional subspaces of V . Hence the basis of V described comes from simply choosing a vector $v_{\mathfrak{p}}$ in each one dimensional space $I_{\mathfrak{p}}$.

Definition 5.3.7. *Given a multiplicity free chain $G_1 = \{1\} \leq G_2 \leq G_3 \leq \dots$ and an irreducible G_n -space V associated with an irreducible representation ρ , the basis produced by the decomposition in 5.5 above is called the Gelfand-Tsetlin basis of V , or GZ-basis of V for short. We denote each vector of this basis by $v_{\mathfrak{p}}$ for $\mathfrak{p} \in \mathfrak{P}^\rho$. A vector in a GZ-basis of V , for any irreducible G_n -space V , is called a GZ-vector of G_n .*

Considering the above, the space V decomposes as

$$V = \bigoplus_{\mathfrak{p} \in \mathfrak{P}^\rho} \langle v_{\mathfrak{p}} \rangle$$

where $\langle v_{\mathfrak{p}} \rangle$ is the vector space over $\mathbb{C}$ generated by the Gelfand-Tsetlin basis element $v_{\mathfrak{p}}$, that is $\langle v_{\mathfrak{p}} \rangle = \mathbb{C}v_{\mathfrak{p}}$. Note this is not a decomposition into G_n -subspaces as the spaces $\mathbb{C}v_{\mathfrak{p}}$ are not necessarily G_n -invariant, this is simply a decomposition of the vector space V into one dimensional subspaces, yet the indexation of the basis elements $v_{\mathfrak{p}}$ and the way they are chosen gives us a lot of power, for instance consider $v_{\mathfrak{p}}$ with $\mathfrak{p} = (\rho_{V_n}^n \rightarrow \rho_{V_{n-1}}^{n-1} \rightarrow \dots \rightarrow \rho_{V_2}^2 \rightarrow \iota_{V_1})$ (note that $V_1 = I_{\mathfrak{p}}$). As each G_i -space V_i is irreducible/simple it is generated by any non-zero element of V_i (recall proposition D.1.9), now note that

$v_p \in V_i$ for all $i \in [n]$, hence if $v_p \neq 0$ then

$$V_i = \mathbb{C}G_i v_p$$

for all $i \in [n]$.

Remark 4. It's also possible to define the branching multigraph with a bottom up approach, that is instead of considering restricted representations we can consider induced representations, this follows from corollary 4.6.9 (Nakayama's reciprocity theorem). With this approach the edges point upwards from a level to the one above, meaning that a particular irreducible representation $\sigma \in \widehat{G}_{n-1}$ is connected to an irreducible representation $\rho \in \widehat{G}_n$ by k edges if $k = \dim_{\mathbb{C}}(\text{Hom}_{G_n}(\sigma \uparrow^{G_n}, \rho))$, thus producing k (directed) edges (σ, ρ) . As by corollary 4.6.9 $\dim_{\mathbb{C}}(\text{Hom}_{G_n}(\sigma \uparrow^{G_n}, \rho)) = \dim_{\mathbb{C}}(\text{Hom}_{G_{n-1}}(\sigma, \rho \downarrow_{G_{n-1}}))$ the resulting multigraph is the same as the one previously defined with the only difference being that now edges point in the opposite direction.

5.3.2 The Gelfand-Tsetlin algebra

Recall the fundamental isomorphism in corollary 4.4.3

$$\mathbb{C}G \cong \bigoplus_{\rho_{V_i} \in \widehat{G}} \text{End}(V_i) \quad (5.6)$$

Assuming G is part of a multiplicity free chain of groups, say $G = G_n$ for some $n \geq 1$, the previous subsection gives us a way to produce a basis for each irreducible G -space V_i with very nice properties, namely the GZ-basis of V_i . With it we can write any element of $\text{End}(V_i)$ as a square matrix and consider the more concrete isomorphism

$$\mathbb{C}G \cong \bigoplus_{\rho_{V_i} \in \widehat{G}} M_{n_i}(\mathbb{C}) \quad (5.7)$$

Where $n_i = \dim_{\mathbb{C}}(V_i)$ for each $\rho_{V_i} \in \widehat{G}$. Considering the algebra of matrices $M_{n_i}(\mathbb{C})$ (with the natural operations), it is well known that the subset of diagonal matrices $D_{n_i}(\mathbb{C}) \subset M_{n_i}(\mathbb{C})$ is a maximal commutative subalgebra of it, indeed it is clearly commutative and one may prove that if $A, B \in M_{n_i}(\mathbb{C})$ with A diagonal with pairwise distinct diagonal entries and $AB = BA$ then B is diagonal, hence $D_{n_i}(\mathbb{C})$ is maximal.

Considering we're mapping $\text{End}(V_i) \rightarrow M_{n_i}(\mathbb{C})$ via the GZ-basis of V_i , call that map Φ , this means that $D_{n_i}(\mathbb{C})$ corresponds to maps in $\text{End}(V_i)$ that act diagonally in the GZ-basis, or

in other words for each map in $\Phi^{-1}(D_{n_i}(\mathbb{C}))$ all its eigenspaces are spanned by vectors in the GZ-basis of V_i .

So first we can conclude that $\bigoplus_{\rho_{V_i} \in \widehat{G}} D_{n_i}(\mathbb{C})$ is a maximal commutative subalgebra of $\bigoplus_{\rho_{V_i} \in \widehat{G}} M_{n_i}(\mathbb{C})$ and given an isomorphism Ψ for 5.7 then $\Psi^{-1}(\bigoplus_{\rho_{V_i} \in \widehat{G}} D_{n_i}(\mathbb{C}))$ is a maximal commutative subalgebra of $\mathbb{C}G$, moreover the isomorphism Ψ induces in a natural way an action of $\mathbb{C}G$ on $\bigoplus_{\rho_{V_i} \in \widehat{G}} V_i$, hence

$$\mathfrak{D} = \Psi^{-1} \left(\bigoplus_{\rho_{V_i} \in \widehat{G}} D_{n_i}(\mathbb{C}) \right) \quad (5.8)$$

is a maximal commutative subalgebra of $\mathbb{C}G$, composed of the elements that act diagonally on all GZ-bases, that is the GZ-basis of V_i for all $\rho_{V_i} \in \widehat{G}$, thus for any element in $\mathfrak{D}$ any vector in any of the GZ-basis is an eigenvector of it, or in other words any vector in any of the GZ-basis is a common eigenvector of all elements in $\mathfrak{D}$.

Couple interesting things to note about the subalgebra $\mathfrak{D}$, first we can easily determine its dimension, as Ψ is an isomorphism then

$$\dim_{\mathbb{C}}(\mathfrak{D}) = \sum_{\rho_{V_i} \in \widehat{G}} \dim_{\mathbb{C}}(D_{n_i}(\mathbb{C})) = \sum_{\rho_{V_i} \in \widehat{G}} n_i = \sum_{\rho_{V_i} \in \widehat{G}} \dim_{\mathbb{C}}(V_i)$$

Hence

$$\mathfrak{D} \cong \bigoplus_{\rho_{V_i} \in \widehat{G}} V_i$$

as vector spaces. Considering this any vector in any of the GZ-bases is uniquely determined, up to a factor, by its respective eigenvalues on each element of $\mathfrak{D}$. Indeed consider a vector v_p in the GZ-basis of $\rho_{V_k} \in \widehat{G}$, if we consider the map

$$\bigoplus_{\rho_{V_i} \in \widehat{G}} D_i \in \bigoplus_{\rho_{V_i} \in \widehat{G}} D_{n_i}(\mathbb{C}) \quad (5.9)$$

such that D_i is the zero map for all $i \neq k$ and D_k has entries all zero except in a position in its diagonal corresponding to the vector v_p , where it has a 1, then

$$f = \Psi^{-1} \left(\bigoplus_{\rho_{V_i} \in \widehat{G}} D_i \right)$$

is an element of $\mathfrak{D}$ with eigenvalues 1 and 0 such that $\langle v_p \rangle$ is the eigenspace associated to the eigenvalue 1 and all other elements in all GZ-bases span the eigenspace of

0, hence we can use f to determine v_p up to a scalar factor by knowing that it's associated with the eigenvalue 1. We'll later focus on $G = S_n$ and use this fact to establish a bijection between eigenvalues of particular elements of $\mathfrak{D}$ and standard Young tableaux of partitions of n , that is $\text{SYT}(n)$, justifying the use of Young tableaux in the representation theory of symmetric groups.

As we now proceed to show there is an alternative way of defining the algebra $\mathfrak{D}$, one that allows us to inductively define $\mathfrak{D}$ for any group G_n in the chain of groups $G_1 \leq G_2 \leq G_3 \leq \dots$, this will be crucial to the inductive reasoning used in the O-V approach.

Definition 5.3.8. Let $G_1 \leq G_2 \leq G_3 \leq \dots$ be a chain of groups and consider for each $k \geq 1$, $Z_k = Z(\mathbb{C}G_k)$, the center of the group algebra $\mathbb{C}G_k$. We define the Gelfand-Tsetlin algebra of order n of the aforementioned chain of groups, denoted GZ_n , to be the algebra generated by the union of the centers Z_i for $i \in [n]$, that is

$$\text{GZ}_n = \left\langle \bigcup_{i \in [n]} Z_i \right\rangle$$

Theorem 5.3.9. Let $G_1 = \{1_G\} \leq G_2 \leq G_3 \leq \dots$ be a multiplicity free chain of groups, then $\text{GZ}_n = \mathfrak{D}$ where $\mathfrak{D}$ is the subalgebra of $\mathbb{C}G_n$ defined previously.

Proof. We first prove that $\text{GZ}_n \subseteq \mathfrak{D}$. Let $f \in Z_k$ for some $k \in [n]$ and v_p a GZ-vector, as f commutes with any element in G_k it correspond to a G_k -linear map, moreover v_p is a vector in some irreducible G_k -space V_k and f must act invariantly on this space being it an element of $\mathbb{C}G_k$, thus by part b) of Schur's lemma f correspond to a map αid_{V_k} for some $\alpha \in \mathbb{C}$, hence f acts diagonally on v_p . As v_p is any GZ-vector, f is any in Z_k and k is any in $[n]$ we have that $Z_k \subseteq \mathfrak{D}$ for all $k \in [n]$, so consequently $\text{GZ}_n \subseteq \mathfrak{D}$.

We now prove that GZ_n includes $\mathfrak{D}$. Consider a path

$$p = (\rho_V^n \rightarrow \rho_{V_{n-1}}^{n-1} \rightarrow \dots \rightarrow \rho_{V_2}^2 \rightarrow \iota_{V_1})$$

in the branching graph of the chain. As we know, for each $k \leq n$

$$\mathbb{C}G_k = \bigoplus_{\rho_{V_i^k} \in \widehat{G}_k} \text{End}(V_i^k)$$

It is immediate that each $\text{End}(V_i^k)$ corresponds to a two sided ideal of $\mathbb{C}G_k$, hence $\mathbb{C}G_k$ decomposes as a direct sum of two sided ideal and so by part a) and b) of proposition D.1.15 there exist a set $\{e_{V_i^k}\}_{i \in [\widehat{G}_k]}$ of central orthogonal idempotent elements in $\mathbb{C}G_k$ such that $\mathbb{C}G_k e_{V_i^k} \cong \text{End}(V_i^k)$. Now for some $i \in [\widehat{G}_k]$, $V_i^k = V_k$ as in the path p , so

consider the particular central idempotents e_{V_k} , for each $k \in [n]$. Note that for $k \in [n]$, $e_{V_k} \in Z_k$, with this in mind consider the following element of GZ_n

$$e_p = e_{V_1} e_{V_2} e_{V_3} \cdots e_{V_n}$$

By the elements e_{V_k} being idempotent and $\mathbb{C}G_k e_{V_i^k} \cong \text{End}(V_i^k)$ it is immediate that these elements correspond to projections

$$P_{V_k} : \bigoplus_{\rho_{V_i^k} \in \widehat{G}_k} \text{End}(V_i^k) \rightarrow \text{End}(V^k)$$

Hence e_p corresponds to a map that projects onto V_1 (which note it is equal to I_p), that is to say that $e_p \rightarrow P_p$ such that, letting v_p be the GZ-basis element corresponding to the path p

$$P_p v_p = v_p$$

and for any other vector v' of a GZ-basis, $P_p v' = 0$. But then for each $\rho \in \widehat{G}_n$ and path $p \in \mathfrak{P}^\rho$ the elements $e_p \in \text{GZ}_n$ correspond precisely to the elements described in 5.9, which correspond to a basis of $\mathfrak{D}$, therefore $\mathfrak{D} \subseteq \text{GZ}_n$.^{*} $\square$

By our previous remarks about the algebra $\mathfrak{D}$ we have the following.

Corollary 5.3.10. *For a multiplicity free chain of groups $G_1 = \{1_G\} \leq G_2 \leq G_3 \leq \cdots$*

- a) $\dim_{\mathbb{C}}(\text{GZ}_n) = \sum_{\rho_{V_i} \in \widehat{G}_n} \dim_{\mathbb{C}}(V_i)$
- b) *Any vector in any of the GZ-bases is uniquely determined, up to a factor, by its respective eigenvalues on each element in GZ_n . In particular if two GZ-vectors u, v have the same eigenvalues on every element of GZ_n then $u = v$.*
- c) *For $\rho_V \in \widehat{G}_n$ and $0 \neq v \in V$, if v is an eigenvector for every elements of GZ_n then, up to a scalar factor, v is an elements of the GZ-basis of V .*

Now that we better know who the algebra $\mathfrak{D}$ is, namely the GZ-algebra, we are better prepared to work with it. The proof of the following result is also immediate by the proof of theorem 5.3.9.

Corollary 5.3.11. *For a multiplicity free chain of groups $G_1 = \{1_G\} \leq G_2 \leq G_3 \leq \cdots$ the set of central idempotents of each algebra $\mathbb{C}G_k$, for $k \leq n$ is a set of generators for the GZ-algebra of order n .*

^{*}Note that, although we went in the opposite direction, what was essentially done was finding the inverse image of the canonical basis of $\bigoplus_{\rho_{V_i} \in \widehat{G}} D_{n_i}(\mathbb{C})$, as in 5.9, and use some cunningness to determine what that algebra is.

It's possible to find a condition for a subset of elements of a GZ-algebra to generate it.

Definition 5.3.12. Let $\{f_i\}_{i \in [k]} \subset \text{GZ}_n$, as these elements act diagonally in the GZ-bases consider for $\rho_V \in \widehat{G}_n$ and $\mathfrak{p} \in \mathfrak{P}^\rho$ the GZ-basis of V , $\{v_{\mathfrak{p}}^\rho\}_{\mathfrak{p} \in \mathfrak{P}^\rho}$ and the corresponding eigenvalues in $f_i v_{\mathfrak{p}}^\rho = \lambda_{i,\rho,\mathfrak{p}} v_{\mathfrak{p}}^\rho$. Additionally consider the map

$$\Upsilon : (\rho, \mathfrak{p}) \mapsto (\lambda_{1,\rho,\mathfrak{p}}, \lambda_{2,\rho,\mathfrak{p}}, \dots, \lambda_{k,\rho,\mathfrak{p}})$$

If the map Υ is injective we say the set $\{f_i\}_{i \in [k]}$ separates the GZ-bases, in the sense that any vector $v_{\mathfrak{p}}^\rho$ corresponds to a unique k -tuple of eigenvalues $(\lambda_{1,\rho,\mathfrak{p}}, \lambda_{2,\rho,\mathfrak{p}}, \dots, \lambda_{k,\rho,\mathfrak{p}})$.

Proposition 5.3.13. A set $\{f_i\}_{i \in [k]} \subset \text{GZ}_n$ separates the GZ-bases if and only if $\{f_i\}_{i \in [k]} \cup \{1_{G_n}\}$ generates GZ_n .

Proof. For each $(\rho, \mathfrak{p})$ define

$$f_{\rho,\mathfrak{p}} = \prod_{\substack{(\sigma,\mathfrak{t}) \neq (\rho,\mathfrak{p}), \\ i \in [k], \\ \lambda_{i,\rho,\mathfrak{p}} \neq \lambda_{i,\sigma,\mathfrak{t}}}} \frac{f_i - \lambda_{i,\sigma,\mathfrak{t}} 1_{G_n}}{\lambda_{i,\rho,\mathfrak{p}} - \lambda_{i,\sigma,\mathfrak{t}}}$$

As the set $\{f_i\}_{i \in [k]}$ separates the GZ-bases then $f_{\rho,\mathfrak{p}}$ is non zero as there must be some $i \in [k]$ for which $\lambda_{i,\rho,\mathfrak{p}} \neq \lambda_{i,\sigma,\mathfrak{t}}$, for any $(\sigma, \mathfrak{t}) \neq (\rho, \mathfrak{p})$. Now note that, as

$$\frac{f_i - \lambda_{i,\sigma,\mathfrak{t}} 1_{G_n}}{\lambda_{i,\rho,\mathfrak{p}} - \lambda_{i,\sigma,\mathfrak{t}}} v_{\mathfrak{p}}^\rho = \frac{f_i v_{\mathfrak{p}}^\rho - \lambda_{i,\sigma,\mathfrak{t}} v_{\mathfrak{p}}^\rho}{\lambda_{i,\rho,\mathfrak{p}} - \lambda_{i,\sigma,\mathfrak{t}}} = \frac{\lambda_{i,\rho,\mathfrak{p}} v_{\mathfrak{p}}^\rho - \lambda_{i,\sigma,\mathfrak{t}} v_{\mathfrak{p}}^\rho}{\lambda_{i,\rho,\mathfrak{p}} - \lambda_{i,\sigma,\mathfrak{t}}} = v_{\mathfrak{p}}^\rho$$

then $f_{\rho,\mathfrak{p}} v_{\mathfrak{p}}^\rho = v_{\mathfrak{p}}^\rho$. Moreover for $v_{\mathfrak{k}}^\eta$ with $(\eta, \mathfrak{k}) \neq (\rho, \mathfrak{p})$ and any $i \in [k]$

$$\frac{f_i - \lambda_{i,\eta,\mathfrak{k}} 1_{G_n}}{\lambda_{i,\rho,\mathfrak{p}} - \lambda_{i,\eta,\mathfrak{k}}} v_{\mathfrak{k}}^\eta = \frac{f_i v_{\mathfrak{k}}^\eta - \lambda_{i,\eta,\mathfrak{k}} v_{\mathfrak{k}}^\eta}{\lambda_{i,\rho,\mathfrak{p}} - \lambda_{i,\eta,\mathfrak{k}}} = \frac{\lambda_{i,\eta,\mathfrak{k}} v_{\mathfrak{k}}^\eta - \lambda_{i,\eta,\mathfrak{k}} v_{\mathfrak{k}}^\eta}{\lambda_{i,\rho,\mathfrak{p}} - \lambda_{i,\eta,\mathfrak{k}}} = 0$$

hence $f_{\rho,\mathfrak{p}} v_{\mathfrak{k}}^\eta = 0$. Therefore the elements $f_{\rho,\mathfrak{p}}$, for $(\rho, \mathfrak{p}) \in \widehat{G}_n \times \mathfrak{P}$ are precisely the elements e_ρ defined in the proof of theorem 5.3.9, hence form a basis of GZ_n . Thus the set $\{f_i\}_{i \in [k]} \cup \{1_{G_n}\}$ generates GZ_n .

The other implication follows immediately by part b) of corollary 5.3.10. □

Later we'll explicitly determine a set of generators for the GZ-algebra of symmetric groups, that will be focal in the O-V approach.

5.3.3 Simple branching

We now focus on determining when a chain of groups is multiplicity free, or in other words when the branching multigraph has simple branching, thus reducing to a branching graph.

We've seen in theorem 5.1.5 when a subgroup H of a group G is multiplicity free, and as we're interested in proving this property for a chain of groups the situation might well be amenable to induction, thus making the aforementioned theorem of practical use. Indeed this will be the case for symmetric groups, yet we'll follow more closely the original paper by Okounkov and Vershik and for this we prove here an alternative way of determining when a subgroup is multiplicity free.

Theorem 5.3.14. *Let $H \leq G$ be groups and consider the algebras $\mathbb{C}G$ and $\mathbb{C}H$, moreover let $C_{\mathbb{C}G}(\mathbb{C}H)$ be the centralizer of $\mathbb{C}H$ in $\mathbb{C}G$, that is*

$$C_{\mathbb{C}G}(\mathbb{C}H) = \{f \in \mathbb{C}G \mid \forall_{g \in \mathbb{C}H} fg = gf\}$$

Then the following statements are equivalent:

- a) H is a multiplicity free subgroup of G .
- b) $C_{\mathbb{C}G}(\mathbb{C}H)$ is a commutative algebra.

Proof. By theorem 5.1.5, H is a multiplicity free subgroup of G if and only if $(G \times H, \tilde{H})$ is a Gelfand pair, and by corollary 4.7.27, $(G \times H, \tilde{H})$ is a Gelfand pair if and only if the algebra $L(\tilde{H} \backslash (G \times H) / \tilde{H})$ is commutative.

We claim that $L(\tilde{H} \backslash (G \times H) / \tilde{H})$ is in fact isomorphic to $C_{\mathbb{C}G}(\mathbb{C}H)$. Indeed consider the map

$$\begin{aligned} \Theta : L(\tilde{H} \backslash (G \times H) / \tilde{H}) &\rightarrow C_{\mathbb{C}G}(\mathbb{C}H) \\ F = \sum_{(g,h) \in G \times H} \alpha_{(g,h)}(g,h) &\mapsto f = |H| \sum_{g \in G} \alpha_{(g,1_G)} g \end{aligned}$$

Given $h \in H$

$$hf h^{-1} = |H| \sum_{g \in G} \alpha_{(g,1_G)} hgh^{-1} = |H| \sum_{h^{-1}gh \in G} \alpha_{(h^{-1}gh,1_G)} g$$

Now $\alpha_{(h^{-1}gh,1_G)} = \alpha_{(h^{-1},h^{-1})(g,1_G)(h,h)}$ and as F is in $L(\tilde{H} \backslash (G \times H) / \tilde{H})$, by its definition, $\alpha_{(h^{-1},h^{-1})(g,1_G)(h,h)} = \alpha_{(g,1_G)}$, thus $hf h^{-1} = f$. Hence $f \in C_{\mathbb{C}G}(\mathbb{C}H)$ and the map makes sense.

It's straightforward to see that Θ is a linear map, and to see it's bijective note that $\alpha_{(g,h)} = \alpha_{(gh^{-1}h,h)} = \alpha_{(gh^{-1},1_G)}$ hence F is uniquely determined by f , thus Θ is injective, and given $f = \sum_{g \in G} \alpha_g g \in C_{\mathbb{C}G}(\mathbb{C}H)$ letting

$$F = \frac{1}{|H|} \sum_{(g,h) \in G \times H} \alpha_{(gh^{-1})}(g,h)$$

one easily sees that $F \in L(\tilde{H} \setminus (G \times H) / \tilde{H})$, thus Θ is surjective. So we are left to prove that Θ preserves products.

Let $F(g,h)$ be the coefficient of (g,h) in F and $f(g)$ the coefficient of g in f , then

$$\begin{aligned} [\Theta(F_1 F_2)](g) &= |H| [F_1 F_2](g, 1_G) = |H| \sum_{s \in G} \sum_{h \in H} F_1(gs, h) F_2(s^{-1}, h^{-1}) = \\ &= |H| \sum_{s \in G} \sum_{h \in H} F_1(gsh^{-1}, 1_G) F_2(hs^{-1}, 1_G) = |H|^2 \sum_{t \in G} F_1(gt, 1_G) F_2(t^{-1}, 1_G) \\ &= [\Theta(F_1) \Theta(F_2)](g) \end{aligned}$$

□

Considering theorem 5.1.5 and the above the next result follows immediately.

Corollary 5.3.15. *For $H \leq G$ groups the following statements are equivalent:*

- a) H is a multiplicity free subgroup of G .
- b) $(G \times H, \tilde{H})$ is a Gelfand pair.
- c) The centralizer $C_{\mathbb{C}G}(\mathbb{C}H)$ is commutative.

5.4. Modern representation theory of S_n

The primary objectives in the development of the modern representation theory of S_n is to, first construct an alternative to the Young graph without resorting to tableaux, namely the branching graph, and then prove that indeed this alternative construction is equivalent to the Young graph. Once that is done we get a strong link between the new approach and the classical one, thus the rest of the theory may follow from classical results.

As stated previously we first prove that the chain of symmetric groups is multiplicity free, that is we have simple branching and so indeed a branching graph. With that done we have a basis for the irreducibles of symmetric groups, namely the GZ-bases, and considering the bases of Specht modules are in bijection with standard Young tableaux, as

stated at the end of Section 5.2.2, there must be a bijection between the GZ-bases and standard Young tableaux. Once we find that bijection we'll use it to finally prove that the branching graph and the Young graph are the same, or in other words they are isomorphic graphs.

5.4.1 Simple branching for S_n

To make full use of the theory developed in the previous section we first need to prove that the chain of symmetric groups $S_1 \leq S_2 \leq S_3 \leq \dots$ is multiplicity free. As stated we'll do it using theorem 5.3.14 so as to follow the original work of Okounkov and Vershik, for a proof that the chain of symmetric groups is multiplicity free using corollary 5.1.6 we refer the reader to [CSST10, 3.2.1].

In fact we'll prove this twice, as in the Okounkov and Vershik's paper. Here using complexification of involutive algebras as in appendices D.2.1 and D.2.2 together with theorem D.2.12, and then in the next subsection we'll see another proof using information we'll gather about a particular generating set for the centralizer, that will follow easily. None the less the proof here is the main one, as it is more general and requires only elementary properties of symmetric groups, thus it can be better applied in other situations.

We first remark that the group algebra of a group G over $\mathbb{C}$ or $\mathbb{R}$ can be made involutive by equipping it with the involution

$$\left(\sum_{g \in G} \alpha_g g \right)^* \mapsto \sum_{g \in G} \bar{\alpha}_g g^{-1} \quad (5.10)$$

It's straightforward to verify this is indeed an involution on $\mathbb{C}G$ or $\mathbb{R}G$, as defined in D.2.8. Moreover, recalling definition D.2.11 it's clear that $\mathbb{C}G$ is the $*$ -complexification of $\mathbb{R}G$. Consequently considering $H \leq G$ then the centralizer $C_{\mathbb{C}G}(\mathbb{C}H)$ is the $*$ -complexification of $C_{\mathbb{R}G}(\mathbb{R}H)$, indeed given f in the $*$ -complexification of $C_{\mathbb{R}G}(\mathbb{R}H)$ and $r \in \mathbb{C}H$ then $f = f_1 + if_2$ for $f_1, f_2 \in C_{\mathbb{R}G}(\mathbb{R}H)$ and we may write r as $r = r_1 + ir_2$ with $r_1, r_2 \in \mathbb{R}H$, with this we have that

$$\begin{aligned} fr &= (f_1 + if_2)(r_1 + ir_2) = f_1r_2 + if_1r_2 + if_2r_1 + i^2f_2r_2 = \\ & r_2f_1 + ir_2f_1 + ir_1f_2 + i^2r_2f_2 = (r_1 + ir_2)(f_1 + if_2) = rf \end{aligned}$$

Hence $f \in C_{\mathbb{C}G}(\mathbb{C}H)$, moreover given $f \in C_{\mathbb{C}G}(\mathbb{C}H)$ then we may write $f = f_1 + if_2$ with $f_1, f_2 \in \mathbb{R}G$ and given $r \in \mathbb{R}H \subset \mathbb{C}H$ then

$$(f_1r + if_2r) = fr = rf = (rf_1 + irf_2)$$

By definition of complexification, just as in complex numbers, 1 and i are linearly independent, hence $f_1r = rf_1$ and $f_2r = rf_2$, hence $f_1, f_2 \in C_{\mathbb{R}G}(\mathbb{R}H)$ and so f is in the $*$ -complexification of $C_{\mathbb{R}G}(\mathbb{R}H)$. With this we can use part c) of theorem D.2.12 to prove that $C_{\mathbb{C}S_{n+1}}(\mathbb{C}S_n)$ is commutative, for all $n \in \mathbb{N}$.

Theorem 5.4.1. *The centralizer $C_{\mathbb{C}S_{n+1}}(\mathbb{C}S_n)$ is commutative for all $n \in \mathbb{N}$.*

Proof. By our remarks above and part c) of theorem D.2.12 it's enough to prove that every element of $C_{\mathbb{R}S_{n+1}}(\mathbb{R}S_n)$ is self-adjoint. As such let

$$f = \sum_{\pi \in S_{n+1}} \alpha_\pi \pi \in C_{\mathbb{R}S_{n+1}}(\mathbb{R}S_n)$$

and consider the involution in 5.10 on $\mathbb{R}S_{n+1}$. If f is self-adjoint then, as $\alpha_\pi \in \mathbb{R}$ and so $\overline{\alpha_\pi} = \alpha_\pi$, it must be the case that $\alpha_\pi = \alpha_{\pi^{-1}}$ for all $\pi \in S_{n+1}$, so this is what we have to prove.

Let $\pi, \sigma \in S_{n+1}$, as σ and σ^{-1} have the same cycle structure they belong to the same conjugacy class, thus σ is conjugate to its own inverse.* Let then $\tau \in S_{n+1}$ be such that $\tau\pi\tau^{-1} = \sigma$, if it is the case that $\sigma \in S_n$ we can also find $\tau \in S_n$ that conjugates σ and σ^{-1} , and as $f \in C_{\mathbb{R}S_{n+1}}(\mathbb{R}S_n)$ then

$$f = \tau\tau^{-1}f = \tau f \tau^{-1} = \sum_{\pi \in S_{n+1}} \alpha_\pi \tau\pi\tau^{-1}$$

Thus it must be the case that $\alpha_\sigma = \alpha_{\sigma^{-1}}$, and so for any $\sigma \in S_n$, $\alpha_\sigma = \alpha_{\sigma^{-1}}$. Hence we now need only to focus in the case where $\sigma \in S_{n+1} \setminus S_n$. We may construct τ explicitly, indeed let

$$\sigma = (a_1 a_2 a_3 \cdots a_{n_1}) \cdots (x_1 x_2 x_3 \cdots x_{n_k})$$

be the cycle decomposition of σ , with $x_{n_k} = n$, then

$$\sigma^{-1} = (a_{n_1} a_{n_1-1} a_{n_1-2} \cdots a_1) \cdots (x_{n_k} x_{n_k-1} x_{n_k-2} \cdots x_1)$$

*This property is usually called *ambivalence*.

This may be rewritten as

$$\sigma^{-1} = (a_{n_1} a_{n_1-1} a_{n_1-2} \cdots a_1) \cdots (x_{n_k-1} x_{n_k-2} \cdots x_1 x_{n_k})$$

Now, using table notation for a permutation, one may verify that the permutation defined as

$$\tau = \begin{pmatrix} a_1 & a_2 & a_3 & \cdots & a_{n_1} & \cdots & x_1 & x_2 & \cdots & x_{n_k-1} & x_{n_k} \\ a_{n_1} & a_{n_1-1} & a_{n_1-2} & \cdots & a_1 & \cdots & x_{n_k-1} & x_{n_k-2} & \cdots & x_1 & x_{n_k} \end{pmatrix}$$

conjugates σ and σ^{-1} , yet note that τ fixes $x_{n_k} = n$, hence $\tau \in S_n$. Now as before we have that

$$f = \tau \tau^{-1} f = \tau f \tau^{-1} = \sum_{\pi \in S_{n+1}} \alpha_\pi \tau \pi \tau^{-1}$$

Therefore we can conclude that for any $\sigma \in S_{n+1}$, $\alpha_\sigma = \alpha_{\sigma^{-1}}$, thus f is self-adjoint. $\square$

The desired result of this subsection now follows from theorem 5.3.14.

Corollary 5.4.2. *For any $n \in \mathbb{N}$, S_n is a multiplicity free subgroup of S_{n+1} , i.e. the chain $S_1 \leq S_2 \leq S_3 \leq \cdots$ is multiplicity free. In particular the branching graph of this chain has simple branching.*

5.4.2 The YJM elements

We now focus on finding a generating set for the GZ-algebras GZ_n of the chain of symmetric groups. As we've seen in proposition 5.3.13 a generating set for GZ_n separates the respective GZ-bases, hence there is a bijection between the GZ-bases and tuples of eigenvalues of the generating set of GZ_n , as in definition 5.3.12. Moreover, as we know from the classical representation theory of symmetric groups, there is a bijection between standard Young tableaux of partitions of n and particular bases of the irreducible representations of S_n (refer to the statements on a basis for Specht modules in Subsection 5.2.2, after definition 5.2.12), thus there must be a bijection between standard Young tableaux of partitions n and those tuples of eigenvalues of the generating set of GZ_n . Our job for the next three subsections will be to find and exploit that bijection.

In order to determine a generating set for the GZ-algebras we first determine a generating set for the centers $Z_n = Z(\mathbb{C}S_n)$ and the centralizers $C_n = C_{\mathbb{C}S_{n+1}}(\mathbb{C}S_n)$.

Definition 5.4.3. Fix $n > 1$ and let, for $i \in \{2, \dots, n\}$, $Y_i \in \mathbb{C}S_n$ be the sum of all cycles of length i (i -cycles) in S_{n-1} . Additionally define $Y'_i \in \mathbb{C}S_n$ to be the sum of all i -cycles in S_n containing n . By convention we let $Y_n = 0$.

Recall now definition 5.2.1.

Definition 5.4.4. Fix $n \in \mathbb{N}$ and let, for $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k) \vdash n$ and $i = \lambda_j$ for some $j \in [k]$, $c_{\lambda,i} \in \mathbb{C}S_n$ be the sum of all permutations $\pi \in S_n$ such that π has cycle structure λ and the cycle that contains n is an i -cycle.

Proposition 5.4.5. Considering the definitions above, $Y_i = c_{\lambda,1}$ and $Y'_i = c_{\lambda,i}$ with $\lambda = (i, 1, \dots, 1) \vdash n$.

Definition 5.4.6. Let $n \in \mathbb{N}$ and $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k) \vdash n$. We define

$$\Sigma\lambda = \sum_{\substack{i \in [k], \\ \lambda_i > 1}} \lambda_i$$

Lemma 5.4.7. Considering the definitions above:

a) For $n > 1$ the set $\{c_{\lambda,i} \mid \lambda \vdash n, i \in [n]\}$ is a basis of C_{n-1} . In particular

$$\langle Y_2, \dots, Y_{n-1}, Y'_2, \dots, Y'_n \rangle \subseteq C_{n-1}$$

b) For $\lambda \vdash n$ such that $\Sigma\lambda = k$ and $i \in [n]$ $c_{\lambda,i} \in \langle Y_2, \dots, Y_k, Y'_2, \dots, Y'_k \rangle$.

c) $C_{n-1} = \langle Y_2, \dots, Y_{n-1}, Y'_2, \dots, Y'_n \rangle$

d) $Z_{n-1} = \langle Y_2, \dots, Y_{n-1} \rangle$

Proof. a) Given $c_{\lambda,i} = \sum \pi$ where π has cycle structure λ and n is in a cycle of length i we know that for any $\tau \in S_{n-1}$, $\tau^{-1}\pi\tau$ has the same cycle structure of π , moreover letting

$$\pi = (a_1 a_2 a_3 \cdots a_{n_1}) \cdots (x_1 x_2 x_3 \cdots x_i)$$

with $x_i = n$ then

$$\tau^{-1}\pi\tau = \tau^{-1}(a_1 a_2 a_3 \cdots a_{n_1})\tau\tau^{-1} \cdots \tau\tau^{-1}(x_1 x_2 x_3 \cdots x_i)\tau$$

Now $\tau^{-1}(x_1 x_2 x_3 \cdots x_i)\tau$ will still be a cycle of length i , moreover because $\tau \in S_{n-1}$ it does not affect n thus it's clear that n will still be in this cycle, hence $\tau^{-1}\pi\tau$ has cycle structure λ and n is in a cycle of length i . With this we can conclude that $c_{\lambda,i}$

is invariant under conjugation by elements of S_{n-1} and so commutes with elements in $\mathbb{C}S_{n-1}$, thus $c_{\lambda,i} \in C_{n-1}$.

It is also clear that the elements $c_{\lambda,i}$ are linearly independent, there is no way to get an element $c_{\lambda,i}$ by a linear combination of elements $c_{\lambda',j}$ as the summands will either have different cycle structures from those in $c_{\lambda,i}$ or n will be in a cycle of different length. So we are left to prove that any element in C_{n-1} can be written as a linear combination of elements in $\{c_{\lambda,i} \mid \lambda \vdash n, i \in [n]\}$.

Let $f \in C_{n-1}$ such that

$$f = \sum_{\pi \in S_n} \alpha_\pi \pi$$

Now $f \in C_{n-1}$ is the same as saying that for any $\tau \in S_{n-1}$, $\tau f = f\tau \Leftrightarrow f = \tau^{-1}f\tau$, hence $\alpha_\pi = \alpha_{\tau^{-1}\pi\tau}$ for any $\pi \in S_n$. By what we've seen above $\tau^{-1}\pi\tau$ has the same cycle structure as π and n will remain in a cycle of the same length as in π , hence the coefficients of f have to be constant not only on the cycle types but also on the cycle lengths of the cycles that include n . Or in other words f decomposes as a sum of elements in $\{c_{\lambda,i} \mid \lambda \vdash n, i \in [n]\}$.

The remaining statement left to prove, namely that

$$\langle Y_2, \dots, Y_{n-1}, Y'_2, \dots, Y'_n \rangle \subseteq C_{n-1}$$

now follows immediately from proposition 5.4.5.

b) We prove this by induction on $\Sigma\lambda = k$. If $k = 2$ then $c_{\lambda,i}$ is the sum of all cycles of length 2 that, depending on whether $i = 1$ or $i = 2$, fix or permute n , hence is equal to Y_2 or Y'_2 respectively. As such we now assume for the induction step that the statement in b) is true for all $\Sigma\lambda \leq k$. Let $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k) \vdash n$ with $\Sigma\lambda = k + 1$, moreover let $\lambda'_1, \lambda'_2, \dots, \lambda'_l \in \lambda$ be the elements greater than 1 in λ . We split the problem in two cases:

i) If $i = 1$ (in $c_{\lambda,i}$) then consider the product $Y_{\lambda'_1} Y_{\lambda'_2} \cdots Y_{\lambda'_l}$, by part a) this is an element of C_{n-1} , moreover by the definition of Y_i it is clearly a sum of permutations that fix n , hence again by part a) it can be written as

$$Y_{\lambda'_1} Y_{\lambda'_2} \cdots Y_{\lambda'_l} = \alpha_{\lambda,1} c_{\lambda,1} + \sum_{\substack{\mu \vdash n, \\ \Sigma\mu \leq k}} \alpha_{\mu,1} c_{\mu,1}$$

We note that $\alpha_{\lambda,1} \neq 0$ as there are permutations in $Y_{\lambda'_1} Y_{\lambda'_2} \cdots Y_{\lambda'_l}$ with cycle structure λ , and the remaining permutations must have cycle structure μ such

that $\Sigma\mu \leq k$, hence by induction, as $\lambda'_i \leq k + 1$ for all $i \in [l]$

$$c_{\lambda,i} \in \langle Y_2, \dots, Y_{k+1}, Y'_2, \dots, Y'_{k+1} \rangle$$

ii) If $i > 1$, then $i = \lambda'_i$ for some $i \in [l]$, with this consider now the product

$Y_{\lambda'_1} Y_{\lambda'_2} \cdots Y'_{\lambda'_i} Y_{\lambda'_{i+1}} \cdots Y_{\lambda'_l}$, as before by part a) we have that

$$Y_{\lambda'_1} Y_{\lambda'_2} \cdots Y'_{\lambda'_i} Y_{\lambda'_{i+1}} \cdots Y_{\lambda'_l} = \alpha_{\lambda,i} c_{\lambda,i} + \sum_{\substack{\mu \vdash n, \\ \Sigma\mu \leq k}} \alpha_{\mu,j} c_{\mu,j}$$

Again it must be the case that $\alpha_{\lambda,i} \neq 0$ and by induction the conclusion follows as before.

c) This now follows immediately by part a) and b).

d) Simply note that $\langle Y_2, \dots, Y_{n-1} \rangle$ generates any element of the form $c_{\lambda,1}$, and the summands in this element are permutation in S_{n-1} , moreover they are all the elements of S_{n-1} with cycle structure λ , hence $c_{\lambda,1}$ is the sum of all elements in the conjugation class of S_{n-1} of permutations with cycle structure λ . Now note that any element in Z_{n-1} must be constant in the conjugation classes, as it must permute with any element in S_{n-1} , hence $\{c_{\lambda,1} \mid \lambda \vdash n\}$ generates Z_{n-1} and as we've seen this set is linearly independent thus it must be a basis of Z_{n-1} .

□

We now move to the generating set for the GZ-algebras.

Definition 5.4.8. Fix $n \in \mathbb{N}$, for $i \in [n]$ define

$$X_i = \sum_{j \in [i-1]} (j \ i)$$

This is a sum of 2-cycles, we convention $X_1 = 0$. These elements are called the Young-Jucys-Murphy elements of $\mathbb{C}S_n$, YJM elements for short.

These elements were already known prior to the work of Okounkov and Vershik, as their name indicates, and already played an important role in the representation theory of symmetric groups [Mur81]. For instance it was known, due to A. Jucys [Juc74], that they could be used to construct a generating set for the centers Z_n , so considering the definition of a GZ-algebra this alone should hint as to why they were considered by Okounkov and Vershik.

Recall that a GZ-algebra GZ_n is commutative and generated by the centers Z_i for $i \leq n$. We remark that a YJM element X_i is the difference between the sum of all 2-cycles in S_i minus the sum of all 2-cycles in S_{i-1} , hence it is the difference between an element of Z_i and an element of Z_{i-1} , therefore $X_i \in \text{GZ}_n$ for all $i \in [n]$, in particular the YJM elements commute. Moreover $X_i \notin Z_i$ for all $i \in [n]$ with $i > 1$, for instance X_i and $(1\ i)$ don't commute, this is very important as we shall see.

Theorem 5.4.9. *Considering the YJM elements defined above:*

- a) For $n > 1$, $C_{n-1} = \langle Z_{n-1}, X_n \rangle$.
- b) For $n > 1$, $\text{GZ}_n = \langle X_1, \dots, X_n \rangle$.

Proof. a) On one hand, noting that $X_n = Y'_2$, by part c) and d) of lemma 5.4.7

$$\langle Z_{n-1}, X_n \rangle \subseteq C_{n-1}$$

Now for the other inclusion we prove that $Y'_k \in \langle Z_{n-1}, X_n \rangle$ for all $k \in \{3, \dots, n\}$, and so the proof follows by part c) of lemma 5.4.7. We prove this by induction, as we already have $Y'_2 \in \langle Z_{n-1}, X_n \rangle$ we assume that $Y'_i \in \langle Z_{n-1}, X_n \rangle$ for all $i \in \{2, \dots, k+1\}$ and attempt to prove that $Y'_{k+2} \in \langle Z_{n-1}, X_n \rangle$. We may write Y'_{k+1} as

$$Y'_{k+1} = \sum_{i_1, \dots, i_k} (i_1 \dots i_k n)$$

summing over all distinct $i_1, \dots, i_k \in [n-1]$. Now consider

$$Y'_{k+1} X_n = \left(\sum_{i_1, \dots, i_k} (i_1 \dots i_k n) \right) \left(\sum_{i \in [n-1]} (i n) \right) = \sum_{\substack{i_1, \dots, i_k, \\ i \in [n-1]}} (i_1 \dots i_k n)(i n)$$

By assumption $Y'_{k+1} X_n \in \langle Z_{n-1}, X_n \rangle$. There are two possibilities for a summand $(i_1 \dots i_k n)(i n)$ in $Y'_{k+1} X_n$; either $i = i_j$ for some $j \in [k]$ or $i \neq i_j$ for all $j \in [k]$.

- i) If $i = i_j$ for some $j \in [k]$ then $(i_1 \dots i_k n)(i n) = (i_1 \dots i_j)(i_{j+1} \dots n)$.
- ii) If $i \neq i_j$ for all $j \in [k]$ then $(i_1 \dots i_k n)(i n) = (i_1 \dots i_k n, i)$.

Hence $Y'_{k+1} X_n$ decomposes into two sums as

$$Y'_{k+1} X_n = \sum_{i, i_1, \dots, i_k} (i, i_1 \dots i_k n) + \sum_{i_1, \dots, i_k} \sum_{j=1}^k (i_1 \dots i_j)(i_{j+1} \dots n)$$

Where the sums are over all distinct $i, i_1, \dots, i_k \in [n-1]$. Now note that the first sum (on the left) is equal to Y'_{k+2} and as $X_n = Y'_2$ we have that the second sum is equal to $Y'_{k+1}Y'_2 - Y'_{k+2} \in C_{n-1}$ hence we may apply part a) of lemma 5.4.7 to rewrite $Y'_{k+1}X_n$ as

$$Y'_{k+1}X_n = Y'_{k+2} + \sum_{\lambda, j} \alpha_{\lambda, j} c_{\lambda, j}$$

Where $\sum \lambda \leq k+1$. Therefore

$$Y'_{k+2} = Y'_{k+1}X_n - \sum_{\lambda, j} \alpha_{\lambda, j} c_{\lambda, j}$$

Now by the induction hypotheses and part b) of lemma 5.4.7 it must be the case that $Y'_{k+2} \in \langle Z_{n-1}, X_n \rangle$, concluding the proof.

- b) This follows by induction on n , first noting that $\text{GZ}_2 = \mathbb{C}S_2 = \langle X_1, X_2 \rangle$ as although $X_1 = 0$ we have $X_2 = (1\ 2)$. Assuming GZ_{n-1} is generated by $\langle X_1, X_2, \dots, X_{n-1} \rangle$ we need to prove that $\text{GZ}_n = \langle \text{GZ}_{n-1}, X_n \rangle$.

On one hand it is clear that $\langle \text{GZ}_{n-1}, X_n \rangle \subseteq \text{GZ}_n$ as we've seen that $X_n \in \text{GZ}_n$. For the other inclusion we need only to prove that $Z_n \subseteq \langle \text{GZ}_{n-1}, X_n \rangle$, but this is straightforward as by part a) of this theorem

$$Z_n \subseteq C_{n-1} = \langle Z_{n-1}, X_n \rangle \subseteq \langle \text{GZ}_{n-1}, X_n \rangle$$

Therefore $\text{GZ}_n = \langle \text{GZ}_{n-1}, X_n \rangle = \langle X_1, X_2, \dots, X_n \rangle$.

□

Corollary 5.4.10. *The algebra C_n is commutative for all $n \in \mathbb{N}$.*

Proof. Indeed by part a) of the previous theorem, $C_n = \langle Z_n, X_{n+1} \rangle$ and as $X_{n+1} \in \text{GZ}_{n+1}$ then X_{n+1} commutes with every element in Z_n , thus C_n is commutative. □

Recall now the remarks made at the beginning of this subsection.

Definition 5.4.11. *For v a GZ-vector of S_n let $\alpha(v) = (a_1, a_2, \dots, a_n)$ where a_i is the eigenvalue of X_i on v . We call $\alpha(v)$ the weight of v , note that $a_1 = 0$ for any GZ-vector v since $X_1 = 0$. Moreover for $n \in \mathbb{N}$ let*

$$\text{Spec}(n) = \{\alpha(v) \mid v \text{ a GZ-vector of } S_n\}$$

We call $\text{Spec}(n)$ the spectrum of the YJM elements of $\mathbb{C}S_n$.

As GZ-vectors correspond uniquely to paths in the branching graph we can see α as the map in definition 5.3.12, in other words given a GZ-vector v_p^ρ we can talk about $\alpha(v_p^\rho)$ or equivalently $\alpha(\rho, p)$, or even more simply just $\alpha(p)$, and as stated this is a bijection, hence given $\alpha \in \text{Spec}(n)$ we may also refer to the respective path in the branching graph as p_α and the corresponding GZ-vector by v_α . Moreover by part a) of corollary 5.3.10

$$|\text{Spec}(n)| = \sum_{\rho V_i \in \widehat{S}_n} \dim_{\mathbb{C}}(V_i) = \dim_{\mathbb{C}}(\text{GZ}_n)$$

By part b) of the same corollary a GZ-vector v is uniquely determined by $\alpha(v)$, up to a scalar factor. In the context of symmetric groups a GZ-basis is usually called a *Young basis*, as A. Young considered it in his work, yet since the notions of GZ-algebra and YJM elements were not known to him he could not describe them as completely as Okounkov and Vershik did.* For this, from now on we refer to a GZ-basis of the chain of symmetric groups as a Young basis and a vector in such a basis as a Young vector.

There is a simple equivalence relation we may define on $\text{Spec}(n)$.

Definition 5.4.12. For $\alpha, \beta \in \text{Spec}(n)$ we define the equivalence relation

$$\alpha \sim \beta \Leftrightarrow p_\alpha, p_\beta \in \mathfrak{P}^\rho$$

for some $\rho \in \widehat{S}_n$. That is the paths p_α, p_β start at the same vertex in the branching graphs, or in other words v_α, v_β belong to the same Young basis, of some irreducible S_n -space V associated with an irreducible representation ρ .

Clearly $|\text{Spec}(n)/\sim| = |\widehat{S}_n|$, we'll come back to this equivalence relation later but for now, recalling definition 5.2.27, one can see that $\sim$ and $\approx$ are very much related.

5.4.3 Coxeter generators II

Recall the definition of Coxeter generators in 5.2.19, in this subsection we'll use the Coxeter relations together with the YJM elements in order to extract some information on the spectrum of the YJM elements, that is $\text{Spec}(n)$ as in definition 5.4.11. As a consequence of this we'll also be able to better understand how the Coxeter generators and the YJM elements act on a Young basis. This will be fundamental in finding the bijection between $\text{Spec}(n)$ and standard tableaux of n , that is $\text{SYT}(n)$.

*Although Young's name appears in YJM this is due to their implicit appearance in his work, he did not consider the YJM elements explicitly and indeed as stated he was not fully aware of their existence.

In definition 5.2.19 we saw that Coxeter generator commute except when they're neighbors (that is $|i - j| = 1$), this is usually stated as saying that the Coxeter generators are *local*, in the sense that they interact locally with each other. Due to their locality and the properties of the GZ-bases they'll act on Young vectors in a "local" way.

Proposition 5.4.13. *Consider the path*

$$\mathbf{p} = (\rho^n \rightarrow \rho^{n-1} \rightarrow \cdots \rightarrow \rho^2 \rightarrow \iota)$$

in the branching graph of the symmetric groups with $\rho^i \in \widehat{S}_i$ and let $k \in [n - 1]$. Then considering the Coxeter generator s_k and the Young vector $v_{\mathbf{p}}$ we have that $s_k v_{\mathbf{p}}$ is a linear combination of Young vectors $v_{\mathbf{p}'}$ where $\mathbf{p}'$ are paths of the form

$$\mathbf{p}' = (\sigma^n \rightarrow \sigma^{n-1} \rightarrow \cdots \rightarrow \sigma^2 \rightarrow \iota)$$

where $\sigma^i = \rho^i$ for $i \neq k$ and σ^k is a possible (for such paths in the branching graph) irreducible in $\widehat{S}_k$.

Proof. For $i > k$ since $s_k \in S_i$ and $v_{\mathbf{p}} \in V_i$, where V_i is the irreducible S_i -space associated with ρ^i , then $s_k v_{\mathbf{p}} \in \mathbb{C}S_i v_{\mathbf{p}}$ and $\mathbb{C}S_i v_{\mathbf{p}} = V_i$ (recall our discussion following definition 5.3.7). Hence for $i > k$, $s_k v_{\mathbf{p}} \in V_i$, thus $s_k v_{\mathbf{p}}$ is a linear combination of Young vectors $v_{\mathbf{p}'}$ where the paths $\mathbf{p}'$, using the notation above, satisfy $\sigma^i = \rho^i$ for $i > k$.

For $i < k$, note that s_k commutes with any element in S_i , hence $\mathbb{C}S_i s_k v_{\mathbf{p}} = s_k \mathbb{C}S_i v_{\mathbf{p}} = s_k V_i$ and s_k is an S_i -linear map. Considering V_i is an irreducible S_i -space then by Schur's lemma $s_k V_i \subseteq V_i$, thus $s_k v_{\mathbf{p}} \in V_i$ for $i < k$ and the conclusion follows. $\square$

The above proposition may also be proven using the simple relations

$$s_i X_j = X_j s_i \text{ for } j \notin \{i, i + 1\} \quad (5.11)$$

These relations are straightforward to prove and with them we get that $s_k v_{\mathbf{p}}$ is an eigenvector of X_j for $j \notin \{k, k + 1\}$ associated with an eigenvalue α_j , call the respective eigenspace E_j . Then clearly $v_{\mathbf{p}}$ also belongs to E_j for all $j \notin \{k, k + 1\}$, now by part b) of theorem 5.4.9, $\text{GZ}_{k-1} = \langle X_1, \dots, X_{k-1} \rangle$ hence by part b) of corollary 5.3.10, $s_k v_{\mathbf{p}}$ is a linear combination of Young vectors $v_{\mathbf{p}'}$ such that

$$\mathbf{p}' = (\sigma^n \rightarrow \sigma^{n-1} \rightarrow \cdots \rightarrow \sigma^2 \rightarrow \iota)$$

where $\sigma^i = \rho^i$ for $i < k$. On the other hand, as we noted in the above proof, $s_k v_{\mathbf{p}} \in V_i$ for all $i > k$, thus the conclusion follows.

We can further add to the above relations the following remarkable, and well known, relation

$$s_i X_i + 1 = X_{i+1} s_i \quad (5.12)$$

valid for all $i \in \mathbb{N}$. If we consider the space $V = \langle v_p, s_i v_p \rangle$ then the Coxeter relations together with 5.11 and 5.12 imply that V is invariant for the action of s_i , X_i and X_{i+1} .

In what follows we are going to fully exploit X_i , X_{i+1} and s_i , yet to do so we need to consider the subalgebra of $\mathbb{C}S_{i+1}$ generated by them.

Definition 5.4.14. Let $H(2)$ be the algebra generated by elements h_1 , h_2 and s subject to the relations*

$$s^2 = 1; \quad h_1 h_2 = h_2 h_1; \quad s h_1 + 1 = h_2 s$$

The algebra $H(2)$ above is the simplest example of what are called *degenerate affine Hecke algebras*, note the generator h_2 is superfluous as by the first and last relations $h_2 = s h_1 s + s$, yet it's more convenient to consider it. The importance of the algebra $H(2)$ here comes from the following fact.

Proposition 5.4.15. The algebra $\mathbb{C}S_{i+1}$ is generated by $\mathbb{C}S_i$ and $H(2)$ where $H(2)$ has generators $h_1 = X_i$, $h_2 = X_{i+1}$ and $s = s_i$.

This is straightforward as we've seen that indeed X_i , X_{i+1} and s_i satisfy the required relations of the generators of $H(2)$ and S_{i+1} is clearly seen to be generated by S_i and s_i . Yet the superfluous generators X_i and X_{i+1} allow us to better use induction, that is at each step going from i to $i + 1$ we reduce the problem to simply focusing on the algebra $H(2)$.

Theorem 5.4.16. Let v_α be a Young vector with respective weights $\alpha = (a_1, a_2, \dots, a_n)$, then

- a) For all $i \in [n - 1]$, $a_i \neq a_{i+1}$.
- b) For $i \in [n - 1]$, $a_{i+1} = a_i \pm 1 \Leftrightarrow s_i v_\alpha$ and v_α are linearly dependent $\Leftrightarrow s_i v_\alpha = \pm v_\alpha$
- c) For $i \in [n - 1]$ it cannot be the case that $a_i = a_{i+1} + 1 = a_{i+2}$ or $a_i = a_{i+1} - 1 = a_{i+2}$.
- d) If $a_{i+1} \neq a_i \pm 1$ then

$$\alpha' = (a_1, a_2, \dots, a_{i+1}, a_i, \dots, a_n) \in \text{Spec}(n)$$

*This is analogous to the concept of group presentations, refer to [Rot02, sec 5.5] for more on this.

and $\alpha \sim \alpha'$, with $\sim$ as in definition 5.4.12. Moreover

$$v_{\alpha'} = z \left(s_i - \frac{1}{a_{i+1} - a_i} \right) v_{\alpha}$$

with $z \in \mathbb{C}$, thus the action of X_i , X_{i+1} and s_i on the basis $(v_{\alpha}, v_{\alpha'})$ is given by

$$X_i \rightarrow \begin{bmatrix} a_i & 0 \\ 0 & a_{i+1} \end{bmatrix}; \quad X_{i+1} \rightarrow \begin{bmatrix} a_{i+1} & 0 \\ 0 & a_i \end{bmatrix}; \quad s_i \rightarrow \begin{bmatrix} \frac{1}{a_{i+1} - a_i} & z \left(1 - \frac{1}{(a_{i+1} - a_i)^2} \right) \\ \frac{1}{z} & \frac{1}{a_i - a_{i+1}} \end{bmatrix}$$

Proof. As stated before the space $V = \langle v_{\alpha}, s_i v_{\alpha} \rangle$ is invariant for the action of X_i , X_{i+1} and s_i , with respective actions, in the basis $(v_{\alpha}, s_i v_{\alpha})$ (assuming of course these vector are linearly independent, to produce a basis), given by

$$X_i \rightarrow \begin{bmatrix} a_i & -1 \\ 0 & a_{i+1} \end{bmatrix}; \quad X_{i+1} \rightarrow \begin{bmatrix} a_{i+1} & 1 \\ 0 & a_i \end{bmatrix}; \quad s_i \rightarrow \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Hence we get a 2-dimensional representation of $H(2)$.

- a) If $a_i = a_{i+1}$ then one may easily verify that the matrix above associated with X_i is not diagonalizable, but X_i is a diagonal operator in the Young bases, hence diagonalizable on any subspace, thus we get a contradiction.
- b) If $a_{i+1} = a_i \pm 1$ then assume $a_{i+1} = a_i + 1$ (the other case is analogous). Consider the subspace of V , $V_1 = \langle v_{\alpha} - s_i v_{\alpha} \rangle$, it is straightforward to verify that this space is invariant for the action of X_i , X_{i+1} and s_i , thus it's a one-dimensional subrepresentation of $H(2)$. Now as we've seen $H(2)$ is semisimple, thus completely reducible by corollary D.1.24, hence there must exist a complementary (to V_1) $H(2)$ -invariant subspace of V , consequently also of dimension 1. We now prove that such space coincides with V_1 and thus it must be the case that V is one dimensional, hence v_{α} and $s_i v_{\alpha}$ are linearly dependent.

Indeed suppose that V_2 is an $H(2)$ -invariant one dimensional subspace of V , as such let it be spanned by $bv_{\alpha} + cs_i v_{\alpha}$ for some $b, c \in \mathbb{C}$. As this space is $H(2)$ -invariant it must be invariant for the action of s_i , hence $c = \pm b \neq 0$. As we may assume without loss of generality that $b = 1$ then $c = \pm 1$, but then either $V_2 = V_1$ or V_2 is spanned by $v_{\alpha} + s_i v_{\alpha}$, which one may easily verify is not an $H(2)$ -invariant subspace of V . Thus $V_2 = V_1$.

Now v_{α} and $s_i v_{\alpha}$ are linearly dependent if and only if $s_i v_{\alpha} = \lambda v_{\alpha}$ for some $\lambda \in \mathbb{C}$. As $s_i^2 = 1$ then $\lambda^2 = 1$ and so $\lambda = \pm 1$, thus $s_i v_{\alpha} = \pm v_{\alpha}$. Hence $s_i v_{\alpha}$ and v_{α} being

linearly dependent is equivalent to stating that $s_i v_\alpha = \pm v_\alpha$. So we now need only to prove that $s_i v_\alpha = \pm v_\alpha \Rightarrow a_{i+1} = a_i \pm 1$, but this follows easily using relation 5.12, indeed by it we have that $(s_i X_i s_i + s_i) v_\alpha = X_{i+1} v_\alpha \Leftrightarrow (a_i \pm 1) v_\alpha = a_{i+1} v_\alpha$.

c) If $a_i = a_{i+1} - 1 = a_{i+2}$ then by part b), $s_i v_\alpha = v_\alpha$ and $s_{i+1} v_\alpha = -v_\alpha$. Now using the Coxeter relation $s_i s_{i+1} s_i = s_{i+1} s_i s_{i+1}$ we get that $v_\alpha = -v_\alpha$ which is not possible as v_α is non-zero. The other case is analogous.

d) Considering a matrix of the form

$$\begin{bmatrix} a & \pm 1 \\ 0 & b \end{bmatrix}$$

one may easily show that, if $a \neq b$ then it has eigenvalues a, b associated with the eigenvectors $(1, 0)$ and $(\pm 1/(b-a), 1)$ respectively. As by part a) we have that $a_i \neq a_{i+1}$ then $v' = s_i v_\alpha - \frac{1}{a_{i+1}-a_i} v_\alpha$ is an eigenvector of X_i and X_{i+1} with respective eigenvalues a_{i+1}, a_i , yet note that we need $a_{i+1} \neq a_i \pm 1$ otherwise $v' = 0$.

Now using relations 5.11 one may easily verify that v' is an eigenvector of X_j , for $j \neq i, i+1$, with respective eigenvalues a_j , hence v' has weights

$$\alpha' = (a_1, a_2, \dots, a_{i+1}, a_i, \dots, a_n)$$

thus v' is an eigenvector of every element of GZ_n and so by part c) of corollary 5.3.10, v' is a scalar multiple of the Young vector associated with the weights α' , thus $\alpha' \in \text{Spec}(n)$ and v' is a scalar multiple of $v_{\alpha'}$. Finally by proposition 5.4.13, $s_i v_\alpha \in V_n$ where V_n is the irreducible S_n -space such that $v_\alpha \in V_n$, thus $s_i v_\alpha - \frac{1}{a_{i+1}-a_i} v_\alpha \in V_n$ and so $\alpha \sim \alpha'$.

With this the matrices of X_i, X_{i+1} and s_i in the basis $(v_\alpha, v_{\alpha'})$ are straightforward to compute, indeed X_i and X_{i+1} are immediate from the above and for s_i we have

$$v_{\alpha'} = z \left(s_i - \frac{1}{a_{i+1}-a_i} \right) v_\alpha \Leftrightarrow s_i v_\alpha = \frac{1}{a_{i+1}-a_i} v_\alpha + \frac{v_{\alpha'}}{z}$$

And $s_i v_{\alpha'} = z \left(1 - \frac{1}{a_{i+1}-a_i} s_i \right) v_\alpha$, as we already know what $s_i v_\alpha$ is the conclusion now follows from a simple substitution.

□

Note in part d) of the above theorem that, if we let the Coxeter generators s_j act on $\alpha = (a_1, a_2, \dots, a_n) \in \text{Spec}(n)$ by

$$s_j \cdot \alpha = (a_{s_j(1)}, a_{s_j(2)}, \dots, a_{s_j(n)})$$

then $s_i \cdot \alpha = \alpha'$, and as we've seen, subject to the condition $a_{i+1} \neq a_i \pm 1$, $\alpha' \in \text{Spec}(n)$, moreover $\alpha \sim \alpha'$.

Definition 5.4.17. Let $\alpha = (a_1, a_2, \dots, a_n) \in \text{Spec}(n)$, if $a_{i+1} \neq a_i \pm 1$ for some $i \in [n-1]$ we say s_i is admissible for α .

As stated if s_i is admissible for α then $\alpha \sim s_i \alpha$. This will also later relate to definition 5.2.27 by the bijection between $\text{Spec}(n)$ and standard tableaux of n .

5.4.4 Content vectors

The bijection between $\text{Spec}(n)$ and standard tableaux of n will require an intermediate step, a suitable set that will facilitate linking the two aforementioned sets. So in fact we'll need to prove two bijections, one from $\text{Spec}(n)$ to that suitable set, and another from that set to the set of standard tableaux of n . Here we'll prove the latter mentioned bijection.

Definition 5.4.18. An n -tuple of integers $(a_1, a_2, \dots, a_n) \in \mathbb{Z}^n$ is said to be a content vector if

- a) $a_1 = 0$
- b) For all $i > 1$, $\{a_i - 1, a_i + 1\} \cap \{a_1, a_2, \dots, a_{i-1}\} \neq \emptyset$, that is the intersection is non-empty.
- c) If $a_i = a_j$ for some $i < j$ then $\{a_i - 1, a_i + 1\} \subseteq \{a_{i+1}, \dots, a_{j-1}\}$. That is, letting $a_i = a$, between any two occurrences of a there also occurs $a - 1$ and $a + 1$.

The set of all content vectors in $\mathbb{Z}^n$ is denoted by $\text{Cont}(n)$.

As an example one may verify that $\text{Cont}(1) = \{(0)\}$ and $\text{Cont}(2) = \{(0, 1), (0, -1)\}$. The set $\text{Cont}(n)$ will be the aforementioned suitable set, we may strengthen conditions b) and c) above as follows.

Proposition 5.4.19. Consider the set $\text{Cont}(n)$ as in the above definition and let $(a_1, a_2, \dots, a_n) \in \text{Cont}(n)$, then

b') For all $i > 1$, if $a_i \geq 0$ then $a_i - 1 = a_j$ for some $j < i$ and if $a_i \leq 0$ then $a_i + 1 = a_j$ for some $j < i$.

c') If $a_i = a_j$ for some $i < j$ and $a_r \neq a_i$ for all $r \in \{i + 1, \dots, j - 1\}$ then there exists unique indices $s^-, s^+ \in \{i + 1, \dots, j - 1\}$ such that $a_{s^-} = a_i - 1$ and $a_{s^+} = a_i + 1$.

Proof.

b') If $a_i \geq 0$ then by part b) of the above definition $\{a_i - 1, a_i + 1\} \cap \{a_1, a_2, \dots, a_{i-1}\} \neq \emptyset$, if we have $a_i - 1 \in \{a_1, a_2, \dots, a_{i-1}\}$ we are done, so assume we have just $a_i + 1 \in \{a_1, a_2, \dots, a_{i-1}\}$ and let then $a_i + 1 = a_j$ with $j < i$, then again by b) we must have either a_i or $a_i + 2$ in $\{a_1, a_2, \dots, a_{j-1}\}$, but if we have a_i then by c) we must have $a_i - 1 \in \{a_1, a_2, \dots, a_{i-1}\}$, so again assume we just have $a_i + 2 \in \{a_1, a_2, \dots, a_{j-1}\}$. Repeating this process we arrive at the contradiction that $a_1 > 0$, hence it must be the case that $a_i - 1 \in \{a_1, a_2, \dots, a_{i-1}\}$.

The case for $a_i \leq 0$ is analogous.

c') The existence of the indices s^-, s^+ is guaranteed by c), the uniqueness comes from the condition $a_r \neq a_i$ for all $r \in \{i + 1, \dots, j - 1\}$. Indeed, for instance, if there exists another s_2^- such that $a_{s_2^-} = a_i - 1$ then, assuming without loss of generality that $s_2^- < s^-$, by c) $a_i = (a_i - 1) + 1 \in \{a_{s_2^-}, \dots, a_{s^-}\}$ which is a contradiction. $\square$

Just as with elements in $\text{Spec}(n)$ we may define an action of Coxeter generators (transpositions) on elements of $\text{Cont}(n)$ by, letting $\alpha = (a_1, a_2, \dots, a_n) \in \text{Cont}(n)$

$$s_j \cdot \alpha = (a_{s_j(1)}, a_{s_j(2)}, \dots, a_{s_j(n)})$$

Definition 5.4.20. Let $\alpha = (a_1, a_2, \dots, a_n) \in \text{Cont}(n)$, if $a_{i+1} \neq a_i \pm 1$ for some $i \in [n - 1]$ we say s_i is admissible for α .

Note this is analogous to definition 5.4.17. With this we may define an equivalence relation on $\text{Cont}(n)$.

Definition 5.4.21. We define the equivalence relation $\propto$ on $\text{Cont}(n)$ as, for $\alpha, \beta \in \text{Cont}(n)$

$$\alpha \propto \beta \Leftrightarrow \beta \text{ can be obtained from } \alpha \text{ by a sequence of admissible transpositions}$$

Note that given $\alpha \in \text{Cont}(n)$ and s_i for some $i \in [n - 1]$ it's not always the case that $s_i\alpha \in \text{Cont}(n)$, for instance $(0, 1) \in \text{Cont}(2)$ yet $s_1(0, 1) = (1, 0) \notin \text{Cont}(2)$. More generally given $\alpha \in \text{Cont}(n)$ there may well exist $\tau \in S_n$ such that $\tau\alpha \notin \text{Cont}(n)$.

As stated content vector are very much related to standard tableaux, in particular they relate by the underlying Ferrers diagrams. To see this we start with some definitions.

Definition 5.4.22. Let t be a Ferrers diagram and consider a particular box in it with coordinates (i, j) , then the content of the box (i, j) is $j - i$, we denote this as $c(i, j) = j - i$.

Considering this we may also apply the same definition to Young tableaux, it is just the case then that the content of each box depends only on the shape of the tableaux, that is the underlying Ferrers diagram. Yet we'll want to organize the contents of each box in a tuple, so as to relate standard tableaux to content vector, we do this as follows.

Definition 5.4.23. Let t be a Young λ -tableau, with $\lambda \vdash n$, we define the content of t , denoted $c(t)$ to be the n -tuples $(a_1, a_2, \dots, a_n)$ where a_i is the context of the box with the number i in it.

Example 5.4.24. Let $\lambda = (4, 3, 1) \vdash n$ and consider the Young tableau

$$t = \begin{array}{|c|c|c|c|} \hline 1 & 2 & 5 & 7 \\ \hline 3 & 4 & 6 & \\ \hline 8 & & & \\ \hline \end{array}$$

Then we may replace the numbers in each box with the content of each box, resulting in the following diagram

$$\begin{array}{|c|c|c|c|} \hline 0 & 1 & 2 & 3 \\ \hline -1 & 0 & 1 & \\ \hline -2 & & & \\ \hline \end{array}$$

Note this is not a Young tableau. It is then clear that $c(t) = (0, 1, -1, 0, 2, 1, 3, -2)$.

The question now is if the content of a standard tableau is a content vector, seeing this will require us to define the content of a standard tableau in an alternative way.

Recall theorem 5.2.17 and the discussion preceding it. By it we have a bijection between paths in the Young graph and standard tableaux given by, considering a path

$$p = (\lambda^n \rightarrow \lambda^{n-1} \rightarrow \dots \rightarrow \lambda^1)$$

of Ferrers diagrams in the Young graph (where λ^1 is the unique Ferrers diagram corresponding to the unique partition $(1) \vdash 1$), inductively populating the box $\lambda^i / \lambda^{i-1}$ of λ^i with

i. Under this setting we let $c(\lambda^i/\lambda^{i-1})$ be the content of the box λ^i/λ^{i-1} in λ^i . Considering this we can define the content of a standard Young tableau t as

$$c(t) = (c(\lambda^1), c(\lambda^2/\lambda^1), \dots, c(\lambda^n/\lambda^{n-1}))$$

It is clear this is consistent with definition 5.4.23.

Lemma 5.4.25. For $t \in \text{SYT}(n)$, $c(t) \in \text{Cont}(n)$.

Proof. Let t be the standard tableau corresponding to the path

$$\mathbf{p} = (\lambda^n \rightarrow \lambda^{n-1} \rightarrow \dots \rightarrow \lambda^1)$$

in the Young graph. To prove this lemma we need only to prove that $c(t) = (a_1, a_2, \dots, a_n) = (c(\lambda^1), c(\lambda^2/\lambda^1), \dots, c(\lambda^n/\lambda^{n-1}))$ satisfies the condition in definition 5.4.18. For starters it is clear that $c(t) \in \mathbb{Z}^n$, also condition a) is clear as the number 1 must be in the box $(1,1)$ for t to be standard, or in other words $a_1 = c(\lambda^1) = 0$, which also follows clearly from the construction of t mentioned above.

For condition b) note that, given $q \in \{2, \dots, n\}$ such that q is in box (i, j) , so $a_q = j - i$ then as t is standard the numbers in boxes $(i-1, j)$ and $(i, j-1)$ (depending on which is applicable as it's possible one of the two boxes does not exist) are less than q , hence for some $p < q$, p is either in the box $(i-1, j)$ or box $(i, j-1)$, that is $a_p = i - j + 1$ or $a_p = i - j - 1$, hence condition b) follows.

Now for condition c), note that if $a_p = a_q$ for some $p < q$ then p, q must be in the same diagonal of t (going from the top left to the bottom right of the tableau), that is, for instance, we are in a situation of the form

	p					
			q			

Now as t is standard note that the boxes to the right of p have values greater than p and the boxes above of q have values less than q , moreover the boxes below p have values greater than p and the boxes to the left of q have values less than q that is considering

	p	a	b			
	a'		c			
	b'	c'	q			

we have that $p < a < b < c < q$ and $p < a' < b' < c' < q$. Now if p is in box (i, j) then a is in box $(i, j + 1)$ and a' is in box $(i + 1, j)$, thus $a_a = j - i + 1$ and $a'_a = j - i - 1$. Hence c) follows. $\square$

With this in mind we now move to establish a bijection between $\text{SYT}(n)$ and $\text{Cont}(n)$.

Theorem 5.4.26. Consider the map $\Phi : \text{SYT}(n) \rightarrow \text{Cont}(n)$ define by, for a standard tableau $t \in \text{SYT}(n)$ with respective path in the Young graph

$$p = (\lambda^n \rightarrow \lambda^{n-1} \rightarrow \dots \rightarrow \lambda^1)$$

letting

$$t \mapsto (c(\lambda^1), c(\lambda^2/\lambda^1), \dots, c(\lambda^n/\lambda^{n-1}))$$

Then Φ is a bijection, moreover it maps $\approx$ -equivalent tableaux to $\propto$ -equivalent content vectors, that is for $t_1, t_2 \in \text{SYT}(n)$

$$t_1 \approx t_2 \Rightarrow \Phi(t_1) \propto \Phi(t_2)$$

Proof. By the previous lemma we know Φ is well defined. To prove it's injective let then $c(t_1) = c(t_2)$ for $t_1, t_2 \in \text{SYT}(n)$, thus

$$(c(\lambda^1), c(\lambda^2/\lambda^1), \dots, c(\lambda^n/\lambda^{n-1})) = (c(\mu^1), c(\mu^2/\mu^1), \dots, c(\mu^n/\mu^{n-1}))$$

Note that the only situation in which two numbers $p < q$ could be assigned to different boxes in t_1 and t_2 is if they appear in the same diagonal but in different places, like for instance

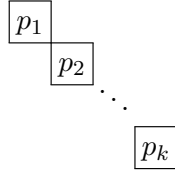
$$t_1 = \begin{array}{|c|c|c|c|c|c|c|} \hline & & & & & & \\ \hline & p & & & & & \\ \hline & & & & & & \\ \hline & & & & & & \\ \hline & & & q & & & \\ \hline & & & & & & \\ \hline \end{array} ; \quad t_2 = \begin{array}{|c|c|c|c|c|c|c|} \hline & & & & & & \\ \hline & p & & & & & \\ \hline & & q & & & & \\ \hline & & & & & & \\ \hline & & & & & & \\ \hline & & & & & & \\ \hline \end{array}$$

Or are simply permuted in that diagonal like

$$t_2 = \begin{array}{|c|c|c|c|c|c|c|} \hline & & & & & & \\ \hline & q & & & & & \\ \hline & & & & & & \\ \hline & & & p & & & \\ \hline & & & & & & \\ \hline \end{array}$$

This latter situation is not possible as by t_1 being standard it must be the case that $p < q$ so t_2 couldn't possibly be a standard tableau, so we may focus on the first situation. If we

consider a full diagonal of t_1



then by t being standard we must have $p_1 < p_2 < \dots < p_k$, and $a_{p_1} = a_{p_2} = \dots = a_{p_k}$. So first note that if $c(t_1) = c(t_2)$ then the numbers that appear in the diagonal containing p, q in t_1 and t_2 must be the same, and then considering this it must be the case that in t_2 we have permuted, say q , with some p_i in that diagonal, hence we get a contradiction of the order $p_1 < p_2 < \dots < p_k$ that is required for t_2 to be standard, or in other words we fall into the previous case of permuted entries.

So in conclusion if $c(t_1) = c(t_2)$ then $t_1 = t_2$ and thus Φ is injective.

Now for surjectivity we use induction on n to show that any content vector gives the construction of a standard tableau. Let $\alpha = (a_1, a_2, \dots, a_n) \in \text{Cont}(n)$, if $n = 1$ then $\alpha = (0)$ by condition a) of definition 5.4.18, thus we get the unique tableau of $\lambda = (1) \vdash 1$ which is trivially standard. Assuming now the assumption is valid for $k \leq n$ let $\alpha' = (a_1, a_2, \dots, a_{n+1}) \in \text{Cont}(n+1)$, we note that $\alpha = (a_1, a_2, \dots, a_n) \in \text{Cont}(n)$ and so it gives a standard tableau $t \in \text{SYT}(n)$, so we need only to see where the box containing $n+1$ is added to t . By condition b') in proposition 5.4.19 if $a_{n+1} \geq 0$ then $a_{n+1} = a_p + 1$ for some $p < n+1$ and if $a_{n+1} \leq 0$ then $a_{n+1} = a_p - 1$ for some $p < n+1$, we can assume there is no other a_q with $p < q < n+1$ such that $a_p = a_q$ (by simply choosing the last such a_p in α) thus letting $a_p = j - i$

i) If $a_{n+1} = \alpha_p + 1$ then $n+1$ may go to box $(i, j+1)$.

ii) If $a_{n+1} = \alpha_p - 1$ then $n+1$ may go to box $(i+1, j)$.

As $n+1$ is bigger than any value in the boxes of t this is a question of those boxes being addable to t , depending on the situation. We first prove that p is an fortuitous location in t , that is the box $(i, j+1)$ or $(i+1, j)$ don't already exist in t .

If $a_{n+1} = \alpha_p + 1$ and the box $(i, j+1)$ already exists then let s be in box $(i, j+1)$, so we have $s > p$, but as $a_s = a_p + 1 = a_{n+1}$ then by condition c) of definition 5.4.18, $a_p \in (a_{s+1}, \dots, a_n)$, but this contradicts the non existence of other a_q with $p < q < n+1$ such that $a_p = a_q$. The case for $a_{n+1} = \alpha_p - 1$ and box $(i+1, j)$ is analogous. So in conclusion those boxes are available to be added, but now we need to prove that they

are indeed addable, that is depending on the situation, once added we get a Young tableau (or in particular a Ferrers diagram).

If $a_{n+1} = \alpha_p + 1$ then as we've seen the box $(i, j + 1)$ is not in t , if it is the case that the box $(i - 1, j)$ is also not in t then we must be in the first row of t and so the box $(i, j + 1)$ is addable, if the box $(i - 1, j)$ exists in t then we have the following sub diagram of t

s	
p	$n+1$

with $s < p < n + 1$ where the missing box above $n + 1$ is the box $(i - 1, j + 1)$ we need to prove it exists in order for $(i, j + 1)$ to be addable. We note that $a_s = a_p + 1 = a_{n+1}$ hence by condition c), $a_r = a_p + 2 \in (a_{s+1}, \dots, a_p, \dots, a_n)$. Now as by this $r > s$ and $a_r = j + 1 - (i - 1)$ it must be the case that r is in the box $(i - 1, j + 1)$ as there are no other possibilities. So the box $(i - 1, j + 1)$ exists in t and so $(i, j + 1)$ is addable. Again the other case is analogous.

So we can conclude that Φ is indeed bijective, so now we need only to prove that Φ maps $\approx$ -equivalent tableaux to $\propto$ -equivalent content vectors. We prove that if a transposition s_i is admissible for $t \in \text{SYT}(n)$ as in definition 5.2.21 then s_i is admissible for $\Phi(t) \in \text{Cont}(n)$ as in definition 5.4.20.

Consider then a young tableau t and a transposition s_i , if in the tableau t , $a_i = a_i \pm 1$ where a_{i+1}, a_i are the content of the boxes with $i + 1$ and i in them respectively, then letting $a_i = j - i$ we have that $a_{i+1} = j - i \pm 1$, so seeing this more visually, for instance if we let $a_{i+1} = j - i + 1$, $i + 1$ can be in one of the following boxes marked by $\times$

	i	$\times$				
			$\times$			
				$\times$		

That is it can be immediately to the right of i or going down along that diagonal. Considering this it is clear that if we permute i and $i + 1$ we won't get a standard tableau. The case where $a_{i+1} = j - i - 1$ is analogous. □

5.4.5 The spectrum of the YJM elements

As stated in the previous subsection we'll prove the bijection between $\text{Spec}(n)$ and $\text{SYT}(n)$ in two steps, using $\text{Cont}(n)$ as an intermediary. A first step has been achieved in theorem

5.4.26, we are now left to prove the bijection between $\text{Spec}(n)$ and $\text{Cont}(n)$. As it turns out these sets are in fact the very same.

Lemma 5.4.27. *For $n \in \mathbb{N}$, $\text{Spec}(n) \subseteq \text{Cont}(n)$.*

Proof. We prove this by induction on n . For $n = 1$ as $X_1 = 0$ we get $\text{Spec}(1) = \text{Cont}(1)$, so assume that for $k < n$, $\text{Spec}(k) \subseteq \text{Cont}(k)$ and consider $\text{Spec}(n)$. We have to prove that for $\alpha = (a_1, a_2, \dots, a_n) \in \text{Spec}(n)$, $\alpha \in \text{Cont}(n)$, that is α satisfies conditions a), b) and c) of definition 5.4.18. Condition a) follows easily as regardless of n , $X_1 = 0$ and thus $a_1 = 0$.

For condition b) assume, for a contradiction, that

$$\{a_n - 1, a_n + 1\} \cap \{a_1, a_2, \dots, a_{n-1}\} = \emptyset$$

Then by part d) of theorem 5.4.16, s_{n-1} is an admissible transposition for α , hence

$$(a_1, a_2, \dots, a_{n-2}, a_n, a_{n-1}) \in \text{Spec}(n)$$

And so by induction hypothesis $(a_1, a_2, \dots, a_{n-2}, a_n) \in \text{Cont}(n-1)$, but by our assumption $\{a_n - 1, a_n + 1\} \cap \{a_1, a_2, \dots, a_{n-1}\} = \emptyset$, thus $\{a_n - 1, a_n + 1\} \cap \{a_1, a_2, \dots, a_{n-2}\} = \emptyset$ which contradicts $(a_1, a_2, \dots, a_{n-2}, a_n) \in \text{Cont}(n-1)$ by contradicting condition b) of definition 5.4.18. Note then that as $\{a_n - 1, a_n + 1\} \cap \{a_1, a_2, \dots, a_{n-1}\} \neq \emptyset$ then $a_n - 1$ or $a_n + 1$ is in $(a_1, a_2, \dots, a_{n-1}) \in \text{Cont}(n-1)$, thus $a_n \in \mathbb{Z}$.

Now for condition c) let $a_p = a_n$ with $p < n$ and such that p is maximal for $\{a_n - 1, a_n + 1\} \not\subseteq \{a_{p+1}, \dots, a_{n-1}\}$, note that in this case $a_p \notin \{a_{p+1}, \dots, a_{n-1}\}$, moreover $a_n + 1$ or $a_n - 1$ can only appear in $\{a_{p+1}, \dots, a_{n-1}\}$ at most once, as by $\{a_1, \dots, a_{n-1}\} \in \text{Cont}(n)$ then by condition c) we'd have $a_n = a_p \in \{a_{p+1}, \dots, a_{n-1}\}$ which contradicts the maximality of p . Now if $a_n + 1$ is in $\{a_{p+1}, \dots, a_{n-1}\}$ we have

$$(a_p, \dots, a_n + 1, \dots, a_n)$$

Using part d) of theorem 5.4.16 we may apply a sequence of admissible transpositions to α producing then

$$\alpha' = (\dots, a_p, a_n + 1, a_n, \dots)$$

But as then $\alpha' \in \text{Spec}(n)$ and $a_p = a_n$ we get a contradiction of part c) of theorem 5.4.16. We may do the same for the case where $a_n - 1$ is in $\{a_{p+1}, \dots, a_{n-1}\}$. Thus it

can't be the case that α doesn't satisfy condition c) of definition 5.4.18. So in conclusion $\alpha \in \text{Cont}(n)$. $\square$

Recall now the equivalence relations in definitions 5.2.27, 5.4.12 and 5.4.21.

Proposition 5.4.28. *Considering the equivalence relations mentioned above*

$$|\text{SYT}(n)/\approx| = |\text{Cont}(n)/\propto| = |\text{Spec}(n)/\smile|$$

Proof. By theorem 5.4.26, $|\text{SYT}(n)| = |\text{Cont}(n)|$ and if $t_1 \approx t_2$ for $t_1, t_2 \in \text{SYT}(n)$ then $c(t_1) \propto c(t_2)$, hence $|\text{Cont}(n)/\propto| \leq |\text{SYT}(n)/\approx| = |\widehat{S}_n|$. Moreover by the previous lemma we have that $\text{Spec}(n) \subseteq \text{Cont}(n)$ and by part d) of theorem 5.4.16 if $\alpha \propto \beta$ for $\alpha \in \text{Spec}(n)$ and $\beta \in \text{Cont}(n)$ then $\alpha \smile \beta$, and in particular $\beta \in \text{Spec}(n)$, hence letting $[\alpha] \in \text{Spec}(n)/\smile$ and $[\beta] \in \text{Cont}(n)/\propto$ then either $[\beta] \subseteq [\alpha]$ or $[\alpha] \cap [\beta] = \emptyset$, therefore $|\text{Spec}(n)/\smile| \leq |\text{Cont}(n)/\propto|$, but as we know $|\text{Spec}(n)/\smile| = |\widehat{S}_n|$, thus

$$|\widehat{S}_n| \leq |\text{Cont}(n)/\propto| \leq |\widehat{S}_n|$$

Hence $|\text{Cont}(n)/\propto| = |\widehat{S}_n|$ and so $|\text{SYT}(n)/\approx| = |\text{Cont}(n)/\propto| = |\text{Spec}(n)/\smile|$. $\square$

With this the results we seek come easily.

Proposition 5.4.29. *For $n \in \mathbb{N}$, $\text{Spec}(n) = \text{Cont}(n)$, moreover the equivalence relations $\smile$ and $\propto$ coincide.*

Proof. Indeed by the previous lemma $\text{Spec}(n) \subseteq \text{Cont}(n)$, moreover by the previous proposition $|\text{Cont}(n)/\propto| = |\text{Spec}(n)/\smile|$ and for $[\alpha] \in \text{Spec}(n)/\smile$, $[\beta] \in \text{Cont}(n)/\propto$ either $[\beta] \subseteq [\alpha]$ or $[\alpha] \cap [\beta] = \emptyset$. Considering this, by the pigeon hole principle, one concludes that $\text{Spec}(n) = \text{Cont}(n)$ and the equivalence classes in each set, corresponding to the relations $\smile$ and $\propto$ respectively, are the same, hence the relations are the same. $\square$

Theorem 5.4.30. *The map Φ in theorem 5.4.26 is a bijection between $\text{SYT}(n)$ and $\text{Spec}(n)$, moreover it preserves the relations $\approx$ and $\smile$, that is for $t_1, t_2 \in \text{SYT}(n)$*

$$t_1 \approx t_2 \Leftrightarrow \Phi(t_1) \smile \Phi(t_2)$$

Proof. That it is a bijection comes immediately from theorem 5.4.26 and the previous proposition. Moreover by those same results we have that Φ sends $\approx$ -equivalent tableaux

to $\sim$ -equivalent elements of $\text{Spec}(n)$, now by proposition 5.4.28

$$|\text{SYT}(n)/\approx| = |\text{Spec}(n)/\sim|$$

hence, again by the pigeon hole principle, Φ also gives a bijection between $\text{SYT}(n)/\approx$ and $\text{Spec}(n)/\sim$ such that for $t_1, t_2 \in \text{SYT}(n)$

$$[t_1] = [t_2] \Leftrightarrow [\Phi(t_1)] = [\Phi(t_2)]$$

□

As stated in Section 5.4, after definition 5.4.11, we have a bijection between Young vectors and $\text{Spec}(n)$, which can also be seen as bijection between paths starting at level n of the branching graph and $\text{Spec}(n)$ as any Young vector is associated uniquely with such a path. So the map Φ above also gives a bijection between Young vectors and standard Young tableaux. With this we can finally make good on our promise that the Young graph given in Section 5.2 and the branching graph of symmetric groups are the same graph. Recall that two graphs $\Lambda = (V, E)$, $\Lambda' = (V', E')$, where V, V' are sets of vertices and E, E' sets of edges, are isomorphic if there exists a bijective map $\Psi : V \rightarrow V'$ such that $(v_1, v_2) \in E$ if and only if $(\Psi(v_1), \Psi(v_2)) \in E'$.

Theorem 5.4.31. *The branching graph of the chain of symmetric groups and the Young graph are isomorphic graphs.*

Proof. First note that each equivalence class in $\text{SYT}(n)/\approx$ is uniquely associated to a partition of n , by the shape of the standard tableaux in that class, or in other words the underlying Ferrers diagram, this comes from corollary 5.2.26 and the discussion thereafter. Hence we have a natural bijective map from $\text{SYT}(n)/\approx$ to $\{\lambda \vdash n\}$, and as $\{\lambda \vdash n\}$ are the vertices at the level n of the Young graph we have by extension a natural bijection from the vertices of the Young graph $\{\lambda \vdash n \mid n \in \mathbb{N}\}$ to $\{\text{SYT}(n)/\approx \mid n \in \mathbb{N}\}$, call it ϕ .

On the other hand we also clearly have a natural bijection between $\text{Spec}(n)/\sim$ and vertices at the level n of the branching graph, times comes immediately from definition 5.4.12. So by extension we have a natural bijection from $\{\text{Spec}(n)/\sim \mid n \in \mathbb{N}\}$ to the vertices of the branching graph $\{\hat{S}_n \mid n \in \mathbb{N}\}$, call it ψ .

Now as stated in the proof of the previous theorem the map Φ therein gives a bijection between $\text{SYT}(n)/\approx$ and $\text{Spec}(n)/\sim$ such that for $t_1, t_2 \in \text{SYT}(n)$, $[t_1] = [t_2] \Leftrightarrow [\Phi(t_1)] =$

$[\Phi(t_2)]$. Again by extension we have a natural bijection from $\{\text{SYT}(n)/\approx \mid n \in \mathbb{N}\}$ to $\{\text{Spec}(n)/\sim \mid n \in \mathbb{N}\}$ that preserves the relations $\approx$ and $\sim$, call it Φ' .

So the composition $\Psi = \psi\Phi'\phi$ is a bijection between vertices of the branching graph and vertices of the Young graph, now we need to prove it preserves edges. Let $(\lambda^n, \lambda^{n-1})$ be an edge in the Young graph and consider some path in that graph

$$\mathbf{p} = (\lambda^n \rightarrow \lambda^{n-1} \rightarrow \dots \rightarrow \lambda^1)$$

This path produces a unique standard tableau t , moreover if we consider the subpath

$$\mathbf{p}' = (\lambda^{n-1} \rightarrow \dots \rightarrow \lambda^1)$$

then we also produce another standard tableau t' , that will be a sub-tableau of t . Now considering the explicit bijection Φ given in theorem 5.4.26 we can see that if $\Phi(t) = (a_1, a_2, \dots, a_n)$ then $\Phi(t') = (a_1, a_2, \dots, a_{n-1})$. Letting $\alpha = \Phi(t)$ and $\alpha' = \Phi(t')$ then by part b) of corollary 5.3.10 the Young vectors v_α and $v_{\alpha'}$ belong to a same irreducible representation ρ^{n-1} on level $n-1$, moreover v_α also belongs to an irreducible representation ρ^n on level n , hence (ρ^n, ρ^{n-1}) is an edge on the branching graph. Now by this we have that $\psi([\Phi(t)]) = \rho^n$ and $\psi([\Phi(t')]) = \rho^{n-1}$ hence $\Psi(\lambda^{n-1}) = \rho^{n-1}$ and $\Psi(\lambda^n) = \rho^n$. Proving that if (ρ^n, ρ^{n-1}) is an edge in the branching graph then $(\Psi^{-1}(\rho^n), \Psi^{-1}(\rho^{n-1}))$ is an edge in the Young graph is done similarly, and so the conclusion follows. $\square$

With the two theorem above we have the following result.

Corollary 5.4.32. *Given an irreducible representation ρ of S_n then the dimension of ρ is equal to the number of standard Young tableaux of shape $\Psi^{-1}(\rho)$, where Ψ is the graph isomorphism in the previous theorem.*

Recall now the Specht modules S^λ , for $\lambda \vdash n$, given in theorem 5.2.11, as stated these modules are precisely the irreducible representations of S_n .

Theorem 5.4.33. *For $\lambda \vdash n$ the Specht module S^λ and the irreducible S_n -space corresponding to the representation $\Psi(\lambda)$ are isomorphic S_n -spaces.*

Proof. This follows by induction on n . Indeed for $n = 1$, one may easily verify that S^λ is the trivial representation of the trivial group, as needed. If we now assume the statement is true for all $k < n$ then given $\lambda \vdash n$ and the respective Specht module S^λ , by the Branching

rule in theorem 5.2.18

$$S^\lambda \downarrow_{S_{n-1}} = \bigoplus_{\lambda \rightarrow \mu} S^\mu$$

But by the induction hypotheses and the previous theorem this becomes

$$S^\lambda \downarrow_{S_{n-1}} = \bigoplus_{\Psi(\lambda) \rightarrow \Psi(\mu)} V_{\Psi(\mu)}$$

Where $V_{\Psi(\mu)}$ is the irreducible S_{n-1} -space associated with the irreducible representation $\Psi(\mu)$, hence S^λ and $V_{\Psi(\lambda)}$ have the same decomposition into irreducible S_{n-1} -spaces, thus S^λ and $V_{\Psi(\lambda)}$ are isomorphic S_{n-1} -spaces. To prove that S^λ and $V_{\Psi(\lambda)}$ are the same irreducible S_n -space we first analyze the case where $n = 2$ and then the case where $n > 2$, so from now on let $\rho = \Psi(\lambda)$.

If $n = 2$ let $\lambda = (1, 1)$, in this case both spaces have just a single basis element corresponding to the unique standard λ -tableau, call that tableau t ; S^λ has as basis vector the polytabloid e_t , as in definition 5.2.8, and V_ρ has basis vector v_α where $\alpha = (0, -1)$. We now prove S_2 acts the same way on both these vector, indeed from part b) of theorem 5.4.16 we have that $s_1 v_\alpha = -v_\alpha$. Now as $s_1 = (12)$ acts on t by only permuting the elements of a column then $s_1 \in C_t$, with C_t as in definition 5.2.7, as such

$$s_1 e_t = \sum_{\pi \in C_t} \text{sgn}(\pi) s_1 \pi \{t\} = \sum_{s_1 \pi \in C_t} \text{sgn}(s_1 \pi) \pi \{t\} = \text{sgn}(s_1) e_t = -e_t$$

With this we can produce the S_2 -isomorphism $\Theta : V_\rho \rightarrow S^\lambda$ such that $\Theta(v_\alpha) = e_t$, hence S^λ and V_ρ are S_2 -isomorphic and so the other Specht module $S^{\lambda'}$ where $\lambda' = (2) \vdash 2$ can only correspond to $V_{\Psi(\lambda')}$.

Now for $n > 2$ we prove that no two non-equivalent irreducible representations of S_n , when restricted to S_{n-1} , can have the same decomposition into irreducible S_{n-1} -spaces, hence as S^λ and V_ρ have the same decomposition into irreducible S_{n-1} -spaces, as we've seen above, they must be the same S_n -space. We do this by looking at the Young graph and proving that no two distinct Ferrers diagrams at level n can cover (that is have an edge to) the same Ferrers diagrams at level $n - 1$. Indeed let λ and λ' be two Ferrers diagram at level n both covering some μ_1 and μ_2 as in figure 5.3 (here we are not saying that λ and λ' cover just these two, we are saying that they cover at least these two and might well cover more equal Ferrers diagrams). Then of course λ has just one more box than μ_1 and μ_2 , moreover μ_1 and μ_2 are sub-diagrams of λ and μ_1 and μ_2 differ by the position of just a single box, thus if we add to μ_1 the box in μ_2 that differs in position we

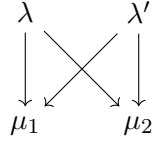


FIGURE 5.3: Two partitions λ, λ' at level n both covering some μ_1 and μ_2 at level $n - 1$.

must get the diagram λ , applying the same reasoning to λ' we get that it must be the case that $\lambda = \lambda'$. Now if it is the case that λ and λ' happen to only cover a single Ferrers diagram μ at level $n - 1$ like in figure 5.4, then as $n > 2$, λ and λ' have at least 3 boxes, and as they are distinct one can easily see that at least one of them must cover more than just a single Ferrers diagram at level $n - 1$ hence this case is not possible. So in

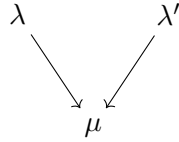


FIGURE 5.4: Two partitions λ, λ' at level n covering a single partition μ at level $n - 1$.

conclusion by theorem 5.4.31, for $n > 2$, no two irreducible representations of S_n , when restricted to S_{n-1} , can have the same decomposition into irreducible S_{n-1} -spaces. $\square$

By the above theorem we don't need to distinguish between the Specht module S^λ and $V_{\Psi(\lambda)}$, for this from now on we let V_λ be the irreducible S_n -space spanned by the Young vectors associated to $\Psi(\lambda)$.

With theorem 5.4.31 we can obtain an upper bound for the multiplicities of irreducibles of S_k in the decomposition of an irreducible of S_n restricted to S_k (so $k < n$). Recall definition 5.3.4, we now adopt the same notations for the Young graph, that is if there exists a path from $\lambda \vdash n$ to $\mu \vdash k$, with $k < n$, in the Young graph we say $\lambda \searrow \mu$, and as before if (λ, μ) is an edge in the Young graph then we say $\lambda \rightarrow \mu$.

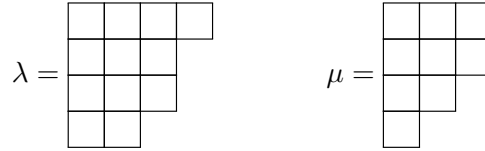
Proposition 5.4.34. *Let $k \in [n - 1]$, $\lambda \vdash n$ and $\mu \vdash k$, then the multiplicity of V_μ in $V_\lambda|_{S_k}$, denoted $m_{\lambda, \mu}$, satisfies $m_{\lambda, \mu} \leq (n - k)!$, and if it is not the case that $\lambda \searrow \mu$ then $m_{\lambda, \mu} = 0$. Moreover this bound is sharp.*

Proof. Indeed by the corollary 5.3.6 we have that $m_{\lambda, \mu}$ is equal to the number of path in the branching graph from λ to μ , that is $m_{\lambda, \mu} = |\mathfrak{P}_\mu^\lambda|$. Now looking at the Young graph this means that $m_{\lambda, \mu}$ is equal to the number of ways of getting the Ferrers diagram of λ from

the one of μ by inductively adding $n - k$ (addable) boxes to μ . Considering this there are at most $(n - k)!$ ways of adding the $(n - k)$ boxes of λ not in μ to μ to obtain λ . $\square$

To see the bound above is indeed sharp we look at the following example.

Example 5.4.35. Consider the following Ferrers diagrams:



With $\mathfrak{P}_\mu^\lambda$ they produce the subgraph of the Young graph in figure 5.5. One notes that there

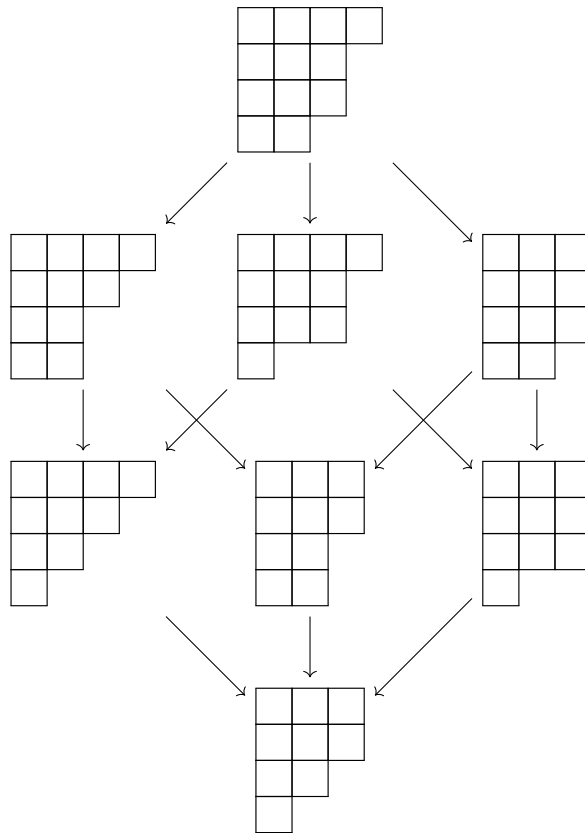


FIGURE 5.5: Subgraph of the Young graph produces by considering two partitions λ, μ with $\lambda \searrow \mu$.

are $6 = (12 - 9)!$ paths from λ to μ , thus the bound in the above proposition is indeed sharp, moreover it is clear that the equality occurs when there are no restrictions to the order in which the boxes can be added to μ , that is any order will produce a sequence of Ferrers diagrams.

Before moving on to the next subsection we provide here a final example exploiting theorem 5.4.31 and proposition 5.4.34.

Example 5.4.36. Let $\lambda \vdash n$ and $\mu \vdash n - 2$, by proposition 5.4.34, $m_{\lambda, \mu} \leq 2$, assuming that $\lambda \searrow \mu$ then there are two possible cases:

- There exists a unique partition $\nu \vdash n - 1$ such that $\lambda \rightarrow \nu \rightarrow \mu$ in the Young graph.

In this case the boxes in λ/μ must be all in the same row or the same column, otherwise ν wouldn't be unique. As such for $\alpha = (a_1, \dots, a_n) \in \text{Spec}(n)$ corresponding to a path $p = (\lambda \rightarrow \nu \rightarrow \mu \rightarrow \lambda^{n-3} \rightarrow \dots \rightarrow \lambda^1)$ it must be the case that $a_n = a_{n-1} \pm 1$, hence by theorem 5.4.16, $s_{n-1}v_\alpha = \pm v_\alpha$. Note this is consistent with proposition 5.4.13.

- There exist two partitions $\nu_1, \nu_2 \vdash n - 1$ such that $\lambda \rightarrow \nu_i \rightarrow \mu$ for $i \in [2]$, as in figure 5.6.

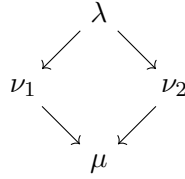


FIGURE 5.6: Two partitions $\nu_1, \nu_2 \vdash n - 1$ such that $\lambda \rightarrow \nu_i \rightarrow \mu$ for $i \in [2]$.

In this case the boxes in λ/μ are not in the same row or column, and as a result for $\alpha = (a_1, \dots, a_n)$ in $\text{Spec}(n)$ corresponding to a path $p = (\lambda \rightarrow \nu_1 \rightarrow \mu \rightarrow \lambda^{n-3} \rightarrow \dots \rightarrow \lambda^1)$, $a_n \neq a_{n-1} \pm 1$. Considering this, by theorem 5.4.16

$$s_{n-1}v_\alpha = \beta v_{\alpha'} - \frac{1}{a_n - a_{n-1}}v_\alpha$$

Where $\beta \in \mathbb{C}$ is some scalar factor and $\alpha' = (a_1, \dots, a_{n-2}, a_n, a_{n-1}) \in \text{Spec}(n)$ correspond to the path $p' = (\lambda \rightarrow \nu_2 \rightarrow \mu \rightarrow \lambda^{n-3} \rightarrow \dots \rightarrow \lambda^1)$, so again we have consistency with proposition 5.4.13.

5.4.6 Young's seminormal and orthogonal forms

By now we have a very good understanding of the vector spaces of irreducible representations of symmetric groups, and a strong link between the classical and modern approach to the representation theory of these groups, yet we are still lacking in how exactly they

act on those vector spaces. For this we now present A. Young's *Seminormal and Orthogonal representations*, they will follow greatly from our work in Subsection 5.4.3, as such the exposition here is closer to the work of G.E. Murphy [Mur81].

To understand how a symmetric group acts on the aforementioned vector spaces we focus on the Coxeter generators and the Young basis, in particular we'll choose the scalar factors of the Young basis to our advantage, so that we may determine how the Coxeter generators acts on this basis, thus fully determining the irreducible representations.

As noted at the end of Subsection 5.4.2, we may index Young vector by paths in the branching graph, as v_p for some path p , yet by theorem 5.4.31 the branching graph and the Young graph are the same, hence we may also consider the path p in the Young graph. As this path uniquely determines a standard tableau we may also index Young vectors by standard tableaux, so if path p produces the standard tableau t we can denote v_p by v_t . Recall also proposition 5.2.25, as per it given a standard tableau t we denote the unique permutation mapping t to t_s by σ_t , and this permutation decomposes as a product of $l(\sigma_t)$ admissible transpositions for t , as in definition 5.2.21.

We now explore the advantages of choosing scalar factor for the Young basis. By this we mean that given a Young basis $\{v_t \mid t \in \text{SYT}(n)\}$, we multiply each Young vector v_t by a scalar factor $\alpha_t \in \mathbb{C}$ so as to produce another Young basis $\{\alpha_t v_t \mid t \in \text{SYT}(n)\}$, with useful properties.

Proposition 5.4.37. *Given a Young basis $\{v'_t \mid t \in \text{SYT}(n)\}$ there exists scalar factors $\{\alpha_t \in \mathbb{C} \mid t \in \text{SYT}(n)\}$ such that for the Young basis $\{v_t = \alpha_t v'_t \mid t \in \text{SYT}(n)\}$, given $t' \in \text{SYT}(\lambda)$ for some $\lambda \vdash n$, with corresponding permutation $\sigma_{t'}$ as in proposition 5.2.25, one has*

$$\sigma_{t'}^{-1} v_{t_s^\lambda} = v_{t'} + \sum_{\substack{r \in \text{SYT}(\lambda), \\ l(\sigma_r) < l(\sigma_{t'})}} \beta_r v_r$$

With $\beta_r \in \mathbb{C}$.

Proof. We prove this by induction on $l(\sigma_{t'})$. First note that $\sigma_{t_s^\lambda}^{-1} = 1_{S_n}$ hence $\sigma_{t_s^\lambda}^{-1} v'_{t_s^\lambda} = v'_{t_s^\lambda}$ so let $\alpha_{t_s^\lambda} = 1$. If $l(\sigma_{t'}) = 1$ then $\sigma_{t'}^{-1}$ is an admissible transposition for t_s^λ as it sends a standard tableau to a standard tableau, thus by proposition 5.4.29 and theorem 5.4.30, $\sigma_{t'}^{-1}$ (which is equal to $\sigma_{t'}$ by being a transposition) is an admissible transposition for $v_{t_s^\lambda}$. So by part d) of theorem 5.4.16 (letting $\sigma_{t'} = s_i$) we have that

$$\sigma_{t'}^{-1} v_{t_s^\lambda} = \alpha v_{t_s} + \frac{1}{a_{i+1} - a_i} v_{t_s^\lambda}$$

where a_{i+1} and a_i are elements of the spectrum of $v_{t_s^\lambda}$, $s \in \text{Spec}(n)$ and $\alpha \in \mathbb{C}$. As $\sigma_{t_s^\lambda} = 1_{S_n}$ then $l(\sigma_{t_s^\lambda}) < l(\sigma_{t'})$ so we have that $\beta_{t_s^\lambda} = 1/(a_{i+1} - a_i)$. Now we need only to check that $v_s = v_{t'}$, and we then let $\alpha_{t'} = \alpha$. But looking at the content vector of $v_{t_s^\lambda}$ one easily sees that $\sigma_{t'}$, being a transposition, will act on it by sending it to the content vector of $v_{t'}$, thus the conclusion follows.

Now if $l(\sigma_{t'}) > 1$ let $\sigma_{t'} = s_{i_1} s_{i_2} \cdots s_{i_k}$ be the decomposition of $\sigma_{t'}$ into a sequence of admissible transpositions per proposition 5.2.25, then $t'_1 = s_{i_k} t'$ is a standard λ -tableau and $\sigma_{t'} = \sigma_{t_1} s_{i_k}$. As $l(\sigma_{t_1}) = l(\sigma_{t'}) - 1$, by the induction hypothesis

$$s_{i_k} \sigma_{t'}^{-1} v_{t_s^\lambda} = \sigma_{t_1}^{-1} v_{t_s^\lambda} = v_{t_1} + \sum_{\substack{r \in \text{SYT}(\lambda), \\ l(\sigma_r) < l(\sigma_{t_1})}} \beta'_r v_r \quad (5.13)$$

But since $t' = s_{i_k} t_1$, letting $s_{i_k} = s_i$, we have as before, by part d) of theorem 5.4.16, that

$$s_i v_{t_1} = \alpha v_{t'} + \frac{1}{a_{i+1} - a_i} v_{t_1} \quad (5.14)$$

where a_{i+1} and a_i are elements of the spectrum of v_{t_1} and $\alpha \in \mathbb{C}$. Hence by 5.13 and 5.14

$$\sigma_{t'}^{-1} v_{t_s^\lambda} = s_j \sigma_{t_1}^{-1} v_{t_s^\lambda} = \alpha v_{t'} + \frac{1}{a_{i+1} - a_i} v_{t_1} + \sum_{\substack{r \in \text{SYT}(\lambda), \\ l(\sigma_r) < l(\sigma_{t_1})}} \beta'_r s_j v_r$$

Finally as $l(\sigma_{t_1}) = l(\sigma_{t'}) - 1$, with another use of theorem 5.4.16 for the $s_j v_r$ in the above sum, noting that by part b) and d) this is at most a sum of two Young vectors indexed by standard tableaux r that satisfy $l(\sigma_r) < l(\sigma_{t'})$, the result follows immediately, with then $\alpha_{t'} = \alpha$. □

With the Young basis $\{v_t = \alpha_t v'_t \mid t \in \text{SYT}(n)\}$ in the above proposition we can now determine the action of the Coxeter generators on this basis. This gives rise to what is called *Young's Seminormal form* or *Young's Seminormal representation*.

Theorem 5.4.38. *Considering the Young basis $\{v_t = \alpha_t v'_t \mid t \in \text{SYT}(n)\}$ in the previous proposition, for a standard tableau $t \in \text{SYT}(\lambda)$, with $\lambda \vdash n$, and with content vector $c(t) = (a_1, \dots, a_n)$, a Coxeter generators s_i , $i \in [n - 1]$, acts on v_t as*

- a) *If $a_{i+1} = a_i \pm 1$ then $s_i v_t = \pm v_t$.*

b) If $a_{i+1} \neq a_i \pm 1$, then letting $t' = s_i t$

$$s_i v_t = \begin{cases} \frac{1}{a_{i+1}-a_i} v_t + v_{t'} & \text{if } l(\sigma_{t'}) > l(\sigma_t) \\ \frac{1}{a_{i+1}-a_i} v_t + \left(1 - \frac{1}{(a_{i+1}-a_i)^2}\right) v_{t'} & \text{if } l(\sigma_{t'}) < l(\sigma_t) \end{cases} \quad (5.15)$$

Proof. Case a) follows immediately from part b) of theorem 5.4.16, for case b) we use part d) of theorem 5.4.16, but here we need to check what the scalar factor of $v_{t'}$ will be. We first note that indeed the Young vector called $v_{\alpha'}$ in part d) of theorem 5.4.16, in our context, is the Young vector $v_{t'}$. This is analogous to what we've seen in the previous proposition, as the action of s_i on the content vector of v_t gives the content vector of $v_{t'}$.

If $l(\sigma_{t'}) > l(\sigma_t)$ then as $t' = s_j t$ we have that $\sigma_{t'} = \sigma_t s_i$. Now by proposition 5.4.37 it is the case that

$$\sigma_t^{-1} v_{t_s^\lambda} = v_t + \sum_{\substack{r \in \text{SYT}(\lambda), \\ l(\sigma_r) < l(\sigma_t)}} \beta_r v_r \quad (5.16)$$

and

$$\sigma_{t'}^{-1} v_{t_s^\lambda} = v_{t'} + \sum_{\substack{r \in \text{SYT}(\lambda), \\ l(\sigma_r) < l(\sigma_{t'})}} \beta'_r v_r$$

Moreover from 5.16 and $\sigma_{t'} = \sigma_t s_i$ we also have that

$$\sigma_{t'}^{-1} v_{t_s^\lambda} = s_i \sigma_t^{-1} v_{t_s^\lambda} = s_i v_t + s_i \sum_{\substack{r \in \text{SYT}(\lambda), \\ l(\sigma_r) < l(\sigma_t)}} \beta_r v_r$$

Hence

$$v_{t'} + \sum_{\substack{r \in \text{SYT}(\lambda), \\ l(\sigma_r) < l(\sigma_{t'})}} \beta'_r v_r = s_i v_t + s_i \sum_{\substack{r \in \text{SYT}(\lambda), \\ l(\sigma_r) < l(\sigma_t)}} \beta_r v_r$$

From this we have that the coefficient of $v_{t'}$ in $s_i v_t$ must be 1.

For the case where $l(\sigma_{t'}) < l(\sigma_t)$ we start by noting that $\sigma_t = \sigma_{t'} s_i$. Using proposition 5.4.37 we have that

$$s_i v_{t'} + s_i \sum_{\substack{r \in \text{SYT}(\lambda), \\ l(\sigma_r) < l(\sigma_{t'})}} \beta'_r v_r = v_t + \sum_{\substack{r \in \text{SYT}(\lambda), \\ l(\sigma_r) < l(\sigma_t)}} \beta_r v_r$$

Hence the coefficient of v_t in $s_i v_{t'}$ is 1, thus using part b) of theorem 5.4.16, specifically the matrix of s_i which in our case is in the basis $\{v_t, v_{t'}\}$, we have that

$$z \left(1 - \frac{1}{(a_{i+1} - a_i)^2} \right) = 1$$

Now, continuing with the notation of part b) of theorem 5.4.16, as $1/z$ is the coefficient of $v_{t'}$ in $s_i v_t$ the conclusion follows immediately. $\square$

Recalling that content vectors are tuples of integers (definition 5.4.18) we have the following result.

Corollary 5.4.39. *Young's seminormal representations have their respective matrix representations defined over $\mathbb{Q}$, the field of rational numbers.*

The theorem above also tells us that the coefficients β_r in proposition 5.4.37 are rational numbers. This is a good time to stop and ponder, that the irreducible representations of symmetric groups may be defined over $\mathbb{Q}$ is nothing short of remarkable.

Having seen Young's Seminormal form we now move to *Young's Orthogonal form/representation*. This representation is characterized by choosing a Young basis that is orthonormal, so implicitly we have inner product spaces. This also somewhat justifies the names of the representations, in this form we have a normal basis, in the previous form we had a "seminormal" basis. So given a Young basis $\{v'_t \mid t \in \text{SYT}(n)\}$ here we'll consider the Young basis

$$\left\{ v_t = \frac{v'_t}{\|v'_t\|} \mid t \in \text{SYT}(n) \right\} \quad (5.17)$$

where $\|\cdot\|$ is the norm induced by the considered inner product on each irreducible space V_λ , $\lambda \vdash n$, as in C.0.4.

Theorem 5.4.40. *Considering the orthonormal Young basis above, given a Coxeter generator s_i , a standard tableau $t \in \text{SYT}(n)$ with content vector $c(t) = (a_1, \dots, a_n)$ and letting $t' = s_i t$ we have that*

$$s_i v_t = \frac{1}{r} v_t + \left(\sqrt{1 - \frac{1}{r^2}} \right) v_{t'}$$

where $r = a_{i+1} - a_i$.

Proof. First recall that in an inner product space representations are unitary, as in definition 4.1.11, by Weyl's unitary trick in theorem 4.1.12, hence considering that S_n -invariant inner product we have that $\langle s_i v, s_i v \rangle = \langle v, v \rangle$ for any vector v in the space under consideration.

Now considering the basis $\{w_t = \alpha_t v'_t \mid t \in \text{SYT}(n)\}$ from proposition 5.4.37, we have by theorem 5.4.38 that if $a_{i+1} \neq a_i \pm 1$ and $l(\sigma_{t'}) > l(\sigma_t)$ then letting $r = a_{i+1} - a_i$

$$\|w_{t'}\|^2 = \|s_i w_t - \frac{1}{r} w_t\|^2$$

Using definition C.0.4 of an induced norm and the linearity of inner products one can easily see that

$$\|s_i w_t - \frac{1}{r} w_t\|^2 = \|s_i w_t\|^2 - \frac{1}{r} \langle s_i w_t, w_t \rangle - \frac{1}{r} \langle w_t, s_i w_t \rangle + \frac{1}{r^2} \|w_t\|^2$$

One notes that as the representations are unitary $\|s_i w_t\|^2 = \|w_t\|^2$, moreover as $s_i w_t = \frac{1}{a_{i+1}-a_i} w_t + w_{t'}$ and w_t and $w_{t'}$ are orthogonal then $\langle s_i w_t, w_t \rangle = \langle w_t, s_i w_t \rangle = \frac{1}{r} \|w_t\|^2$, where the first equality comes again from the fact that representations here are unitary and $s_i^{-1} = s_i$. Therefore we have that

$$\|w_{t'}\|^2 = \|w_t\|^2 - \frac{1}{r^2} \|w_t\|^2 = \left(1 - \frac{1}{r^2}\right) \|w_t\|^2$$

With this we may consider the orthonormal basis

$$\left\{ v_t = \frac{w_t}{\|w_t\|}, v_{t'} = \frac{w_{t'}}{\left(\sqrt{1 - \frac{1}{r^2}}\right) \|w_t\|} \right\}$$

And by 5.15 we have that

$$s_i v_t = \frac{s_i w_t}{\|w_t\|} = \frac{1}{r} \frac{w_t}{\|w_t\|} + \left(\sqrt{1 - \frac{1}{r^2}}\right) \frac{w_{t'}}{\left(\sqrt{1 - \frac{1}{r^2}}\right) \|w_t\|} = \frac{1}{r} v_t + \left(\sqrt{1 - \frac{1}{r^2}}\right) v_{t'}$$

The case where $l(\sigma_{t'}) < l(\sigma_t)$ is analogous and produces the same conclusion for the action of s_i on v_t , moreover if $a_{i+1} = a_i \pm 1$ then

$$s_i v_t = \frac{s_i w_t}{\|w_t\|} = \pm \frac{w_t}{\|w_t\|} = \pm v_t$$

The important thing to note here is that when $a_{i+1} = a_i \pm 1$ then

$$\frac{1}{r} v_t + \left(\sqrt{1 - \frac{1}{r^2}}\right) v_{t'} = \pm v_t$$

Hence the formula

$$s_i v_t = \frac{1}{r} v_t + \left(\sqrt{1 - \frac{1}{r^2}}\right) v_{t'}$$

is consistent with all cases. □

We may extend the above theorem to make a parallel with part d) of theorem 5.4.16, and call it the sought Young's Orthogonal representation (although we could just as well call the previous theorem that).

Corollary 5.4.41. *Consider the orthonormal Young basis in 5.17, a Coxeter generator s_i and a standard tableau $t \in \text{SYT}(n)$ with content vector $c(t) = (a_1, \dots, a_n)$, then letting $t' = s_i t$ and $r = a_{i+1} - a_i$*

$$s_i v_t = \frac{1}{r} v_t + \left(\sqrt{1 - \frac{1}{r^2}} \right) v_{t'}$$

$$s_i v_{t'} = \left(\sqrt{1 - \frac{1}{r^2}} \right) v_t - \frac{1}{r} v_{t'}$$

Hence in the basis $\{v_t, v_{t'}\}$, s_i has the matrix form

$$\begin{bmatrix} \frac{1}{r} & \sqrt{1 - \frac{1}{r^2}} \\ \sqrt{1 - \frac{1}{r^2}} & -\frac{1}{r} \end{bmatrix}$$

We finish this subsection with a simple example.

Example 5.4.42. Consider the partition $\lambda = (n) \vdash n$, this has a unique standard tableau

$$t = \boxed{1} \boxed{2} \boxed{3} \cdots \boxed{n}$$

with content vector $c(t) = (0, 1, 2, \dots, n-1)$, hence by the previous corollary given s_i , for $i \in [n-1]$, $s_i v_t = v_t$. Thus V_λ is the trivial representation.

If we consider now the partition $\lambda' = (1, 1, 1, \dots, 1) \vdash n$ then again we have a unique standard tableau

$$t = \begin{array}{c} \boxed{1} \\ \boxed{2} \\ \boxed{3} \\ \vdots \\ \boxed{n} \end{array}$$

with content vector $c(t) = (0, -1, -2, \dots, -n+1)$. In this case we have $s_i v_t = -v_t$, thus $V_{\lambda'}$ is the alternating representation.

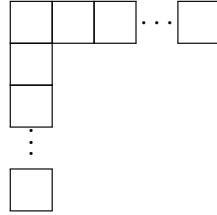
5.4.7 The hook-length formula

Before moving to the character theory of symmetric groups there's a final useful result regarding the dimension of irreducible representations of those groups. It follows from corollary 5.4.32, as there is already a rich theory on the combinatorics of partitions and Young tableaux. In fact the number of standard tableaux of a given shape, as we know it today in representation theory, was first proven by *J. S. Frame*, *G. de B. Robinson*, and *R. M. Thrall* in 1954, in their infamous paper "*The hook graphs of the symmetric group*" [FBRT54], since then many proofs of their result emerged, varying in level of complexity and approach.

While some proofs are combinatorial in nature, others resort to probability theory [GNW79], here we'll go for the former and give a simple combinatorial proof.

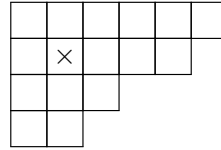
The proof of *Frame, Robinson, and Thrall's* result given here is due to *J. Bandlow* [Ban08], it requires only elementary observations and follows easily by induction, we start with a couple simple definition regarding tableaux.

Definition 5.4.43. A Ferrers diagram is a hook if it has shape $\lambda = (n - k, 1, 1, \dots, 1) \vdash n$, for some $k \in [n - 1]$. That is it's of the form

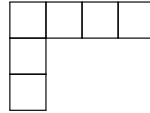


The value k is called the height of the hook, as it gives the number of rows in it, minus the first one.

In a Ferrers diagram every box defines a hook, for instance considering the following diagram



with the box $(2, 2)$ marked with an $\times$. This box produces a hook by considering the boxes to the right of $\times$ in the same row, and the boxes below $\times$ in the same column, that is the hook



The same principle applies to Young tableaux.

Definition 5.4.44. Given a Young tableau t and a box (i, j) within it, we define the hook-length of (i, j) , denoted $h_\lambda(i, j)$, to be the number of boxes in the hook associated with the box (i, j) , as seen above, in the underlying Ferrers diagram of t , that is

$$h_\lambda(i, j) = |\{(i', j) \text{ box of } t \mid i' \geq i\} \cup \{(i, j') \text{ box of } t \mid j' \geq j\}|$$

To simplify notation, from now on we let, for $\lambda \vdash n$

$$f_\lambda = |\text{SYT}(\lambda)|$$

Moreover we refer to boxes in a Ferrers diagram of shape λ , and by extension Young λ -tableaux, as "members" of the tuple λ , so if a box in a Ferrers diagram of λ has coordinates

(i, j) we say $(i, j) \in \lambda$, this avoids us having to specify notation for a Ferrers diagram distinct from the notation for partitions, as this would be redundant. The sought result of this subsection follows.

Theorem 5.4.45. *For $n \in \mathbb{N}$ consider $\lambda \vdash n$, then*

$$f_\lambda = \frac{n!}{\prod_{(i,j) \in \lambda} h_\lambda(i, j)} \quad (5.18)$$

Proof. Considering the Young graph, one notes that for $\lambda \vdash n$

$$f_\lambda = \sum_{\lambda \rightarrow \mu} f_\mu$$

Indeed as we've seen in subsection 5.2.3 the standard Young tableaux of shape λ follow uniquely from the standard Young tableaux of shape μ for $\lambda \rightarrow \mu$ in the Young graph. Our objective here is to prove an analogous result for the right-hand side of 5.18, we'll prove that

$$\frac{n!}{\prod_{(i,j) \in \lambda} h_\lambda(i, j)} = \sum_{\lambda \rightarrow \mu} \frac{(n-1)!}{\prod_{(i,j) \in \mu} h_\mu(i, j)}$$

With this we may use induction on n to prove the theorem, as we may easily verify that 5.18 is true for the (unique) trivial partition $\lambda = (1) \vdash 1$, and then assuming 5.18 is true for any $\mu \vdash n-1$ the above formula guarantees it will also be true for any $\lambda \vdash n$.

To prove the above equation is true we prove the equivalent statement

$$\sum_{\lambda \rightarrow \mu} \frac{\prod_{(i,j) \in \lambda} h_\lambda(i, j)}{\prod_{(i,j) \in \mu} h_\mu(i, j)} = n$$

This is done in three parts, in the first part we define values x_i, y_i , with $0 \leq i \leq m$ for some $m \in \mathbb{N}$, such that

$$\sum_{\lambda \rightarrow \mu} \frac{\prod_{(i,j) \in \lambda} h_\lambda(i, j)}{\prod_{(i,j) \in \mu} h_\mu(i, j)} = - \sum_{i \in [m]} \frac{\prod_{j=0}^m x_i - y_j}{\prod_{j \in [m] \setminus i} x_i - x_j} \quad (5.19)$$

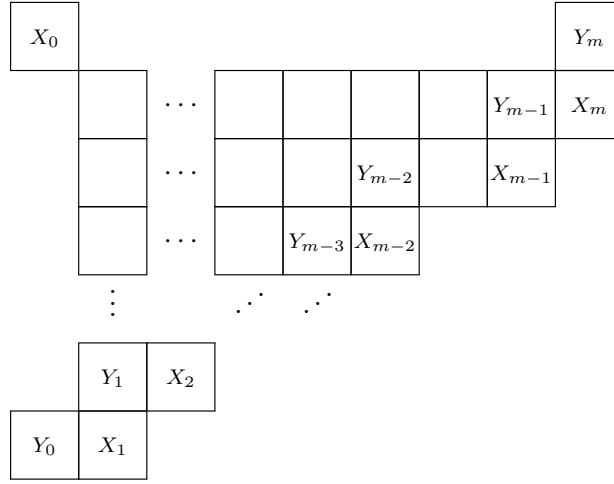
In the second part we prove that

$$- \sum_{i \in [m]} \frac{\prod_{j=0}^m x_i - y_j}{\prod_{j \in [m] \setminus i} x_i - x_j} = -\frac{1}{2} \sum_{i=0}^m (x_i^2 - y_i^2) \quad (5.20)$$

And finally in the third part we prove that

$$-\frac{1}{2} \sum_{i=0}^m (x_i^2 - y_i^2) = n \quad (5.21)$$

We start off then with the first part, equation 5.19, for this consider the removable boxes of the Ferrers diagram of λ , we label them from bottom to top as $X_i = (a_i, b_i)$, for $i \in [m]$, where (a_i, b_i) are the coordinates and m the total number of said boxes. From this we consider the boxes $Y_i = (a_{i+1}, b_i)$ for $0 \leq i \leq m$ where $b_0 = a_{m+1} = 0$, these boxes are called the *inner corners* of the Ferrers diagram, and X_i the *outer corners*. Additionally we let $X_0 = (a_0, b_0) = (0, 0)$, from this we note that the boxes X_0 , Y_0 and Y_m are not really defined on the Ferrers diagram λ , but we may consider them nonetheless as seen in the illustrative example below



Given $\mu \vdash n-1$ such that $\lambda \rightarrow \mu$ in the Young graph (the Ferrers diagram of) μ is obtained by removing one of the boxes X_i , $i \in [m]$, from λ . For this we let μ^i be the partition/Ferrers diagram resulting from removing X_i from λ .

For $0 \leq i \leq m$, recalling definition 5.4.22 of content of a box, let $x_i = c(X_i) = b_i - a_i$ and $y_i = c(Y_i) = b_i - a_{i+1}$. We note that

$$\sum_{i=0}^m x_i = \sum_{i=0}^m y_i \quad (5.22)$$

Indeed

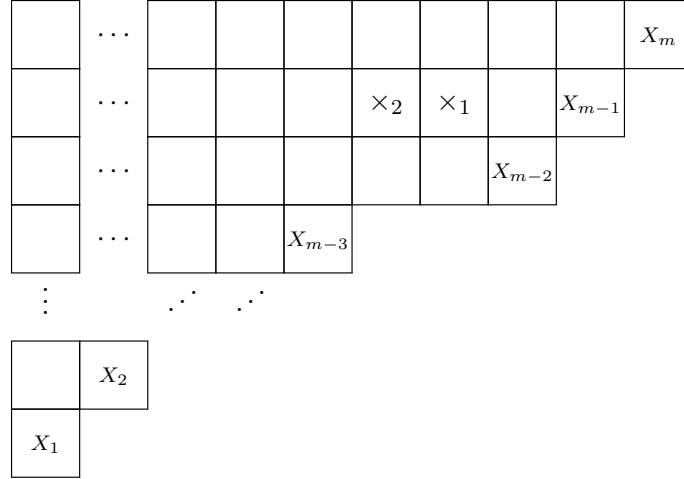
$$\sum_{i=0}^m x_i - y_i = \sum_{i=0}^m b_i - a_i - b_i + a_{i+1} = -a_0 + a_{m+1} = 0$$

Now for a fixed $i \in [m]$ consider

$$\frac{\prod_{(i,j) \in \lambda} h_{\lambda}(i, j)}{\prod_{(i,j) \in \mu^i} h_{\mu^i}(i, j)} \quad (5.23)$$

Our goal is to write the above quotient in terms of the x_i and y_i defined previously, for this we note that as μ^i is obtained from λ by removing the box X_i then every box not in the row or column of X_i in λ will have the same hook-length whether seen in λ or μ^i , hence there will occur massive cancellation of factors in the above quotient. Cancellation may

also occur in another situation, say $X_i = (i, j)$ and $X_{i-1} = (i', j')$ such that $|j - j'| > 2$, then given a box b in a row l in λ and between columns j and j' if the box immediately to left of it is also between columns j and j' it will have the same hook length as b when seen as a box of μ^l , this is illustrated in the example below.



Indeed letting λ be the Ferrers diagram above we observe that the hook-length of the box $\times_1$ will be the same as the hook-length of the box $\times_2$ when this latter is seen as a box in μ^{m-1} , where μ^{m-1} is λ without the box X_{m-1} .

An analogous observation can be made regarding elements in the same columns instead of rows, where then the situation is dependent on the existence of $X_i = (i, j)$ and $X_{i-1} = (i', j')$ such that $|i - i'| > 2$.

By the observations made above most factor will cancel in the quotient 5.23, so we now look at what factor won't cancel. To that end observe the general shape of the Ferrers diagram of λ given in figure 5.7, with the box X_i labeled and marked in back. We exemplify there the canceling pairs in the row and column of X_i mentioned above, marked as grey octagons.

We need only to look at the row and column of X_i as any other box's hook-length cancels in the quotient 5.23, moreover we have the canceling pairs in the row and column of X_i , hence we can see that the only boxes in λ that have no match in μ^i are the grey boxes in figure 5.7. Similarly the boxes immediately to left of the grey boxes in the row of X_i and the boxes immediately above the grey boxes in the column of X_i , when seen as boxes of μ^i , have no matching box in λ , these boxes are marked in figure 5.8 as tilted grey squares.

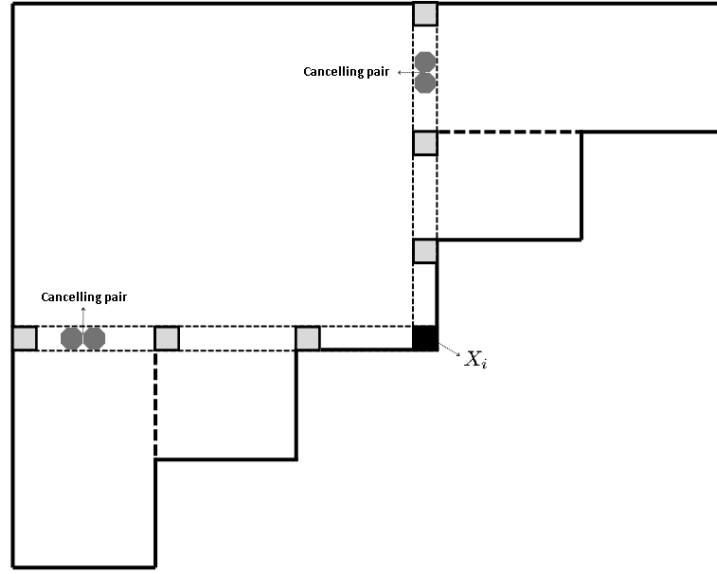


FIGURE 5.7: High level view of a Ferrers diagram with canceling boxes and boxes of interest.

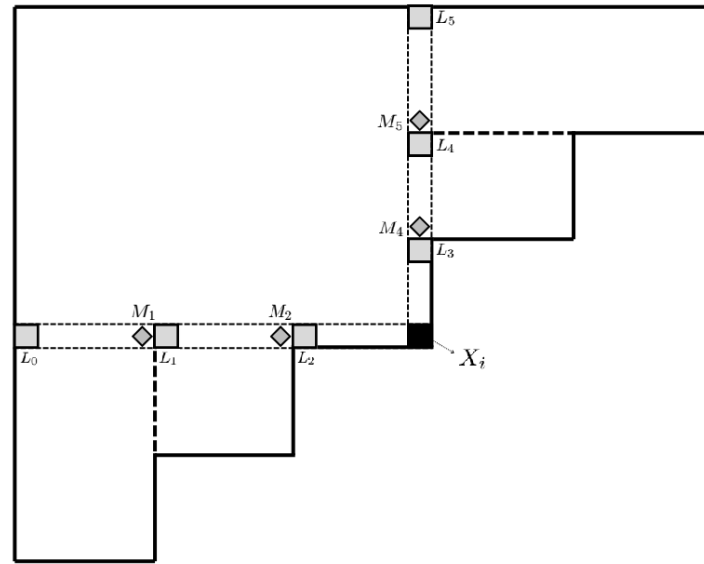


FIGURE 5.8: High level view of a Ferrers diagram with boxes of interest.

We denote the grey boxes as seen in figure 5.8 by L_j and the tilted grey boxes by M_j . Note that, recalling $X_i = (a_i, b_i)$, then the grey boxes in the row of X_i are

$$L_j = (a_i, b_j + 1) \quad 0 \leq j < i$$

and the grey boxes on the column of X_i are

$$L_j = (a_{j+1} + 1, b_i) \quad i \leq j \leq m$$

Moreover for the tilted boxes in the row of X_i we have

$$M_j = (a_i, b_j) \quad 1 \leq j < i$$

and for the tilted boxes in the column of X_i

$$M_j = (a_j, b_i) \quad i < j \leq m$$

Considering this we have that 5.23 reduces to

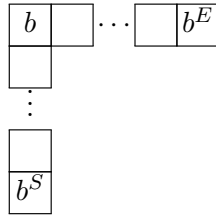
$$\frac{\prod_{j=0}^m h_\lambda(L_j)}{\prod_{j \in [m] \setminus i} h_\mu(M_j)}$$

And so we may extend this to all X_i getting

$$\sum_{\lambda \rightarrow \mu} \frac{\prod_{(i,j) \in \lambda} h_\lambda(i, j)}{\prod_{(i,j) \in \mu} h_\mu(i, j)} = \sum_{i \in [m]} \frac{\prod_{j=0}^m h_\lambda(L_j)}{\prod_{j \in [m] \setminus i} h_\mu(M_j)}$$

Where in this case each L_j and M_j correspond to some particular X_i , $i \in [m]$.

We now compute the hook-lengths above, for that we first note that a hook length may be computed as follows. For a box b let b^S be the southeast box below b in its hook and let b^E be the box most to east of b in its hook, as shown below in hook of b .



Then one may easily verify that

$$h_\lambda(b) = c(b^E) - c(b^S) + 1$$

where $c(b)$ is the content of the box b as in definition 5.4.22. Now note that for $j \in [i - 1]$, $y_j = c(L_j^S) - 1$ and $x_j = c(L_j^E)$, thus

$$h_\lambda(L_j) = c(L_j^E) - c(L_j^S) + 1 = x_j - y_j$$

Similarly, for $j \in [i - 1]$, as X_i is one unit to the right of M_j^E in μ^i we have that

$$h_{\mu^i}(L_j) = c(M_j^E) + 1 - c(M_j^S) = x_i - x_j$$

Analogously, for $i \leq j \leq m$

$$h_\lambda(L_j) = y_j - x_i$$

And for $i < j \leq m$

$$h_{\mu^i}(M_j) = x_j - x_i$$

Therefore

$$\sum_{i \in [m]} \frac{\prod_{j=0}^m h_\lambda(L_j)}{\prod_{j \in [m] \setminus i} h_\mu(M_j)} = \sum_{i \in [m]} \frac{\prod_{j=0}^{i-1} (x_i - y_j) \prod_{j=i}^m -(x_i - y_j)}{\prod_{j=1}^{i-1} (x_i - x_j) \prod_{j=i+1}^m -(x_i - x_j)} = - \sum_{i \in [m]} \frac{\prod_{j=0}^m x_i - y_j}{\prod_{j \in [m] \setminus i} x_i - x_j}$$

This concludes the first part of the proof, we now move to the second part, namely proving equation 5.20. This follows from several simple observations, first consider the polynomial function

$$P(t) = - \sum_{i \in [m]} \frac{\prod_{j=0}^m x_i - y_j}{\prod_{j \in [m] \setminus i} x_i - x_j} \prod_{j \in [m] \setminus i} t - x_j$$

One may observe that $P(t)$ is essentially the *Lagrange* interpolation polynomial [Atk89, 3.1.6] of the function $S(t) = - \prod_{j=0}^m t - y_j$ at x_k , for $k \in [m]$, indeed

$$P(x_k) = - \prod_{j=0}^m x_k - y_j$$

And it is clear $P(t)$ has the required form, so we are simply multiplying the left hand side of 5.20 by $\prod_{j \in [m] \setminus i} t - x_j$ to give it the form of a *Lagrange* interpolation polynomial. With this in mind we proceed.

Note that $P(t)$ is a polynomial of degree $m-1$ with leading coefficient, that is the coefficient of the term t^{m-1} in $P(t)$

$$- \sum_{i \in [m]} \frac{\prod_{j=0}^m x_i - y_j}{\prod_{j \in [m] \setminus i} x_i - x_j}$$

As this is the left hand side of 5.20 this part of the proof follows by evaluating the coefficient of t^{m-1} in $P(t)$ in a different manner. For this, considering the value of $P(t)$ at x_k , $k \in [m]$, we may produce a polynomial having the additive inverse of those values also at x_k , that being

$$Q(t) = -S(t) = \prod_{j=0}^m t - y_j$$

and use this to produce a polynomial with zeros at x_k , for $k \in [m]$. Moreover note that $Q(t)$ is a polynomial of degree $m+1$, thus $Q(t) + P(t)$ has degree $m+1$ and zeros at x_k for $k \in [m]$, hence $Q(t) + P(t)$, being a polynomial over $\mathbb{C}$, an algebraically closed field,

may be decomposed into linear factors as

$$(t - \alpha) \prod_{j=1}^m t - x_j$$

Where $\alpha \in \mathbb{C}$ is the remaining undetermined zero of $Q(t) + P(t)$. From this we have that

$$\begin{aligned} P(t) &= (t - \alpha) \left(\prod_{j=1}^m t - x_j \right) - Q(t) = (t - \alpha) \left(\prod_{j=1}^m t - x_j \right) - \prod_{j=0}^m t - y_j = \\ &= \left(-\alpha - \sum_{i=1}^m x_i + \sum_{i=0}^m y_i \right) t^m + \left(\alpha \sum_{i=1}^m x_i + \sum_{i,j=1}^m x_i x_j - \sum_{i,j=0}^m y_i y_j \right) t^{m-1} + \dots \end{aligned}$$

Now as $P(t)$ is a polynomial of degree $m - 1$ it must be the case that the coefficient of t^m above equals 0, and using 5.22 we conclude that $\alpha = 0$, moreover as $x_0 = 0$ we may write the coefficient of t^{m-1} above as

$$\sum_{i,j=0}^m x_i x_j - y_i y_j$$

To conclude note that

$$\sum_{i,j=0}^m x_i x_j - y_i y_j = \frac{1}{2} \left(\left(\sum_{i=0}^m x_i \right)^2 - \left(\sum_{i=0}^m y_i \right)^2 - \sum_{i=0}^m (x_i^2 - y_i^2) \right)$$

Again using 5.22 the above equals

$$-\frac{1}{2} \sum_{i=0}^m (x_i^2 - y_i^2)$$

This concludes the second part of the proof, we now proceed to the third and final part, namely equation 5.21. Recalling the definitions of x_i and y_i , as differences of coordinates of boxes in λ , we have that

$$\begin{aligned} -\frac{1}{2} \sum_{i=0}^m ((b_i - a_i)^2 - (b_i - a_{i+1})^2) &= a_0^2 - a_{m+1}^2 - \frac{1}{2} \sum_{i=0}^m -2b_i a_i + 2b_i a_{i+1} = \\ &= \sum_{i=1}^m b_i (a_i - a_{i+1}) \end{aligned}$$

Looking at figure 5.9 we may interpret the above sum as the sum of the number of boxes in each delimited rectangle in the Ferrers diagram, this considering also that $a_{m+1} = 0$, and as the total number of boxes in Ferrers diagram of shape $\lambda \vdash n$ is n we have that

$$\sum_{i=1}^m b_i (a_i - a_{i+1}) = n$$

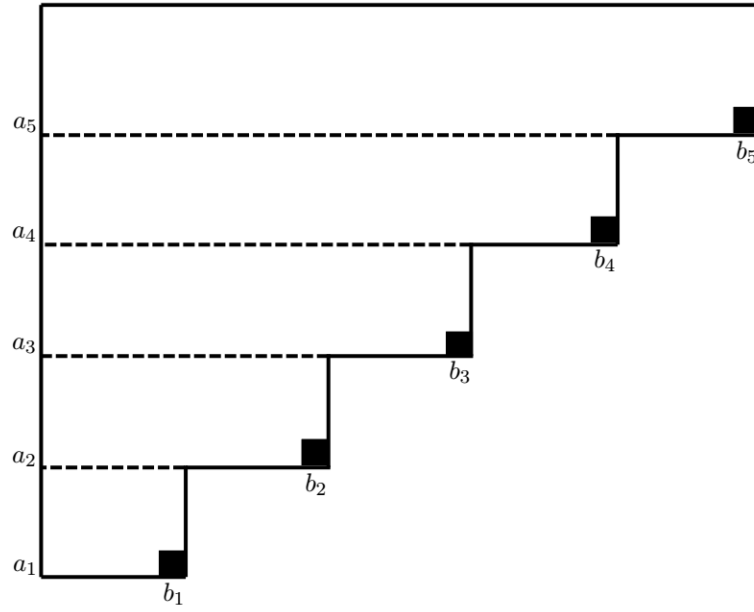


FIGURE 5.9: Ferrers diagram segmented by the boxes X_i , $i \in [m]$.

This concludes the proof of this theorem. □

As stated, by corollary 5.4.32, the above theorem gives us a way of computing the dimension of an irreducible representation V^λ from the hook-lengths of the boxes in the Ferrers diagram of λ , thus we have the following immediate result.

Corollary 5.4.46. *Let V_λ , with $\lambda \vdash n$, be an irreducible representation of S_n , then*

$$\dim_{\mathbb{C}}(V_\lambda) = \frac{n!}{\prod_{(i,j) \in \lambda} h_\lambda(i,j)}$$

where $h_\lambda(i,j)$ is the hook-length of the box with coordinates (i,j) in the Ferrers diagram of λ .

5.5. Character theory of S_n

The character theory of symmetric groups, as we know it today, relies heavily on the combinatorics of diagrams and tableaux, yet this has not always been the case. In the early work of *Frobenius*, in his paper "*Über die Charaktere der symmetrischen Gruppe*", the characters of irreducible representations of S_n were computed by building certain generating functions from the values of the characters, one may see a modern statement of his result in [FH04, 4.10], called the *Frobenius formula*.

The next step in the history of the group characters of S_n came from "D. E. Littlewood and A. R. Richardson" in [LR34, sec. 11], their method for computing the character values came as a corollary of the *Frobenius formula*. Slightly later yet independently F. D. Murnaghan gave a similar derivation of the same result [Mur37], and three year later T. Nakayama gave a more concise proof, still resorting to the *Frobenius formula*, citing the work of F. D. Murnaghan. The result of this endeavor came to be known as the *Murnaghan-Nakayama rule* and it is now the staple result characterizing the group characters of symmetric groups.

The last notable advancement came from J. D. James in his book [Jam78, 21.1], where he proved the Murnaghan-Nakayama rule using the *Littlewood-Richardson rule*, which pertains to the decomposition into irreducibles of S_n of the outer tensor product of two representations of S_p and S_k respectively, with $p + k = n$, induced to S_n . The theory given here is based on his work, as seen in [Sag01].

5.5.1 The case of a cycle

Let $\lambda \vdash n$, we denote the character of an irreducible S_n -space V^λ (as in 5.4.33 and the subsequent statement) by χ^λ , moreover as characters are constant on the conjugacy classes of a group, as we've seen in part a) of proposition 4.3.9, we let χ_C^λ be the value of χ^λ at any element in a conjugacy class C of S_n .

In this section we'll focus on the values of a character at n -cycle permutations of S_n , that is permutations that are a single cycle of length n , like for instance the permutation $(2\ 1\ 3\ 4)$ of S_4 , hence C will be the class with representative, say $(1\ 2\ 3 \cdots n)$. The following result may be seen as a special case of the Murnaghan-Nakayama rule, but will be used as a lemma in its proof.

Recall definition 5.4.43.

Lemma 5.5.1. *Let $\lambda \vdash m$ and C be the conjugacy class of S_m with representative $(1\ 2 \cdots m)$. Consider the character χ_C^λ of the (irreducible) S_m -space V^λ , then*

$$\chi_C^\lambda = \begin{cases} (-1)^k & \text{if } \lambda = (m - k, 1, \dots, 1) \vdash m, \text{ i.e. } \lambda \text{ is a hook of height } k \\ 0 & \text{if } \lambda \text{ is not a hook} \end{cases}$$

Proof. * Recalling the YJM elements, in definition 5.4.8, we note that $X_2 X_3 \cdots X_m$ equals the sum of all m -cycles of S_m , each m -cycle appearing only once in the sum. Indeed, by

*The following proof comes from [CSST10, 3.4.10].

induction on m , this is the case for $m = 2$ as $X_2 = (1\ 2)$ is the only 2-cycle of S_2 , now assuming this is true for all $k \in [m - 1]$ let $(\alpha_1\ \alpha_2\ \cdots\ \alpha_{m-1})$ be an $(m - 1)$ -cycle, and a summand, in $X_2X_3\cdots X_{m-1}$, then in $X_2X_3\cdots X_m$ we'll have the sum

$$\sum_{i \in [m-1]} (\alpha_1\ \alpha_2\ \cdots\ \alpha_{m-1})(i\ m)$$

As i must be somewhere in the cycle $(\alpha_1\ \alpha_2\ \cdots\ \alpha_{m-1})$, but not m , we may rewrite the above as

$$\sum_{i \in [m-1]} (\alpha_1\ \alpha_2\ \cdots\ \alpha_a\ i\ \alpha_b\ \cdots\ \alpha_{m-1})(i\ m)$$

But then $(\alpha_1\ \alpha_2\ \cdots\ \alpha_a\ i\ \alpha_b\ \cdots\ \alpha_{m-1})(i\ m) = (\alpha_1\ \alpha_2\ \cdots\ \alpha_a\ i\ m\ \alpha_b\ \cdots\ \alpha_{m-1})$, which is an m -cycle, moreover it is clear that no two m -cycles in the above sum will be equal, and it is also clear by this that $X_2X_3\cdots X_m$ is the sum of all m -cycles of S_m each appearing only once, as we have $(m - 1)!$ summands and they are all different, which is the number of all m -cycles of S_m .

Now note that by the (natural) isomorphism between the group algebra $\mathbb{C}S_n$ and the space of complex functions $\mathbb{C}[S_n]$, $X_2X_3\cdots X_m$ may be identified with the characteristic function of the conjugacy class C , that is the function

$$\delta(\pi) = \begin{cases} 1 & \text{if } \pi \in C \\ 0 & \text{otherwise} \end{cases}$$

And this function is a class function, that is constant on the conjugacy classes of S_n , hence by proposition 4.4.7, when considering the action of $X_2X_3\cdots X_m$ on V^λ , we have that

$$X_2X_3\cdots X_m \equiv \frac{m!}{\dim_{\mathbb{C}}(V^\lambda)} \langle \delta, \overline{\chi^\lambda} \rangle_{S_m} id_{V^\lambda}$$

And recalling definition 4.3.14 of inner product we get

$$X_2X_3\cdots X_m \equiv \frac{(m - 1)!}{\dim_{\mathbb{C}}(V^\lambda)} \overline{\chi_C^\lambda} id_{V^\lambda} \quad (5.24)$$

With this we have all we need to prove the lemma. First consider the case where λ is not a hook, and let $t \in \text{SYT}(\lambda)$. If j is the number in the box with coordinates $(2, 2)$ in t , then for the content of t , $c(t) = (a_1, a_2, \dots, a_m)$, we have that $a_j = 0$ as the content of the box $(2, 2)$ is $c(2, 2) = 2 - 2 = 0$ (recall definition 5.4.22 and 5.4.23), thus, recalling definition 5.4.11, by proposition 5.4.29 if v_t is the Young vector associated with the standard tableau t , we have that

$$X_j v_t = a_j v_t = 0$$

Therefore

$$X_2 X_3 \cdots X_m v_t = 0$$

As t is any standard tableaux in $\text{SYT}(\lambda)$ we have that

$$X_2 X_3 \cdots X_m \equiv 0$$

Thus by equation 5.24 it must be the case that $\overline{\chi_C^\lambda} = 0$, thus $\chi_C^\lambda = 0$.

If λ is a hook the let $\lambda = (m - k, 1, \dots, 1)$, then is is clear that $c(t) = (0, 1, 2, \dots, m - k - 1, -1, -2, \dots, -k)$ (or some permutation of this) as we may write the contents of each box in it as below

0	1	2	...	$\begin{matrix} m-k \\ -1 \end{matrix}$
-1				
-2				
$\vdots$				
$-k$				

Hence for $t \in \text{SYT}(\lambda)$ with associated Young vector v_t we have that

$$X_2 X_3 \cdots X_m v_t = a_2 a_3 \cdots a_m v_t = (-1)^k k! (m - k - 1)! v_t \quad (5.25)$$

Now, going back to equation 5.24, by corollary 5.4.32

$$\dim_{\mathbb{C}}(V^\lambda) = \binom{m-1}{k}$$

as any such (standard) hook is uniquely determined by the numbers placed in the boxes with coordinates $(2, 1), (3, 1) \dots, (k+1, 1)$, hence equation 5.24 becomes

$$X_2 X_3 \cdots X_m \equiv k! (m - k - 1)! \overline{\chi_C^\lambda} id_{V^\lambda}$$

Comparing this with equation 5.25 we get that $\overline{\chi_C^\lambda} = (-1)^k$, thus the conclusion follows immediately. □

5.5.2 The Littlewood-Richardson rule

In corollary 5.3.6 we saw how the multiplicity of an irreducible representation in some corresponding restricted representation could be computed by counting paths in the branching graph, and subsequently in the case of symmetric groups, by theorem 5.4.31, we saw that this could be restated in terms of the combinatorics of partitions/Ferrers diagrams

(refer to the proof of proposition 5.4.34). These results could be reformulated to consider induced representations instead of restricted representations, by remark 4.

The Littlewood-Richardson rule generalizes these ideas to tensor products of irreducibles of symmetric groups, in particular, letting $\mu \vdash p$ and $\nu \vdash q$ with $p + q = n$, it is interested in the decomposition of $(V^\mu \otimes V^\nu)|^{S_n}$ into irreducibles of S_n . As $S_p \times S_q$ may be seen as a subgroup of S_n this indeed makes sense, and the decomposition will have the form

$$(V^\mu \otimes V^\nu)|^{S_n} = \bigoplus_{\lambda \vdash n} c_{\mu\nu}^\lambda V^\lambda \quad (5.26)$$

where $c_{\mu\nu}^\lambda \in \mathbb{N}$. These coefficients are called the *Littlewood-Richardson coefficients* and knowing them amounts to knowing the sought decomposition. The Littlewood-Richardson rule characterizes these coefficients in terms of the combinatorics of *generalized tableaux*.

Definition 5.5.2. A *generalized tableau of shape $\lambda \vdash n$* is a Ferrers diagram of λ where each box is populated with positive integers, and repetitions are allowed.

A *generalized tableau* is said to be *semistandard* if its rows weakly increase and its columns strictly increase, that is, considering we allow the repetition of integers, given a box in some row the integer in the box immediately to its right must be bigger or equal, and the integer in the box immediately below it must be bigger.

Example 5.5.3. Both tableaux below are generalized tableaux

$$t_1 = \begin{array}{|c|c|c|} \hline 1 & 1 & 2 \\ \hline 2 & 3 & \\ \hline \end{array} \quad t_2 = \begin{array}{|c|c|c|} \hline 2 & 1 & 1 \\ \hline 3 & 2 & \\ \hline \end{array}$$

yet only the tableau t_1 is semistandard.

Definition 5.5.4. Let t be a generalized tableau, then the *type of t* is an m -tuple, $\mathfrak{k} = (k_1, k_2, \dots, k_m)$ where k_i equals the number of times the integer i appears in t , thus we set m to be the greatest integer that appears in t .

Continuing with the previous example, the type of t_1 is $\mathfrak{k} = (2, 2, 1)$, which is equal to the type of t_2 .

The above definitions apply *verbatim* to shapes like λ/μ with $\lambda \vdash n$ and $\mu \vdash k$, $k \in [n-1]$. So if we consider $\lambda = (4, 3, 1)$ and $\mu = (2, 1, 1)$ then the following is a generalized tableau of shape λ/μ

$$t = \begin{array}{|c|c|} \hline & 4 & 2 \\ \hline 2 & 3 & \\ \hline \end{array}$$

Here we'll state the Littlewood-Richardson rule without proof as it requires concepts and results that go beyond our intended scope, referring the interested reader to [Sag01, 4.9.4], yet before doing so we need a couple additional definitions to understand what the Littlewood-Richardson coefficients are counting, from the generalized tableaux perspective. First, given a sequence of integers $(a_1, a_2, \dots, a_n)$ we define its concatenation as the "word" $a_1 a_2 a_3 \dots a_n$, so for instant the concatenation of $(1, 4, 5, 3, 10, 17)$ is 14531017.

Definition 5.5.5. Let t be a tableau, then the row word of t , denoted π_t , is the concatenation of the integers in each row of t going from the last row to the first, that is if t has shape $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k)$ then

$$\pi_t = R_k R_{k-1} \dots R_1$$

where R_i is the concatenation of the integers in the row i .

Example 5.5.6. Consider the tableau

$$t = \begin{array}{|c|c|c|c|} \hline 1 & 3 & 5 & 7 \\ \hline 2 & 6 & & \\ \hline 4 & & & \\ \hline \end{array}$$

then the row word of t is $\pi_t = 4261357$.

Definition 5.5.7. A ballot sequence, or lattice permutation, is a concatenation of positive integers $\pi = i_1 i_2 \dots i_n$ such that for any prefix of π , $\pi_k = i_1 i_2 \dots i_k$, $k < n$, and any integer l , the number of times l appears in the π_k is greater or equal to the number of times $l + 1$ appears in π_k . A reverse ballot sequence is a concatenation of positive integers $\pi = i_1 i_2 \dots i_n$ such that $\pi^r = i_n i_{n-1} \dots i_1$ is a ballot sequence.

With this we can now state the Littlewood-Richardson rule.

Theorem 5.5.8. For $\mu \vdash p$, $\nu \vdash q$ with $p + q = n$ and $\lambda \vdash n$, the Littlewood-Richardson coefficient $c_{\mu\nu}^\lambda$ equals the number of semistandard (generalized) tableaux t such that

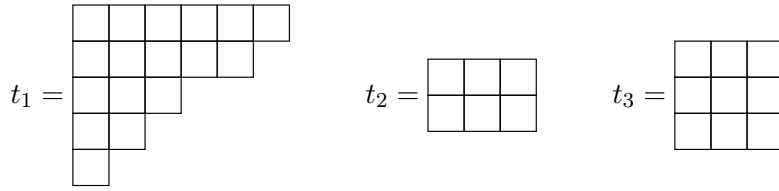
- a) t has shape λ/μ and type ν .
- b) The row word of t , π_t , is a reverse ballot sequence.

The usefulness of the above theorem to us comes when we consider characters, as for $\mu \vdash p$, $\nu \vdash q$ with $p + q = n$ and V^μ, V^ν recalling part a) of proposition 4.3.12 and definition 4.6.2 it is the case that

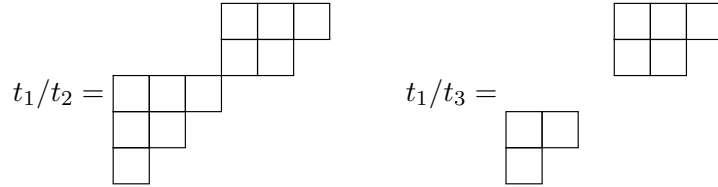
$$\chi_{V^\mu \otimes V^\nu}^{S_n} = \sum_{\lambda \vdash n} c_{\mu\nu}^\lambda \chi^\lambda$$

Under more restrictive conditions the Littlewood-Richardson coefficients may be straightforwardly computed, we next present such a case, yet we'll need to be mindful of a particular type of diagram.

When considering diagrams of the form λ/μ for some $\lambda \vdash n$, $\mu \vdash m$ one thing that can occur is that the end result may be regarded as two separate diagrams, for instance if



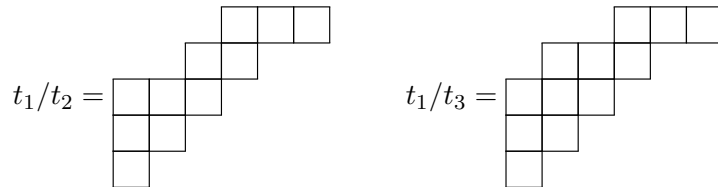
Then



When this happens we say that the diagram (t_1/t_2 or t_2/t_3 in this case) is *not connected*, or *disconnected*. Yet if for instance



Then



And in this case we say that the diagram is *connected*, suffice to say any Ferrers diagram is a connected diagram.

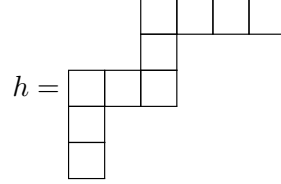
Definition 5.5.9. A diagram of the form λ/μ for some $\lambda \vdash n$, $\mu \vdash m$ is said to be *connected* if at most two boxes in it have a single neighbor, otherwise the diagram is said to be *disconnected*. When a diagram t is disconnected we may decompose it into several separate diagrams $\{t_i\}_{i \in [n]}$ and in this case we say that t is the (disjoint) union of $\{t_i\}_{i \in [n]}$ and denote this decomposition by $t = \uplus_{i \in [n]} t_i$.

Definition 5.5.10. A skew hook, or rim hook, is a diagram of the form λ/μ for some $\lambda \vdash n$, $\mu \vdash m$ that is connected and does not contain the Ferrers diagram $(2, 2) \vdash 4$ as a subdiagram. As with hooks we

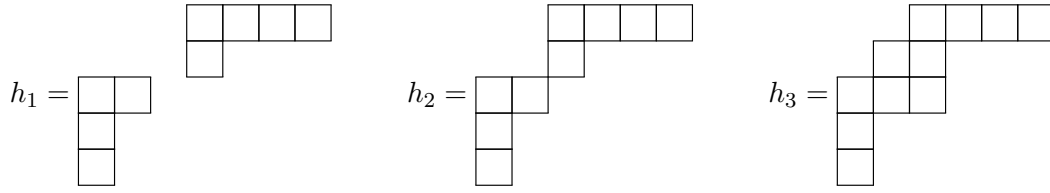
define the height of a skew hook h to be

$$ll(h) = (\text{number of rows of } h) - 1$$

Example 5.5.11. The following diagram is a skew hook



Yet the following three diagrams are not skew hooks.



Lemma 5.5.12. Let $\lambda \vdash n$, $\mu \vdash n - m$ and $\nu = (r, 1, \dots, 1) \vdash m$, then $c_{\mu\nu}^\lambda = 0$ unless λ/μ decomposes as an union of skew hooks, and in that case if λ/μ decomposes as an union of k skew hooks having a combined total of c columns then

$$c_{\mu\nu}^\lambda = \binom{k-1}{c-r}$$

Proof. By the Littlewood-Richardson rule, in theorem 5.5.8, $c_{\mu\nu}^\lambda$ equals the number of semi-standard tableaux t of shape λ/μ and type ν , that is t has r boxes populated with a 1 and $m - r$ boxes bijectively populated with the numbers in $\{2, 3, \dots, m - r + 1\}$, such that the row word of t , π_t , is a reverse ballot sequence. By definition 5.5.7 of ballot sequence it must be the case that the numbers greater than 1 in π_t^r appear in increasing order, thus π_t has the numbers in $\{2, 3, \dots, m - r + 1\}$ occurring in decreasing order with the possibility of several 1's appearing in the midst of the aforementioned numbers, but then as this is the row word of t , which is semistandard thus the rows weakly increase, it must be the case that

- a) If a box in t has a neighbor to its right it must have a 1 in it.

Moreover as the columns strictly increase

- b) If a box in t has a neighbor above it must have a number greater than 1 in it.

Now if t contains the Ferrers diagram $(2, 2) \vdash 4$ anywhere within it there is no way so satisfy both conditions a) and b) above simultaneously, thus there are no such semistandard tableaux and so $c_{\mu\nu}^\lambda = 0$.

On the other hand if t does not contain the Ferrers diagram $(2, 2) \vdash 4$ then by definition 5.5.10 t is either a skew hook or decomposes as an union of skew hooks. Let then $t = \uplus_{i \in [k]} h_i$ be the decomposition of t into skew hooks, then conditions a) and b) imply that h_i , $i \in [k]$, must have the form, for instance

$$h_i = \begin{array}{|c|c|c|c|} \hline & & 1 & 1 & 1 & b \\ \hline & & a & & & \\ \hline 1 & 1 & a+1 & & & \\ \hline a+2 & & & & & \\ \hline a+3 & & & & & \\ \hline \end{array}$$

where $a > 1$ is the smallest number that has not yet appeared in the skew hooks below h_i and b is either 1 or $a - 1$, thus all the entries in h_i are determined once we choose the value of b . By our remarks above regarding π_t (decreasing order) the b in h_1 must be a 1, hence we have $c - (k - 1)$ ones already fixed in t . Thus there are $r - c + k - 1$ ones left to distribute among the $k - 1$ boxes market with a b in t , and this can be done in

$$\binom{k-1}{r-c+k-1} = \binom{k-1}{c-r}$$

ways, thus $c_{\mu\nu}^\lambda = \binom{k-1}{c-r}$. □

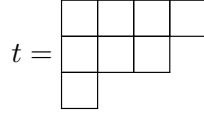
5.5.3 The Murnhagan–Nakayama rule

We now arrive at the staple result in the character theory of symmetric groups. By our early remarks at the beginning of this section the previous two subsection provide the necessary results to prove the Murnhagan–Nakayama rule, which gives a recursive way of computing the character χ_λ , with $\lambda \vdash n$, at a conjugacy class C of S_n , so χ_C^λ .

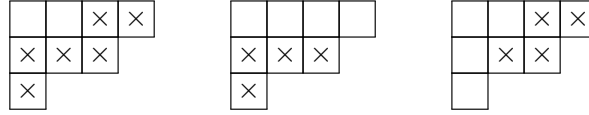
The focal objects will be the skew hooks presented in the previous subsection.

Definition 5.5.13. Given a Ferrers diagram $\lambda \vdash n$, a skew hook of λ is a diagram of the form λ/μ for some $\mu \vdash m < n$ satisfying definition 5.5.10.

Example 5.5.14. Considering the Ferrers diagram



below we show three skew hooks of t , whose boxes are marked by a $\times$ on the diagram of t for easy reference



As is known there is a natural bijection between partitions of n and the conjugacy classes of S_n , as any conjugacy class of S_n contains only elements with a particular cycle structure [Rot95, 3.5] and cycle structures are in particular partitions of n . Thus we may denote a conjugacy class C of S_n by its corresponding partition of n .

Just as with Ferrers diagrams we may define a subtraction between similar partitions as, for $\zeta_1 = (a_1, a_2, \dots, a_k)$ and $\zeta_2 = (a_1, a_2, \dots, a_p) *$ with $p < k$

$$\zeta_1 - \zeta_2 = (a_{p+1}, a_{p+2}, \dots, a_k)$$

Moreover if $h = \lambda/\mu$ is a skew hook we let $\mu = \lambda \setminus h$. We are now ready to state the Murnhagan–Nakayama rule.

Theorem 5.5.15. For $\zeta = (a_1, a_2, \dots, a_k)$, $\lambda \vdash n$

$$\chi_\zeta^\lambda = \sum_h (-1)^{ll(h)} \chi_{\zeta - (a_1)}^{\lambda \setminus h}$$

Where the above sum is over all possible skew hooks h of λ having a_1 boxes.

Proof. Let $m = a_1$ and consider $\pi\sigma \in S_{n-m} \times S_m \subseteq S_n$, where π has cycle structure $(a_2, a_2, \dots, a_k)$ and σ is an m -cycle, note that $\pi\sigma$ has cycle structure ζ . By theorem 4.3.34, theorem 4.3.32 and part b) of proposition 4.3.12, $\{\chi^\mu \chi^\nu \mid \mu \vdash n-m, \nu \vdash m\}$ is a basis of the space of class functions $C(S_{n-m} \times S_m)$, and as χ_λ is a class function on S_n then when restricted to $S_{n-m} \times S_m$ it's a class function in $C(S_{n-m} \times S_m)$. Now, using definition 4.6.7, note that

$$\chi_\zeta^\lambda = \chi^\lambda(\pi\sigma) = \chi_{S_{n-m} \times S_m}^\lambda(\pi\sigma) = \sum_{\substack{\mu \vdash n-m, \\ \nu \vdash m}} \alpha_{\mu\nu} \chi^\mu(\pi) \chi^\nu(\sigma) \quad (5.27)$$

*The reason we use such an outstanding symbol as ζ for partitions here is because we want to stress that these pertain to conjugacy classes and not Ferrers diagrams as usual.

with $\alpha_{\mu\nu} \in \mathbb{C}$.

In fact $\alpha_{\mu\nu} \in \mathbb{N}$ as it's the multiplicity of $V^\mu \otimes V^\nu$ in the decomposition of $V^\lambda \downarrow_{S_{n-m} \times S_m}$ into irreducibles, and with this in mind recalling 5.26, namely the Littlewood-Richardson coefficients, and Frobenius reciprocity (theorem 4.6.8) we have that

$$\begin{aligned} \alpha_{\mu\nu} &= \langle \chi_{S_{n-m} \times S_m}^\lambda, \chi^\mu \chi^\nu \rangle_{S_{n-m} \times S_m} = \\ &= \langle \chi^\lambda, (\chi^\mu \chi^\nu)^{S_n} \rangle_{S_n} = c_{\mu\nu}^\lambda \end{aligned}$$

Thus we may rewrite 5.27 as

$$\chi^\lambda(\pi\sigma) = \sum_{\mu \vdash n-m} \chi^\mu(\pi) \sum_{\nu \vdash m} c_{\mu\nu}^\lambda \chi^\nu(\sigma) \quad (5.28)$$

but as σ is a m -cycle, by lemma 5.5.1

$$\chi^\nu(\sigma) = \begin{cases} (-1)^{m-r} & \text{if } \nu = (r, 1, \dots, 1) \vdash m \\ 0 & \text{if } \nu \text{ is not a hook} \end{cases}$$

Therefore, in 5.28, we need only to consider the cases where $\nu \vdash m$ is a hook, and consequently we need only to determine the Littlewood-Richardson coefficients $c_{\mu\nu}^\lambda$ when $\lambda \vdash n$, $\mu \vdash n-m$ and $\nu = (r, 1, \dots, 1) \vdash m$, which was done in lemma 5.5.12, thus 5.28 becomes

$$\chi^\lambda(\pi\sigma) = \sum_{\mu} \chi^\mu(\pi) \sum_{r \in [m]} \binom{k_\mu - 1}{c_\mu - r} (-1)^{m-r} \quad (5.29)$$

where the $\mu \vdash n-m$ are such that λ/μ decomposes as an union of skew hooks and c_μ, k_μ are as in lemma 5.5.12 (there simply named k and c).

Analyzing closely the proof of lemma 5.5.12 we note that it must the case that $r \leq c_\mu$ and $c_\mu - r \leq k_\mu - 1 < m$, thus we let $\binom{k_\mu - 1}{c_\mu - r} = 0$ when either $r > c_\mu$ or $c_\mu - r > k_\mu - 1$, therefore

$$\begin{aligned} &\sum_{r \in [m]} \binom{k_\mu - 1}{c_\mu - r} (-1)^{m-r} = \\ &= (-1)^{m-1} \binom{k_\mu - 1}{c_\mu - 1} + (-1)^{m-2} \binom{k_\mu - 1}{c_\mu - 2} + \dots + (-1)^{m-c_\mu} \binom{k_\mu - 1}{0} = \\ &= (-1)^{m-c_\mu} \left(\binom{k_\mu - 1}{0} - \binom{k_\mu - 1}{1} + \binom{k_\mu - 1}{2} - \dots \pm \binom{k_\mu - 1}{k_\mu - 1} \right) = \\ &= (-1)^{m-c_\mu} \sum_{i=0}^{k_\mu-1} \binom{k_\mu - 1}{i} (-1)^i \end{aligned}$$

Now recalling the (straightforward to determine) identity $(1+x)^n = \sum_{i=0}^n \binom{n}{i} x^i$ we get that

$$\sum_{r \in [m]} \binom{k_\mu - 1}{c_\mu - r} (-1)^{m-r} = \begin{cases} (-1)^{m-c} & \text{if } k_\mu = 1 \\ 0 & \text{otherwise} \end{cases}$$

But if $k_\mu = 1$ then λ/μ is a skew hook h with m boxes and c_μ columns, hence $m - c_\mu = ll(h)$, the height of the skew hook h , and so 5.29 becomes

$$\chi^\lambda(\pi\sigma) = \sum_h (-1)^{ll(h)} \chi^{\lambda \setminus h}(\pi)$$

where then h are skew hooks of λ with m boxes. □

As stated in the beginning of this subsection the above theorem gives a recursive way of computing the character of an irreducible representation of S_n . This is done, using the notation of the above theorem, by first finding all possible skew hooks h with a_1 boxes of λ , computing their height and then removing them from λ , obtaining the diagrams $\lambda \setminus h$. Then we find all possible skew hooks with a_2 boxes of $\lambda \setminus h$, compute their height and remove them from $\lambda \setminus h$, and so on until we either can no longer remove rim hooks from the resulting diagrams with the required number of boxes, thus giving a contribution of 0 to the sum, or ending up removing all boxes resulting in what we define to be the character $\chi_{(0)}^{(0)} = 1$. The following example shows how this goes about.

Example 5.5.16. Let $\lambda = (4, 4, 3), \zeta = (5, 4, 2) \vdash 11$, we compute $\chi_{(5,4,2)}^{(4,4,3)}$. The process of recursively removing skew hooks can be visualized as a tree, seen in figure 5.10. We mark the skew hook in the respective diagrams with a $\times$ as in the previous example, thus seeing both required diagrams for the computation simultaneously.

We first remove from λ the possible skew hooks with five boxes. There are two skew hooks, h_1 seen on the left of figure 5.10 and h_2 on the right, with $ll(h_1) = -1$ and $ll(h_2) = 1$, for which $\mu_1 = \lambda \setminus h_1 = (4, 2)$ and $\mu_2 = \lambda \setminus h_2 = (3, 2, 1)$, resulting in the characters $\chi_{(4,2)}^{(4,2)}$ and $\chi_{(4,2)}^{(3,2,1)}$. For μ_2 one notes that now it's not possible to find a skew hook of it with four boxes, thus $\chi_{(4,2)}^{(3,2,1)} = 0$, for μ_1 there is one possible skew hook $h_{1,1}$ with four boxes, we have that $ll(h_{1,1}) = -1$ and with it we get that $\mu_{1,1} = \mu_1 \setminus h_{1,1} = (1, 1)$ with respective character $\chi^{(1,1)}$. Finally we note that there is one possible skew hook of $\mu_{1,1}$ with two boxes, namely $\mu_{1,1}$ itself, thus we get the character $\chi_{(0)}^{(0)} = 1$.

Per the Murnhagan-Nakayama rule the above observations produce

$$\chi_{(5,4,2)}^{(4,4,3)} = -\chi_{(4,2)}^{(4,2)} + \chi_{(4,2)}^{(3,2,1)} = -(-\chi_{(2)}^{(1,1)}) + 0 = -(-(-\chi_{(0)}^{(0)})) = (-1)^3$$

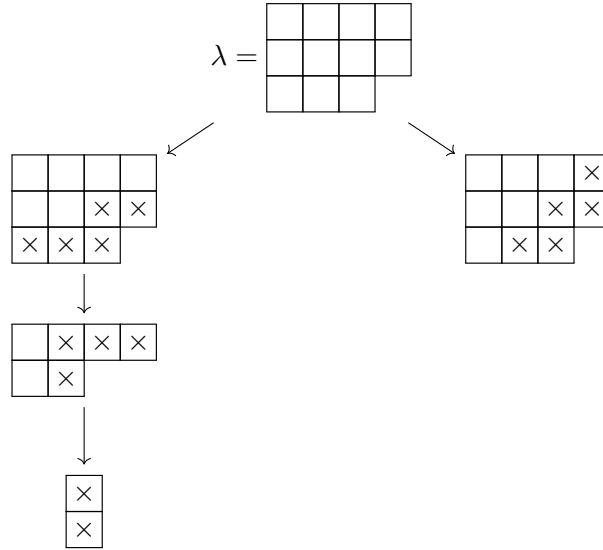


FIGURE 5.10: Graph of recursive skew hook elimination starting at the Ferrers diagram of $\lambda = (4, 4, 3)$, for the conjugation class $\zeta = (5, 4, 2)$.

The above example illustrates an useful characteristic of the Murnhagan-Nakayama rule. One notes that as the process of recursively finding and eliminating skew hooks always results in either 0 or $(-1)^k$ for some $k \in \mathbb{N}$, the character of an irreducible of S_n may be written as a sum of the form

$$\sum_{k \in T} (-1)^k$$

where T is some finite set of natural numbers. The trick to do this is labeling the boxes of the Ferrers diagram of λ according to the possible ways of removing skew hooks from it, per the Murnhagan-Nakayama rule, thus creating a combinatorial object to determine each k in T above.

Definition 5.5.17. A skew hook tableau t is a generalized tableau, as in definition 5.5.2, where the boxes are populated with positive integers such that

- a) Rows and columns of t weakly increase.
- b) All occurrences of an $i \in \mathbb{N}$ in t lie in a single skew hook.

For instance the following is a skew hook tableau

$$t = \begin{array}{|c|c|c|c|c|c|} \hline 1 & 1 & 2 & 2 & 3 & 3 \\ \hline 2 & 2 & 2 & 3 & 3 & \\ \hline 4 & 4 & 4 & 4 & & \\ \hline 4 & & & & & \\ \hline \end{array}$$

Looking at figure 5.10 we may produce a skew hook tableau t associated with the left path of the graph, seen below.

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

Thus one may easily see that the process of removing skew hooks from a Ferrers diagram determines a skew hook tableau, by labeling the last skew hook removed with a 1, the second to last one with a 2, and so on.

Definition 5.5.18. Given a skew hook tableau t let h_i be the skew hook within t whose boxes are populated with the number i . With this we define the sign of t , denoted $\text{sgn}(t)$, to be

$$\text{sgn}(t) = \prod_{h_i \in t} (-1)^{\text{ll}(h_i)}$$

Given $\zeta = (a_1, a_2, \dots, a_k), \lambda \vdash n$ we want the skew hook tableaux t of shape λ such that the skew hook populated with k has a_1 boxes, the skew hook populated with $k-1$ has a_2 boxes, and so on. The resulting skew hook tableau, if it exists, determines a path in the graph associated with the elimination of skew hooks from λ , and thus $\text{sgn}(t)$ is a summand of the character χ_ζ^λ as in the Murnhagan-Nakayama rule. With this the next result follows immediately.

Corollary 5.5.19. Let $\zeta = (a_1, a_2, \dots, a_k), \lambda \vdash n$, then

$$\chi_\zeta^\lambda = \sum_t \text{sgn}(t)$$

Where t are skew hook tableaux of shape λ and type $(a_k, a_{k-1}, \dots, a_1)$.

6. On the log rank conjecture - Part I

Now that we've become familiar with the representation theory of finite groups, and in particular the symmetric groups, part one of the Raz and Spieker paper comes quite easily. We recall that in this part of the paper the objective is to determine the rank of the matrix $M_{\mathfrak{F}}$ as in 2.2.5, with predicate $\mathfrak{F}$ as defined in 2.1, and as one might have guessed this will be done through the gained insight into the representation theory of symmetric groups.

Why the symmetric groups one might ask. As we've initially remarked on the proof of theorem 3.0.1, a perfect matching in the complete bipartite graph $K_{n,n}$ defines a bijection from $[n]$ to $[n]$, which in turn defines a permutation of $[n]$. Recalling then the set $\mathcal{M}$ of all perfect matchings on $K_{n,n}$ we have that as $\mathcal{M}$ can be seen as the set of all bijections on $[n]$ it can also be seen as the set of all permutations of $[n]$, which when equipped with the operation of composition of permutations becomes the symmetric group S_n . With this in mind we're able to redefine the predicate $\mathfrak{F}$ purely in terms of permutations, and consequently the same goes for the matrix $M_{\mathfrak{F}}$.

Let $X_1, X_2 \in \mathcal{M}$ be matchings with respective permutations $\sigma, \tau \in S_n$. We may consider as an example for $X_1 \cup X_2$ the graph in figure 6.1, edges of X_1 are in black and of X_2 in gray, thus we are considering $n = 5$.

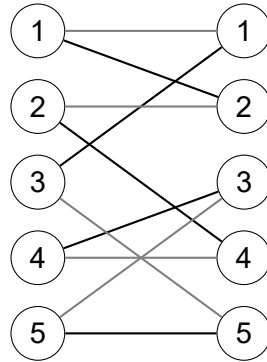


FIGURE 6.1: The union of two perfect matchings X_1 and X_2 of the graph $K_{5,5}$, edges of X_1 are in black and of X_2 in gray.

With this example we can see that the permutation associated with X_1 is $\sigma = (1\ 2\ 4\ 3)$ and with X_2 is $\tau = (3\ 5)$. Now say we want to see what cycle we get by starting at the vertex 1 on the left, then we just follow the path

$$1 \rightarrow \sigma(1) \rightarrow \tau^{-1}(\sigma(1)) \rightarrow \sigma(\tau^{-1}(\sigma(1))) \rightarrow \tau^{-1}(\sigma(\tau^{-1}(\sigma(1)))) \rightarrow \dots \quad (6.1)$$

until it repeats, which clearly must eventually occur. Now as far as determining if $X_1 \cup X_2$ is a Hamiltonian cycle all we have to do is see if the above path passes through all vertices on the left (this considering that on a cycle it does not matter where we start), for if that is the case then it must also pass through all vertices on the right, thus actually we need only to look at the sequence

$$1 \rightarrow \tau^{-1}(\sigma(1)) \rightarrow \tau^{-1}(\sigma(\tau^{-1}(\sigma(1)))) \rightarrow \dots \quad (6.2)$$

That is we need only to study the permutation $\tau^{-1}\sigma$, in particular if this is an n -cycle then the cycle in 6.2 passes through all vertices on the left, and so the cycle in 6.1 passes through all vertices of $K_{n,n}$. Indeed note that $\tau^{-1}\sigma = (1\ 2\ 4\ 5\ 3)$, thus $X_1 \cup X_1$ is a Hamiltonian cycle. In other words:

Proposition 6.0.1. *Given perfect matchings X_1, X_2 of $K_{n,n}$ with respective permutations $\sigma, \tau \in S_n$, $X_1 \cup X_2$ is a Hamiltonian cycle if and only if $\tau^{-1}\sigma$ is an n -cycle.*

Proof. If $\tau^{-1}\sigma$ contains a cycle of length $k < n$ then there would exist a cycle of length k in $X_1 \cup X_2$, thus a repetition of vertices and so it could not be a Hamiltonian cycle. On the other hand if $\tau^{-1}\sigma$ is an n -cycle then as σ and τ define a pair of perfect matchings on $K_{n,n}$, and $\tau^{-1}\sigma$ determines a path in $K_{n,n}$ that is a cycle and passes through all vertices only once, the union of the matchings is a Hamiltonian cycle. $\square$

For an example when the graph is not a Hamiltonian cycle, and thus $\tau^{-1}\sigma$ is not an n -cycle, consider the union of perfect matchings in figure 6.2. By it we have that $\tau^{-1}\sigma = (1\ 2)(3\ 4\ 5)$.

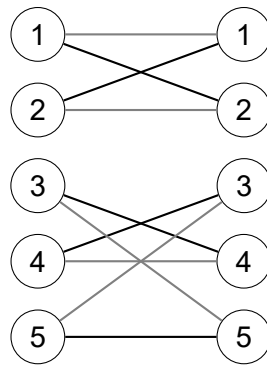


FIGURE 6.2: The union of two perfect matchings of the graph $K_{5,5}$, edges of X_1 are in black and of X_2 in gray.

Finally, as a permutation's cycle structure is not affected by conjugation in S_n [Rot95, 3.5] we have that $\tau^{-1}\sigma$ is an n -cycle if and only if $\tau\tau^{-1}\sigma\tau^{-1} = \sigma\tau^{-1}$ is an n -cycle (or any other

conjugate of $\tau^{-1}\sigma$ for that matter), hence by the above proposition and our remarks that $\mathcal{M}$ is essentially S_n we can now redefine the predicate $\mathfrak{F}$ as

$$\begin{aligned} \mathfrak{F} : S_n \times S_n &\rightarrow \{0, 1\} \\ (\sigma, \tau) &\mapsto \begin{cases} 1 & \text{if } \sigma\tau^{-1} \text{ is an } n\text{-cycle} \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{6.3}$$

With this we are ready to tackle the rank of $M_{\mathfrak{F}}$.

Theorem 6.0.2. *For the predicate $\mathfrak{F}$ above*

$$\text{rank}(M_{\mathfrak{F}}) = \binom{2n-2}{n-1}$$

Proof. The proof follows from a remarkable observation connecting the matrix $M_{\mathfrak{F}}$ to the representations of S_n . Consider the regular representation λ of S_n , as defined in 4.1.18, in its matrix form, so we have the representation

$$\begin{aligned} \lambda : S_n &\rightarrow \text{GL}_{n!}(\mathbb{C}) \\ \pi &\mapsto R_{\pi} \end{aligned}$$

where R_{π} are permutation matrices determined by the action of π , by left multiplication, on S_n , so $(R_{\pi})_{\sigma\tau} = 1$ if and only if $\pi\tau = \sigma$, that is $\pi = \sigma\tau^{-1}$. Note that given $\pi, \nu, \sigma, \tau \in S_n$ with $\pi \neq \nu$ we can't have that $(R_{\pi})_{\sigma\tau} = (R_{\nu})_{\sigma\tau} = 1$ for this would then mean that $\pi = \sigma\tau^{-1} = \nu$.

Let $C_{(n)}$ be the conjugacy class of S_n with representative $(1\ 2\ 3\ \dots\ n)$, that is $C_{(n)}$ is the class composed of all n -cycles of S_n . By 6.3 we have that $(M_{\mathfrak{F}})_{\sigma\tau} = 1$ if and only if $\sigma\tau^{-1} \in C_{(n)}$. Now if we consider the matrix

$$T = \sum_{\pi \in C_{(n)}} R_{\pi}$$

we have by the above that given entry $\sigma\tau$ only one of the matrices R_{π} can have a 1 in it, hence $T_{\sigma\tau} = 1$ if and only if $\sigma\tau^{-1} = \pi$ for some $\pi \in C_{(n)}$, or in other words $\sigma\tau^{-1} \in C_{(n)}$, which precisely matches the definition of $M_{\mathfrak{F}}$, hence

$$M_{\mathfrak{F}} = \sum_{\pi \in C_{(n)}} R_{\pi}$$

Crucially by this we have that $M_{\mathfrak{F}}$ defines an S_n -homomorphism, indeed given $\tau \in S_n$

$$R_{\tau} M_{\mathfrak{F}} R_{\tau^{-1}} = \sum_{\pi \in C_{(n)}} R_{\tau\pi\tau^{-1}} = \sum_{\pi \in C_{(n)}} R_{\pi} = M_{\mathfrak{F}}$$

as cycle structures are invariant under conjugation, that is $\tau C_{(n)} \tau^{-1} = C_{(n)}$. Moreover if we consider the decomposition of $\mathbb{C}^{n!}$ into irreducible S_n -spaces, $\mathbb{C}^{n!} = \bigoplus_i W_i$, then as these spaces are S_n -invariant, that is $R_\tau(W_i) \subseteq W_i$, $\tau \in S_n$, and $M_{\mathfrak{S}}$ is a sum of matrices R_τ , $\tau \in C_{(n)}$, we have that the spaces W_i are invariant for $M_{\mathfrak{S}}$, thus we may decompose $M_{\mathfrak{S}}$ as a sum of maps that are the restriction of the map $M_{\mathfrak{S}}$ to the spaces W_i [†]

$$M_{\mathfrak{S}} = \sum_i M_i$$

Since the maps M_i are just restrictions of $M_{\mathfrak{S}}$ they are also S_n -homomorphism, and as they are maps between the same irreducible S_n -spaces, that is $M_i : W_i \rightarrow W_i$, by part b) of Shur's lemma we have that $M_i = c_i \text{id}_{W_i}$ with $c_i \in \mathbb{C}$. Recalling that the rank of a matrix M is equal to the dimension of the image of the linear map it defines, that is if $M : V \rightarrow V$ then $\text{rank}(M) = \dim(M(V))$ [Gre75, 2.34], we have that

$$\text{rank}(M_{\mathfrak{S}}) = \sum_i \dim_{\mathbb{C}}(M_i(W_i))$$

As $M_i = c_i \text{id}_{W_i}$, $\dim_{\mathbb{C}}(M_i(W_i)) = 0$ if $c_i = 0$ and $\dim_{\mathbb{C}}(M_i(W_i)) = \dim_{\mathbb{C}}(W_i)$ otherwise, hence letting $d_i = \dim_{\mathbb{C}}(W_i)$

$$\text{rank}(M_{\mathfrak{S}}) = \sum_{c_i \neq 0} d_i$$

Now note that, as

$$M_{\mathfrak{S}} = \sum_{\pi \in C_{(n)}} R_\pi$$

we may use the decomposition of λ into irreducible subrepresentations, per the decomposition of $\mathbb{C}^{n!}$ above, to decompose M_i . As M_i is just the restriction of $M_{\mathfrak{S}}$ to W_i , letting $\lambda_i = \lambda_{W_i}$ [‡]

$$\text{tr}(M_i) = \sum_{\pi \in C_{(n)}} \text{tr}(\lambda_i(\pi)) = \sum_{\pi \in C_{(n)}} \chi_i(\pi)$$

where χ_i is the character of λ_i . Since $C_{(n)}$ is a conjugacy class and characters are class functions then

$$\text{tr}(M_i) = |C_{(n)}| \chi_i(\nu)$$

^{*}By $R_\tau(W_i)$ we mean the image of W_i by the map R_τ .

[†]To be precise, by restriction of the map $M_{\mathfrak{S}}$ to a space W_i we mean the matrix $M_i = M_{\mathfrak{S}} P_i$ where P_i is the projection of $\mathbb{C}^{n!}$ onto W_i .

[‡]Where recall λ_{W_i} is then the subrepresentation of λ defined by $\lambda_{W_i}(\pi) = R_\pi P_i$ where P_i is the projection of $\mathbb{C}^{n!}$ onto W_i .

where ν may be any in $C_{(n)}$. Hence as $\text{tr}(M_i) = \text{tr}(c_i \text{id}_{W_i}) = c_i d_i$ and $d_i, |C_{(n)}| > 0; c_i \neq 0$ if and only if $\chi_i(\nu) \neq 0$, therefore

$$\text{rank}(M_{\mathfrak{F}}) = \sum_{\chi_i(\nu) \neq 0} d_i$$

We now use our knowledge of representation theory of symmetric groups. To determine when $\chi_i(\nu) \neq 0$, as π is an n -cycle we may use lemma 5.5.1, χ_i being an irreducible character of S_n is associated to some partition μ of n , that is $\chi_i = \chi^\mu$, as by the mentioned lemma we have that $\chi^\mu(\nu) \neq 0$ if μ is a hook, as in definition 5.4.43, and 0 otherwise. Moreover we may use the hook-length formula, as in corollary 5.4.46 to determine the d_i .

First note that as the multiplicity of an irreducible representation in the decomposition into irreducibles of the regular representations equals its dimension, letting $d_i = d_\mu$ such that μ is the partition associated with χ_i as above, we have that

$$\text{rank}(M_{\mathfrak{F}}) = \sum_{\mu \text{ is a hook}} d_\mu^2$$

Now it's straightforward to see that there are n hooks of n , each with a unique height (5.4.43) in $[n-1]$, and given the hook t of height k , following the hook-length formula

$$d_\mu = \frac{n!}{\prod_{(i,j) \in \mu} h_\mu(i,j)} = \binom{n-1}{k}$$

where $h_\mu(i,j)$ is the hook-length of the box (i,j) in the hook t , as defined in 5.4.44. Hence we have that

$$\text{rank}(M_{\mathfrak{F}}) = \sum_{k=0}^{n-1} \binom{n-1}{k}^2$$

Finally we may use generation functions to better determine the right hand side of the above equation. Let $r_{n-1} = \sum_{k=0}^{n-1} \binom{n-1}{k}^2$ and

$$f_{n-1}(x) = \sum_{k=0}^{n-1} \binom{n-1}{k} x^k$$

Recall that $f_{n-1}(x) = (1+x)^{n-1}$ [CM09, 2.1.4]. We note that

$$f_{n-1}(x)^2 = \sum_{j=0}^{2n-2} \left(\sum_{k=0}^j \binom{n-1}{k} \binom{n-1}{j-k} \right) x^j$$

and when $j = n-1$

$$\sum_{k=0}^j \binom{n-1}{k} \binom{n-1}{j-k} = \sum_{k=0}^j \binom{n-1}{k}^2$$

Now as $f_{n-1}(x) = (1+x)^{n-1}$ then

$$f_{n-1}(x)^2 = (1+x)^{2n-2} = \sum_{j=0}^{2n-2} \binom{2n-2}{j} x^j$$

Therefore

$$r_{n-1} = \binom{2n-2}{n-1}$$

finishing the proof. □

Applying this to the communication complexity of $\mathfrak{F}$ we first note, and one may easily verify, that given $n, k \in \mathbb{N}$ with $k \leq n$

$$\binom{n}{k} < \left(\frac{en}{k}\right)^k$$

Therefore

$$\binom{2n-2}{n-1} < (2e)^{n-1}$$

And so

$$\lg(\text{rank}(M_{\mathfrak{F}})) = O(n)$$

Combining this with theorem 3.0.1 we have the final conclusion of the Raz and Spieker paper.

Theorem 6.0.3. *There exists a non-constant gap between the deterministic communication complexity of $\mathfrak{F}$ and the log of the rank of $M_{\mathfrak{F}}$.*

7. Conclusion

Starting with the second part of the Raz and Spieker paper, we note that, very much to our dismay, we were unable to find an intuition for why $\alpha = \lg(n)^{1/1000}$. The exact reason as to why α is set to this didn't transpire from the proof. Beyond that several attempts were made to modify the proof in order to strengthen the consequent result, to no avail.

For the first part of the paper, as stated in the introduction to this dissertation, it comes straightforwardly from the representation theory of symmetric groups.

As a final remark to the Raz and Spieker paper, one notes the creativity in the proofs. Noting, in the first part of the paper, that the matrix of the predicate may be described in terms of the regular representation of S_n was indeed a stroke of genius.

For the second part, the exploitation of the two found properties of the set Y , the use of the potential function coupled with the assumption that the set Y is large, and the overall inductive reasoning used, gave in the end a very rich proof.

For the Okounkov and Vershik paper, the key again is inductive reasoning, in particular that the symmetric groups form an inductive chain which allows "the theory to be structured inductively". Once the branching graph and the GZ-algebra and basis is found the theory then follows by considering classical results. Of course by this we don't mean to say the paper is just a small step ahead of previous results, quite the contrary.

Although built using an algebraic approach, we note that the theory may be presented using mostly results from linear algebra, as was done in the book [CSST10], and for this we congratulate the authors of said book. Yet presenting the theory as originally done by Okounkov and Vershik gives a richer approach to the theory of symmetric groups, and representation theory of finite groups in general.

Undeniably there's more work to be done here. The Raz and Spieker paper shows that representation theory can give powerful tools to determine the communication complexity of functions, and so a deep exploration for those tools and classes of functions to which they may be applied is in order. For instance functions that result from considering group actions, as is the case for the predicate in the Raz and Spieker paper, may be remarkably related to permutation representations.

Who knows, maybe one day, we'll see representation theory playing a role in the P vs NP problem.

Bibliography

- [AA05] A.M.Vershik and A.Yu.Okounkov. “A New Approach to the Representation Thoery of the Symmetric Groups. 2, Translated from Zapiski Nauchnykh Seminarov POMI, Vol. 307, 2004, pp. 57–98”. In: *J Math Sci* Vol. 131 (2005), pp. 5471—5494. DOI: <https://doi.org/10.1007/s10958-005-0421-7>.
- [AB09] Sanjeev Arora and Boaz Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, 2009. ISBN: 9780511804090.
- [AL06] Krishna B. Athreya and Soumendra N. Lahiri. *Measure Theory and Probability Theory*. Springer Texts in Statistics. Springer New York, NY, 2006. ISBN: 9780387329031.
- [Atk89] Kendall E. Atkinson. *An Introduction to Numerical Analysis, 2nd Edition*. John Wiley and Sons, Inc, 1989. ISBN: 9780471624899.
- [Ban08] Jason Bandlow. “An Elementary Proof of the Hook Formula”. In: *The Electronic Journal of Combinatorics* Volume 15 (2008), R45. DOI: <https://doi.org/10.37236/769>.
- [Bea99] John A. Beachy. *Introductory Lectures on Rings and Modules*. London Mathematical Society Student Texts (47). Cambridge University Press, 1999. ISBN: 9781139173315.
- [BH19] Joseph K. Blitzstein and Jessica Hwang. *Introduction to Probability*. Texts in Statistical Science. CHAPMAN and HALL/CRC, 2019. ISBN: 9781138369917.
- [Bre19] Matej Bresar. *Undergraduate Algebra: A Unified Approach*. Springer Undergraduate Mathematics Series. Springer Cham, 2019. ISBN: 9783030140533.
- [Chu16] Dexter Chua. *Representation Theory Lecture Notes based on the lectures by S.Martin, Mathematical Tripos Part-II, Cambridge University*. 2016. URL: https://dec41.user.srcf.net/notes/II_L/representation_theory.pdf.
- [CM09] Sebastian M. Cioaba and M.Ram Murty. *A First Course in Graph Theory and Combinatorics*. Texts and Readings in Mathematics. Hindustan Book Agency, India, 2009. ISBN: 9789386279392.
- [Cor+09] Thomas H. Cormen et al. *Introduction to Algorithms*. The MIT Press, 2009. ISBN: 9780262533058.

- [CSST10] Tullio Ceccherini-Silberstein, Fabio Scarabotti, and Filippo Tolli. *Representation Theory of the Symmetric Groups The Okounkov-Vershik Approach, Character Formulas, and Partition Algebras*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2010. ISBN: 9781139192361.
- [FBRT54] J. S. Frame, G. de B. Robinson, and R. M. Thrall. "The Hook Graphs of the Symmetric Group". In: *Canadian Journal of Mathematics* Volume 6 (1954), Pages 316–324. DOI: <https://doi.org/10.4153/CJM-1954-030-1>.
- [FH04] W. Fulton and J. Harris. *Representation Theory, A First Course*. Graduate Texts in Mathematics. Springer New York, NY, 2004. ISBN: 9780387975276.
- [Ful96] William Fulton. *Young Tableaux, With Applications to Representation Theory and Geometry*. London Mathematical Society Student Texts (35). Cambridge University Press, 1996. ISBN: 9780511626241.
- [GNW79] Curtis Greene, Albert Nijenhuis, and Herbert S Wilf. "A probabilistic proof of a formula for the number of Young tableaux of a given shape". In: *Advances in Mathematics* Volume 31, Issue 1 (1979), Pages 104–109. DOI: [https://doi.org/10.1016/0001-8708\(79\)90023-9](https://doi.org/10.1016/0001-8708(79)90023-9).
- [Gre75] Werner H. Greub. *Linear Algebra*. Graduate Texts in Mathematics. Springer-Verlag New York, NY, 1975. ISBN: 9783662006726.
- [Hal74] Paul R. Halmos. *Finite-Dimensional Vector Spaces*. Undergraduate Texts in Mathematics. Springer-Verlag New York Inc. 1958, 1974. ISBN: 9789896163723.
- [Hup98] Bertram Huppert. *Character Theory of Finite Groups*. De Gruyter Expositions in Mathematics Volume 25. De Gruyter, 1998. ISBN: 9783110809237.
- [Jam78] G. D. James. *The Representation Theory of the Symmetric Groups*. Lecture Notes in Mathematics. Springer-Verlag Berlin, 1978. ISBN: 9783540089483.
- [Juc74] A.-A.A. Jucys. "Symmetric polynomials and the center of the symmetric group ring". In: *Reports on Mathematical Physics* Volume 5, Issue 1 (1974), Pages 107–112. DOI: [https://doi.org/10.1016/0034-4877\(74\)90019-6](https://doi.org/10.1016/0034-4877(74)90019-6).
- [KMS15] Ida Kantor, Jirí Matousek, and Robert Sámal. *Mathematics++: Selected Topics Beyond the Basic Courses*. Student Mathematical Library 75. American Mathematical Society, 2015. ISBN: 9781470422615.
- [KN96] Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, 1996. ISBN: 9780511574948.

- [Lam98] T. Y. Lam. *Representations of Finite Groups : A Hundred Years , Part I*. 1998. URL: <https://www.ams.org/notices/199803/lam.pdf>.
- [LR34] D. E. Littlewood and A. R. Richardson. “Group characters and algebra”. In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* Volume 233 (1934), Pages 99–141. DOI: <https://doi.org/10.1098/rsta.1934.0015>.
- [Mac03] David J.C. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003. ISBN: 9780521642989.
- [Mar05] Paula M. Martins. *Um Curso de Álgebra Linear II*. número onze. Universidade do Minho, 2005. ISBN: 9728810105.
- [Meh16] David Mehrle. *Representation Theory Lecture Notes based on lecture by S.Martin, Mathematical Tripos Part-III, Cambridge University*. 2016. URL: https://pi.math.cornell.edu/~dmehrle/notes/partiii/reptheory_partiii_notes.pdf.
- [MS82] K. Mehlhorn and Erik M. Schmidt. “Las Vegas is better than determinism in VLSI and distributed computing (Extended Abstract)”. In: *STOC '82. Association for Computing Machinery*, 1982, pp. 330–337. ISBN: 0897910702.
- [Mur37] F. D. Murnaghan. “On the Representations of the Symmetric Group”. In: *American Journal of Mathematics* Volume 59, No. 3 (1937), Pages 437–488. DOI: <https://doi.org/10.2307/2371574>.
- [Mur81] G.E Murphy. “A new construction of Young’s seminormal representation of the symmetric groups”. In: *Journal of Algebra* Volume 69, Issue 2 (1981), Pages 287–297. DOI: [https://doi.org/10.1016/0021-8693\(81\)90205-2](https://doi.org/10.1016/0021-8693(81)90205-2).
- [Nak40] T. Nakayama. “On some modular properties of irreducible representations of a symmetric group, I”. In: *Japanese journal of mathematics* Volume 17 (1940), Pages 165–184. DOI: https://doi.org/10.4099/jjm1924.17.0_165.
- [NW94] N. Nisan and A. Wigderson. “On rank vs. communication complexity”. In: *Proceedings of 35th FOCS*. 1994, pp. 831–836.
- [Raz92] A A Razborov. “The gap between the chromatic number of a graph and the rank of its adjacency matrix is superlinear”. In: *Discrete Mathematics* Volume 108 (1992), Pages 393–396. DOI: [https://doi.org/10.1016/0012-365X\(92\)90691-8](https://doi.org/10.1016/0012-365X(92)90691-8).
- [Rot02] Joseph J. Rotman. *Advanced Modern Algebra*. Prentice Hall, 2002. ISBN: 0130878685.

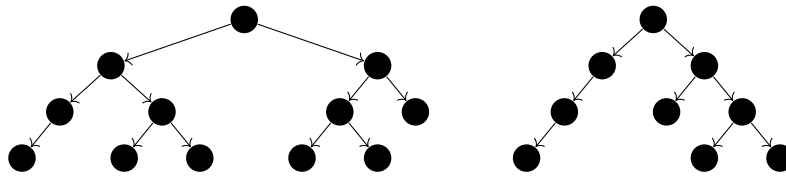
- [Rot95] Joseph J. Rotman. *An Introduction to the Theory of Groups*. Graduate Texts in Mathematics. Springer New York, NY, 1995. ISBN: 9781461241768.
- [RS95] R. Raz and B. Spieker. “On the “log rank”-conjecture in communication complexity”. In: *Combinatorica* Volume 15 (1995), Pages 567–588. DOI: <https://doi.org/10.1007/BF01192528>.
- [Sag01] Bruce E. Sagan. *The Symmetric Group; Representations, Combinatorial Algorithms, and Symmetric Functions*. Graduate Texts in Mathematics. Springer New York, NY, 2001. ISBN: 9781475768046.
- [Sip12] Michael Sipser. *Introduction to the Theory of Computation*. Cengage Learning, 2012. ISBN: 9781133187790.
- [SQ10] Ana P. Santos and João F. Queiró. *Introdução à Álgebra Linear*. Trajectos Ciência. Gradiva Publicações, 2010. ISBN: 9789896163723.
- [ST23] Benny Sudakov and István Tomon. *Matrix discrepancy and the log-rank conjecture*. 2023. URL: <https://arxiv.org/abs/2311.18524>.
- [Yao79] Andrew Chi-Chih Yao. “Some complexity questions related to distributive computing(Preliminary Report)”. In: Association for Computing Machinery, 1979. ISBN: 9781450374385.

A. Binary trees

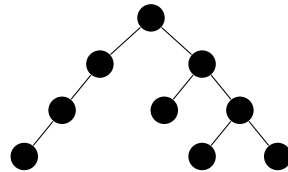
Binary trees are a fundamental structure in computer science, here we define this types of graphs and present a couple useful results about them.

Definition A.0.1. *A binary tree is a directed acyclical graph, that is a directed graph with no cycles in it, where each node as one ingoing edge, with the exception of a single node called the root node, and at most two outgoing edges. The nodes that have no outgoing edges are called the leaves of the tree.*

Example A.0.2. As a binary tree is perhaps best seen by example, below we present two such examples.

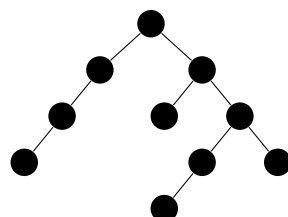


In both examples we can see the root node on top, and the binary tree on the left has 6 leaves while the one on the right has 4. One usually omits the direction of the edges, being then implied that one can only go down from root to leaves and never traverse an edge in the upwards direction, thus we'd represent, for instance the binary tree above on the right, as



Definition A.0.3. *The height of a binary tree is the length of it longest path from root to a leave, where by length we mean the number of edges traversed.*

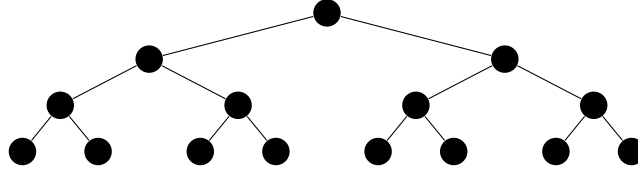
For instance in the above example both trees had height equal to 3, the tree below has height 4.



A particular type of binary tree is of interest to us.

Definition A.0.4. A full binary tree is a binary tree where all nodes have two outgoing edges, with the exception of the leaves. We usually specify what full binary tree we are referring to by stating its number of leaves or its height.

Example A.0.5. The following is a full binary tree with 8 leaves



The reason we may specify a full binary tree by its height or number of leaves is of course that each of these uniquely determine the tree. Indeed one may easily verify that if a full binary tree has k leaves then it must have height equal to $\lg(k)$, and from this we can construct the required tree by simply adding nodes until we reach the required height, in a way that satisfies the definition of a full binary tree above. Note also that we can easily get the number of leaves from the height of the full binary tree.

An interesting property of binary trees is that given a node v in it there exist a single path from the root to v , indeed if there were two then there would exist a node with two ingoing edges, which is never the case, thus we have the following.

Proposition A.0.6. Given a binary tree and a node v in it there exist one and only one path from the root of the tree to v .

Finally we have the following useful result.

Theorem A.0.7. Let T be a binary tree with k leaves, then its height h satisfies

$$h \geq \lg(k)$$

Proof. Given a binary tree T with k leaves let it have height h . We may add nodes to the tree T so that it becomes a full binary tree T' with height h , yet T' will then have 2^h leaves. Now note that $2^h \geq k$ as the best way to maximize the number of leaves in a binary tree is to maximize the number of nodes with two outgoing edges, thus

$$h = \lg(2^h) \geq \lg(k)$$

□

B. Entropy

Shannon entropy is widely used in proving lower bound for the communication complexity of functions, and is for this extensively used in the paper [RS95] here revisited. Originally introduced by *Claude Shannon* in his acclaimed 1948 paper "A Mathematical Theory of Communication", to which the origin of Information Theory is attributed, it quantifies the average information content of a discrete random variable as the (average) number of bits required to describe an outcome of said random variable. We proceed to make this clearer.

For what follows we recall the definition of probability space and discrete random variable, one may properly review these concept in [BH19] and [AL06]. For our purposes we may restrict ourselves to the finite case.

Definition B.0.1. A (finite) probability space is a triple $\Upsilon = (\Omega, \mathcal{F}, P)$ where Ω is a finite set called the sample space, whose elements are called outcomes, $\mathcal{F}$ is subset of the power set of Ω , that is $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ *, whose elements are called events, and $P : \mathcal{F} \rightarrow [0, 1]$ is the probability function, such that

- If $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$, where A^c is the complement of A in Ω .
- $A_1, A_2 \in \mathcal{F} \Rightarrow A_1 \cup A_2 \in \mathcal{F}$.
- For all $A \in \mathcal{F}$, $P(A) \in [0, 1]$, $P(\emptyset) = 0$ and $P(\Omega) = 1$.
- For $A_1, A_2 \in \mathcal{F}$ with $A_1 \cap A_2 = \emptyset$ then $P(A_1 \cup A_2) = P(A_1) + P(A_2)$.

We additionally define a random variable on $(\Omega, \mathcal{F}, P)$ as a function $X : \Omega \rightarrow \mathbb{R}$, with $\mathbb{R}$ the set of real numbers, so that given $A = \{a_1, a_2, \dots, a_n\} \in \mathcal{F}$

$$P(A) = P(\{X(a_i) \mid a_i \in A\})$$

With this we may now proceed to define the *Shannon information content of an outcome*. From now on we refer to outcomes of the set Ω directly by its image by the random variable, that is we let $x = X(a)$ for $a \in \Omega$ and $X = X(\Omega)$

Definition B.0.2. Given a probability space Υ as in the above definition and an outcome $x \in X$ we define the *Shannon information content of x* as

$$h(x) = \lg \left(\frac{1}{P(x)} \right)$$

*The power set of a set A , denoted $\mathcal{P}(A)$, is the set composed of all subsets of A .

Moreover we define the entropy of X as

$$H(X) = \sum_{x \in X} P(x) h(x) = \sum_{x \in X} P(x) \lg \left(\frac{1}{P(x)} \right)$$

Both the information content and the entropy are measured in bits. As $P(x) \in [0, 1]$ we have that $H(X) \geq 0$ with equality if and only if $P(x) = 1$ for one $x \in X$. The following property of the entropy is what makes it so useful to us in communication complexity theory.

Proposition B.0.3. *Given a probability space Υ*

$$H(X) \leq \lg(|X|)$$

With equality if and only if $P(x) = 1/|X|$ for all $x \in X$.

The above proposition follows directly from the fact that logarithmic functions are concave, thus given $\lambda_i \in [0, 1]$, $i \in [n]$, such that $\sum_{i \in [n]} \lambda_i = 1$

$$\sum_{i \in [n]} \lambda_i \log(x_i) \leq \log \left(\sum_{i \in [n]} \lambda_i x_i \right)$$

And it essentially says that entropy is maximized when the distribution is uniform, moreover it allows us to lower bound the size of X with its entropy. This results is usually called the *entropy bound*.

Example B.0.4. Let $X = \{0, 1, 2\}$ and consider the distribution on X following from an experiment where we first flip a coin to determine if $x = 0$, say if heads comes out, and then if x is not 0 we flip a coin again to determine if $x = 1$ or $x = 2$. We get that

$$P(x = 0) = \frac{1}{2}; \quad P(x = 1) = \frac{1}{4}; \quad P(x = 2) = \frac{1}{4}$$

Then entropy of X is

$$H(X) = \frac{1}{2} \lg(2) + 2 \frac{1}{4} \lg(4) = \frac{1}{2} + 1 = \frac{3}{2}$$

The entropy of a probability space may be decomposed into a liner combination of entropies, for instance in the previous example B.0.4 we note that

$$H(X) = \frac{1}{2} \lg(2) + \frac{1}{2} \lg(2) + \frac{1}{2} \left(\frac{1}{2} \lg(2) + \frac{1}{2} \lg(2) \right) = H(X_1) + \frac{1}{2} H(X_2)$$

Where X_1 and X_2 are then random variables with $|X_1| = |X_2| = 2$ uniformly distributed. This may be seen intuitively by noting that in our experiment we first learn if $x = 0$ or not, with probability $1/2$, thus we may attribute to this X_1 , and then if x is not 0 we learn if x is 1 or 2, to which we may attribute X_2 , But as X_2 will only be realized half the time due to the probability of x not being 0 being $1/2$ we get that

$$H(X) = H(X_1) + \frac{1}{2}H(X_2)$$

More generally, and replacing $X = \{x_1, x_2, \dots, x_n\}$ with $P(X) = \{p_1, p_2, \dots, p_n\}$ to simplify notation, we have that

$$H(\{p_1, p_2, \dots, p_n\}) = H(p_1, 1 - p_1) + (1 - p_1)H\left(\left\{\frac{p_2}{1 - p_1}, \frac{p_3}{1 - p_1}, \dots, \frac{p_n}{1 - p_1}\right\}\right)$$

Or we may say that, for $k \in [n]$

$$\begin{aligned} H(\{p_1, p_2, \dots, p_n\}) &= H((p_1 + p_2 + \dots + p_k), (p_{k+1} + p_{k+2} + \dots + p_n)) + \\ &+ (p_1 + p_2 + \dots + p_k)H\left(\left\{\frac{p_1}{p_1 + p_2 + \dots + p_k}, \dots, \frac{p_k}{p_1 + p_2 + \dots + p_k}\right\}\right) + \\ &+ (p_{k+1} + p_{k+2} + \dots + p_n)H\left(\left\{\frac{p_{k+1}}{p_{k+1} + p_{k+2} + \dots + p_n}, \dots, \frac{p_n}{p_{k+1} + p_{k+2} + \dots + p_n}\right\}\right) \end{aligned}$$

Given two probability spaces Υ_1, Υ_2 we may talk about their joint probability space $\Upsilon = (\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2, P)$, refer to [BH19] for more on this. With this in mind we define the entropy of a joint probability space.

Definition B.0.5. Given two probability spaces Υ_1, Υ_2 we define their joint entropy as

$$H(X, Y) = \sum_{(x,y) \in X \times Y} P(x, y) \lg\left(\frac{1}{P(x, y)}\right)$$

The entropy is additive if the random variables X and Y are independent, that is if $P(x, y) = P(x)P(y)$ then

$$H(X, Y) = H(X) + H(Y)$$

We may extend entropy to the concept of conditional probability, recall that given two random variables X, Y , not necessarily distinct and two events A and B in their respective probability spaces, we have that

$$P(A|B) = \frac{P(A, B)}{P(B)} \tag{B.1}$$

Definition B.0.6. The conditional entropy of the random variable X given random variable Y is

$$H(X|Y) = \sum_{y \in Y} P(y) \left(\sum_{x \in X} P(x|y) \lg \left(\frac{1}{P(x|y)} \right) \right) = \sum_{(x,y) \in X \times Y} P(x,y) \lg \left(\frac{1}{P(x|y)} \right)$$

Moreover we define the entropy of X given an outcome y of Y as

$$H(X|y) = \sum_{x \in X} P(x|y) \lg \left(\frac{1}{P(x|y)} \right)$$

From equation B.1 one notes that $P(A, B) = P(B)P(A|B)$, this is usually referred to the chain rule for probability, results in what is called the chain rule for information content and for entropy. For information content applying this identity gives

$$h(x, y) = h(x) + h(y|x)$$

where of course $h(y|x) = \lg(1/P(y|x))$, thus for entropy one may verify that

$$H(X, Y) = H(X) + H(Y|X) = H(Y) + H(X|Y) \quad (\text{B.2})$$

This may then be easily generalized to

$$H(X_1, \dots, X_n) = H(X_1) + H(X_2 | X_1) + \dots + H(X_n | X_1, X_2, \dots, X_{n-1}) \quad (\text{B.3})$$

by generalizing the chain rule for probability.

Finally an useful property of conditional entropy is as follows.

Proposition B.0.7. Given two probability spaces Υ_1, Υ_2

$$H(X | Y) \leq H(X)$$

The proof of this result may be found in [Mac03, p.142]. With this proposition one may also easily verify that

$$H(X | Y_1, Y_2) \leq H(X | Y_1) \quad (\text{B.4})$$

by noting that $H((X | Y) | Z) = H(X | Y, Z)$.

C. Linear Algebra

As we assume the reader has a basic knowledge of linear algebra (definition of finite dimensional vector space, subspaces, existence of basis, linear transformations, etc...) this will be a short exposition of some basic results used throughout. We refer the reader to [Hal74] and [Gre75] for an extensive treatment of the subject. In what follows we always assume V is a finite dimensional vector space.

Recall the conjugation map on the field $\mathbb{C}$ of complex numbers.

$$\begin{aligned}\bar{\cdot} : \mathbb{C} &\rightarrow \mathbb{C} \\ \alpha + i\beta &\mapsto \alpha - i\beta\end{aligned}$$

When restricted to the field $\mathbb{R}$ of real numbers this map is simply the identity map, more generally conjugation can be defined on any field $\mathbb{F}$, yet we need not to worry about it as we focus on the field $\mathbb{C}$.

Definition C.0.1. Let V be a vector space over $\mathbb{C}$, an inner product on V is a map

$$\begin{aligned}\langle \cdot, \cdot \rangle : V \times V &\rightarrow \mathbb{C} \\ (v, w) &\mapsto \langle v, w \rangle\end{aligned}$$

Such that, for all $v, w, u \in V$

- $\langle v, v \rangle \geq 0$
- $\langle v, v \rangle = 0 \Leftrightarrow v = 0$
- $\langle \alpha v + \beta u, w \rangle = \langle \alpha v, w \rangle + \langle \beta u, w \rangle$
- $\langle v, w \rangle = \overline{\langle w, v \rangle}$

For a general field $\mathbb{F}$ the existence of an inner product on the vector space V over $\mathbb{F}$ is not guaranteed, yet when $\mathbb{F} = \mathbb{C}$ it exists.

Definition C.0.2. Let V be a vector space over $\mathbb{C}$ and $B = (e_1, e_2, \dots, e_n)$ a basis of V . Then for any $v \in V$ there exist $\alpha_i \in \mathbb{C}$, $i \in [n]$, such that

$$v = \sum_{i \in [n]} \alpha_i e_i$$

We define an inner product on V as, for $v = \sum_{i \in [n]} \alpha_i e_i$ and $w = \sum_{i \in [n]} \beta_i e_i$

$$\langle v, w \rangle = \sum_{i \in [n]} \alpha_i \overline{\beta_i}$$

It's straightforward to verify that this indeed defines an inner product on V .

Definition C.0.3. If V is a vector space equipped with an inner product then we say V is an inner product space/unitary space.

When dealing with inner product spaces we can define another binary map on V .

Definition C.0.4. If V is an inner product space with inner product $\langle \cdot, \cdot \rangle$ then we define the following map as a norm on V

$$\|v\| = \sqrt{\langle v, v \rangle}$$

with $v \in V$.

By our remarks in the beginning of Section 4.1 any linear map θ on a vector space V can be represented by a square matrix T_θ once a basis of V is chosen, the matrix coefficients of T_θ can be obtained using the inner product defined on V (if it exists of course). We recall that an orthonormal basis $B = (e_1, e_2, \dots, e_n)$ of an inner product vector space V is such that

- $\langle e_i, e_j \rangle = 1$ if $i = j$ and 0 otherwise.
- $\|e_i\| = 1$ for all $i \in [n]$.

Proposition C.0.5. Let V be an inner product space of dimension n and $\theta : V \rightarrow V$ a linear map, then for an orthonormal basis $B = (v_1, v_2, \dots, v_n)$ of V the corresponding matrix of θ , T_θ is given by

$$T_\theta = (\langle T e_i, e_j \rangle)_{j, i \in [n]}$$

Proof. Indeed if B is an orthonormal basis then if we let $T e_i = \sum_{i \in [n]} \alpha_i e_i$ then by linearity in the first argument of an inner product

$$\langle T e_i, e_j \rangle = \sum_i \alpha_i \langle e_i, e_j \rangle = \alpha_j$$

Thus this gives precisely the matrix T_θ as discussed in the beginning of Section 4.1, and so it is a matrix representation of the map θ . □

We now define the adjoint of a linear map and unitary linear maps.

Definition C.0.6. Let V be an inner product space and $T : V \rightarrow V$ a linear map, then the adjoint of T , denoted T^* is defined as the map that satisfies the equality

$$\langle T(v), w \rangle = \langle v, T^*(w) \rangle$$

for all $v, w \in V$.

Proposition C.0.7. If $T : V \rightarrow V$ is a linear map on the vector space V then $(T^*)^* = T$

Proof. Indeed if

$$\langle T(v), w \rangle = \langle v, T^*(w) \rangle$$

Then for $(T^*)^*$ we have

$$\langle T^*(w), v \rangle = \langle w, (T^*)^*(v) \rangle$$

But

$$\langle T^*(w), v \rangle = \overline{\langle v, T^*(w) \rangle} = \overline{\langle T(v), w \rangle} = \langle w, T(v) \rangle$$

hence

$$\langle (T^*)^*(v), w \rangle = \langle T(v), w \rangle$$

As this is true for any $v, w \in V$ then by the previous proposition T and $(T^*)^*$ are the same linear maps as they will have the same matrix representations over an orthonormal basis. □

Recall the definition of transposition of matrix and complex conjugate of a matrix (which essentially conjugates every coefficient of the matrix), the combination of these two operations is called the conjugate transpose of a matrix.

Proposition C.0.8. If T is the matrix of a linear map on V , then the conjugate transpose of T is the matrix of the adjoint map.

Proof. This is clear as

$$\langle T^*(v), w \rangle = \overline{\langle T(w), v \rangle}$$

□

Definition C.0.9. Let V be an inner product space, a linear map $T : V \rightarrow V$ is said to be unitary if

$$\langle T(v), T(w) \rangle = \langle v, w \rangle$$

for all $v, w \in V$.

Hence an unitary linear map preserves inner products. If we now consider the adjoint of T then note that if T is unitary then

$$\langle v, w \rangle = \langle T(v), T(w) \rangle = \langle T^*T(v), w \rangle$$

Proposition C.0.10. *If $T : V \rightarrow V$ is an unitary map then $T^* = T^{-1}$.*

Proof. Indeed by what we said above, for all $v, w \in V$ $\langle v, w \rangle = \langle T^*T(v), w \rangle$ And so by proposition C.0.5 $T^*T = id_V$. □

When a linear map is injective it displays a very useful property.

Proposition C.0.11. *Let $T : V \rightarrow W$ be an injective linear map from a vector space V to a vector space W , then it preserves linear independence. That is if $B = (v_1, v_2, \dots, v_n)$ is a linearly independent set in V then $T(B)$ is also a linearly independent set in W .*

Proof. Indeed let $B_V = (v_1, v_2, \dots, v_n)$ be a linearly independent set in V and consider its image by T , $B_W = (T(v_1), T(v_2), \dots, T(v_n))$, then if there exist $\alpha_i \in \mathbb{C}$, $i \in [n]$, not all zero such that

$$\sum_{i \in [n]} \alpha_i T(v_i) = 0$$

Then by linearity of T

$$T \left(\sum_{i \in [n]} \alpha_i v_i \right) = 0$$

As T is injective it must be the case that $\sum_{i \in [n]} \alpha_i v_i = 0$ which is a contradiction, hence all α_i must be equal to zero and so B_W is a linearly independent set in W . □

We present the following results without proof, referring the reader to [Gre75].

Proposition C.0.12. *Let V be a vector space over $\mathbb{C}$ and $T : V \rightarrow V$ a linear map, if there exists a polynomial f (with coefficients in $\mathbb{C}$) such that $f(T) = 0$ and f factors into distinct linear factors then T is diagonalizable.*

This result follows from the fact that if $f(T) = 0$ then the characteristic polynomial of T divides f and so if f factors into distinct linear factors then all eigenvalues of T are distinct. With that said the following proposition is clear.

Proposition C.0.13. *Let V be a vector space over $\mathbb{C}$, $T : V \rightarrow V$ a linear map and f a polynomial with coefficients in $\mathbb{C}$ such that $f(T) = 0$, then the eigenvalues of T are roots of f .*

Corollary C.0.14. *Let V be a vector space over $\mathbb{C}$ and $T : V \rightarrow V$ a linear map, if $T^n = id_V$ for some $n \in \mathbb{N}$ minimal with this property then the eigenvalues of T are n th-roots of unity. In particular if $n = \dim_{\mathbb{C}}(V)$ then the eigenvalues of T are all the n th-roots of unity.*

D. Modules and Algebras

D.1. Module theory

We'll use module theory extensively to develop a theory of G -spaces, that is vector spaces equipped with a linear group action, and consequently representations. As such the reader needs to get acquainted with fundamental results in module theory, moreover modules are fundamental to understand our exposition of the Okounkov-Vershik Approach in Chapter 5. Here we'll give the required results and definitions not focusing on proofs, for a more detailed exposition refer to [Bea99] and [Rot02]. The information given here is based on [Bea99].

For what follows recall the definitions of noncommutative rings and ideals.

D.1.1 Modules

Cayley's theorem 4.0.1 can be easily generalized to rings. We recall that letting G be an abelian group (thus we refer to the operation as addition) then the set of endomorphisms of G , $\text{End}(G)$, is a ring under addition and composition of maps.

Theorem D.1.1. *Given a ring R there exists an abelian group G such that R is isomorphic to a subring of the ring of endomorphisms of G , that is $\text{End}(G)$.*

The proof is analogous to Cayley's theorem, by letting G be the underlying additive group in R we can define for each $r \in R$ a map $\lambda_r : R \rightarrow R$ such that $\lambda_r(s) = rs$. It's then straightforward to see that the map $\rho : R \rightarrow \text{End}(R)$ such that $r \mapsto \lambda_r$ is an injective ring homomorphism thus producing an embedding of R into $\text{End}(R)$.

From this we note that the map ρ above produces in a natural way an action of R on R , and more generally given a homomorphism $\rho : R \rightarrow \text{End}(G)$ with G an abelian group then we get an action of R on G as, for $r \in R$ and $g \in G$

$$r \cdot g = [\rho(r)](g)$$

This follows our line of thought in the beginning of Subsection 4.1.1. The map ρ is then called a representation of the ring R . As G is abelian the operation on it is denoted as addition.

Considering ρ is a ring homomorphism we get a few interesting properties on the resulting action of R on G . Let $r_1, r_2 \in R$ and $g_1, g_2 \in G$, then

- $1_R \cdot g_1 = g_1$
- $r \cdot (g_1 + g_2) = r \cdot g_1 + r \cdot g_2$
- $(r_1 + r_2) \cdot g_1 = r_1 \cdot g_1 + r_2 \cdot g_1$
- $r_1 \cdot (r_2 \cdot g_1) = (r_1 r_2) \cdot g_1$

This motivates the following definition.

Definition D.1.2. Let R be a ring and M an abelian group. If there exists an action of R on M

$$\begin{aligned} \mu : R \times M &\rightarrow M \\ (r, m) &\mapsto r \cdot m \end{aligned}$$

such that for $r_1, r_2 \in R$ and $g_1, g_2 \in M$

- a) $1_R \cdot g_1 = g_1$
- b) $r \cdot (g_1 + g_2) = r \cdot g_1 + r \cdot g_2$
- c) $(r_1 + r_2) \cdot g_1 = r_1 \cdot g_1 + r_2 \cdot g_1$
- d) $r_1 \cdot (r_2 \cdot g_1) = (r_1 r_2) \cdot g_1$

Then M is called a *left R -module* (considering R acts on G on the left) and it's denoted by ${}_R M$ or simply M if it's clear from context.

If the action of R on M was on the right, that is $\mu : M \times R \rightarrow M$ then we'd have a *right R -module*, denoted M_R .

Note that the map μ is essentially a scalar product of R on M and thus an R -module is a generalization of a vector space where instead of using scalar products with fields we relax this to using rings, in particular any vector space V over a field $\mathbb{F}$ is a left $\mathbb{F}$ -module.

Example D.1.3. When we consider the action of R on itself by left multiplication (as discussed after theorem D.1.1) then we get what is called the *regular left R -module*, denoted ${}_R R$. We could equivalently produce the regular right R -module by considering the action of R on itself by right multiplication, denoted R_R .

We note that the regular left/right R -module coincides with the original ring R as the action we consider is already the product operation on the ring R so we don't gain anything new here, thus ${}_R R = R = R_R$.

Given an R -module M it's straightforward to define a submodule.

Definition D.1.4. *Let M be an R -module, then N is a submodule if it's an R -module itself when equipped with the action of R on M restricted to N .*

It's easy to prove that a subset N of M is a submodule if and only if it's non-empty, closed form sums and closed for the action of R defined on M . A submodule of ${}_R R$ is called a left ideal of R and likewise a submodule of R_R is called a right ideal of R , it's straightforward to see this is consistent with the definition of left and right ideals of a noncommutative ring. It's then clear that a subset of R is an ideal if and only if it's both a left and right ideal of R .

For an element m in M we can consider the smallest submodule of M that contain m , denoted Rm . It's easy to see this module is then

$$Rm = \{rm \mid r \in R\}$$

justifying the notation.

Definition D.1.5. *A left module M is called cyclic if $M = Rm$ for some $m \in M$.*

Given N a submodule of ${}_R M$ we can define the quotient R -module M/N . If we consider the quotient group M/N (which is indeed a group as M is abelian and so N is a normal subgroup) then as N is an R -submodule it is closed for the action of R and so for all $r \in R$, $r \cdot N = N$ thus we define an action of R on M/N using the action of R on M as, for $r \in R$ and $m + N \in M/N$

$$r(m + N) = r \cdot m + N$$

Now that we have defined a new algebraic structure we now need to define maps between them that preserve their operations, that is a class of morphisms between them.

Definition D.1.6. Let M, N be left R -modules, then a function $f : M \rightarrow N$ is said to be an R -homomorphism if it is a group homomorphism and for $r \in R$ and $m \in M$

$$f(rm) = r(f(m))$$

We denote the set of R -homomorphism from M to N as $\text{Hom}_R(M, N)$. If $N = M$ then we say f is an R -endomorphism of M and denote the set of R -endomorphism of M as $\text{End}_R(M)$. If f is both injective and surjective we say f is an R -isomorphism and if f is a bijective R -endomorphism then we say f is an R -automorphism.

We define the kernel of an R -homomorphism to be the kernel of the underlying group homomorphism, that is

$$\ker(f) = \{m \in M \mid f(m) = 0\}$$

The group isomorphism theorem extends easily to modules.

Theorem D.1.7. Let M, N be left R -modules and $f : M \rightarrow N$ an R -homomorphism, then

$$f(M) \cong M / \ker(f)$$

The proof is analogous to the one known for groups, one need only to verify that the homomorphism defined is an R -homomorphism using the fact that f is one.

We now define and characterize simple modules. We denote the trivial module, that is a module where the group M is composed only of the identity, by (0) .

Definition D.1.8. A non-zero module ${}_R M$ is said to be simple or irreducible if its only submodules are (0) and M itself.

Proposition D.1.9. Let M be a left R -module, then M is simple if and only if $M = Rm$ for any non-zero $m \in M$.

Proof. If $0 \neq m \in M$, Rm is a non-zero submodule of M hence if M is simple considering $m \neq 0$ then $Rm = M$. On the other hand if N is a non-zero submodule of M then there exists $0 \neq n \in N$, thus $M = Rn \subseteq N \subseteq M$ by assumption and so $N = M$, thus M must be simple. □

D.1.2 Direct sums and products

As is known any subspace of a vector space has a complement, thus one can decompose a vector space into a direct sum of two subspaces. One now wonders what can be said in this regard about modules.

Definition D.1.10. Let M, N be left R -modules, we define their direct sum as the set

$$M \oplus N = \{(m, n) \mid m \in M, n \in N\}$$

with addition as $(m_1, n_1) + (m_2, n_2) = (m_1 + m_2, n_1 + n_2)$ and action by R as $r(m, n) = (rm, rn)$. We refer to these operations as component wise addition and scalar multiplication respectively.

The Cartesian product of modules $\{M_\alpha\}_{\alpha \in I}$ can be described as the set of function $m : I \rightarrow \cup_{\alpha \in I} M_\alpha$ where for each $\alpha \in I$, $m(\alpha) \in M_\alpha$. The element $m(\alpha)$ is then usually denoted as m_α .

Definition D.1.11. Let $\{M_\alpha\}_{\alpha \in I}$ be a collection of left R -modules indexed by the set I .

a) The direct product of the modules $\{M_\alpha\}_{\alpha \in I}$ is the Cartesian product

$$\prod_{\alpha \in I} M_\alpha$$

With component wise addition and scalar multiplication by R .

a) The submodule of $\prod_{\alpha \in I} M_\alpha$ consisting of elements m such that $m_\alpha = 0$ for all but finitely many components m_α is called the direct sum of the modules $\{M_\alpha\}_{\alpha \in I}$, and is denoted by

$$\bigoplus_{\alpha \in I} M_\alpha$$

It's straightforward to prove that $\prod_{\alpha \in I} M_\alpha$ and $\bigoplus_{\alpha \in I} M_\alpha$ are indeed left R -modules. If the set of indices I is finite then the direct sum coincides with the direct product.

For each index $\iota \in I$ we define the projection map

$$p_\iota : \prod_{\alpha \in I} M_\alpha \rightarrow M_\iota$$

$$m \mapsto m_\iota$$

and the inclusion map

$$i_\iota : M_\iota \rightarrow \prod_{\alpha \in I} M_\alpha$$

$$m_\iota \mapsto m$$

where m has components equal 0 in all indices except in the index ι where it has m_ι . These are both R -homomorphism as can be easily seen, moreover the image of an inclusion map is clearly contained in $\bigoplus_{\alpha \in I} M_\alpha$ and so in reality $i_\iota : M_\iota \rightarrow \bigoplus_{\alpha \in I} M_\alpha$

Proposition D.1.12. *Let $\{M_\alpha\}_{\alpha \in I}$ be a collection of left R -modules indexed by the set I , N a left R -module and consider the projection and inclusion maps defined above, then*

- a) *For any set $\{f_\alpha\}_{\alpha \in I}$ of R -homomorphisms such that $f_\alpha : N \rightarrow M_\alpha$ for each $\alpha \in I$ there exists a unique R -homomorphism $f : N \rightarrow \prod_{\alpha \in I} M_\alpha$ such that $p_\alpha f = f_\alpha$ for each $\alpha \in I$.*
- b) *For any set $\{f_\alpha\}_{\alpha \in I}$ of R -homomorphisms such that $f_\alpha : M_\alpha \rightarrow N$ for each $\alpha \in I$ there exists a unique R -homomorphism $f : \bigoplus_{\alpha \in I} M_\alpha \rightarrow N$ such that $f i_\alpha = f_\alpha$ for each $\alpha \in I$.*

The proof of the above proposition is straightforward, one needs only to construct the most obvious R -homomorphisms that satisfy the required conditions and prove their uniqueness.

Direct sums can be seen from two equivalent perspectives. On one hand we have the perspective as in definition D.1.10 where we consider its elements as n -tuples, this is called an *outer direct sum*.

On the other hand if we consider an R -module M with two submodules M_1, M_2 such that $M_1 \cap M_2 = (0)$ and $M = M_1 + M_2$, that is any element of M can be written as a sum of an element of M_1 and an element of M_2 , then $M \cong M_1 \oplus M_2$, indeed we first note that any element of M can be written uniquely as sum of an element of M_1 and an element of M_2 and if for $m \in M$, $m = m_1 + m_2$ and $m = m'_1 + m'_2$ with $(m_1, m_2) \neq (m'_1, m'_2)$ and $m_1, m'_1 \in M_1, m_2, m'_2 \in M_2$ then

$$M_1 \ni m_1 - m'_1 = m'_2 - m_2 \in M_2$$

hence, as $(m_1, m_2) \neq (m'_1, m'_2)$, it must be the case that $M_1 \cap M_2 \neq (0)$ which is a contradiction. Therefore we can map any element $m \in M$ uniquely to a 2-tuple $(m_1, m_2) \in M_1 \oplus M_2$ and it's easy to see this produces an R -isomorphism. So in this case we can still say that M is the direct sum of M_1 and M_2 but as these are submodules of M instead

of writing (m_1, m_2) we can write $m_1 + m_2$ and say we have an *inner direct product*.

Note that if $M = M_1 \oplus M_2$ then any element of M can be written uniquely as $(m_1, 0) + (0, m_2)$ for $m_1 \in M$ and $m_2 \in M_2$ so the ideas of inner and outer direct sum are equivalent, they are two faces of the same coin, yet it's important to be able to use these two ideas interchangeably.

Definition D.1.13. Let M be a left R -module and M_1 a submodule of M . If there exists M_2 submodule of M such that $M_1 \cap M_2 = (0)$ and $M_1 + M_2 = M$ then we say M_1 is a *direct summand* of M .

Next we'll see that R -homomorphisms display good behaviors when considering direct sums of modules.

Proposition D.1.14. Let M, M_1, M_2 be left R modules, then

- a) $\text{Hom}_R(M, M_1 \oplus M_2) \cong \text{Hom}_R(M, M_1) \oplus \text{Hom}_R(M, M_2)$
- b) $\text{Hom}_R(M_1 \oplus M_2, M) \cong \text{Hom}_R(M_1, M) \oplus \text{Hom}_R(M_2, M)$

The above result follows from D.1.12 and is easily extended to arbitrary direct sums of modules, by induction.

An element $e \in R$ is called *idempotent* if $e^2 = e$, moreover a set of idempotent elements of R is said to be *orthogonal* if $ef = 0$ for any two distinct idempotents e, f of said set. Note that if $e \in R$ is idempotent then the set $\{e, 1 - e\}$ is orthogonal.

Proposition D.1.15. Let $\{A_i\}_{i \in [n]}$ be left ideals of a ring R .

- a) The ring R can be written as a direct sum $R = A_1 \oplus A_2 \oplus \cdots \oplus A_n$ if and only if there exists a set $\{e_1, e_2, \dots, e_n\}$ of orthogonal idempotent elements of R such that $A_i = Re_i$ for $i \in [n]$ and $e_1 + e_2 + \cdots + e_n = 1$.
- b) The left ideals A_i in part a) are two sided ideals if and only if the corresponding idempotent elements e_i belong to the center of R .
- c) If condition b) holds then every R -module M can be written as a direct sum $M = M_1 \oplus M_2 \oplus \cdots \oplus M_n$ where $M_i = e_i M$ is a module over the ring A_i , for $i \in [n]$.

The above proposition is not hard to prove, yet we refer the reader to [Bea99, 2.2.7] if you desire so.

Definition D.1.16. An R -module M is said to be *completely reducible* if every submodule of M is a direct summand of M .

Later when we talk about *semisimple* modules we'll see what the above definition really means.

D.1.3 Tensor products

Here we'll briefly discuss tensor products, refer to Subsection 4.1.6 for more. We'll essentially give the general definition of tensor product of modules and a useful result about them.

Definition D.1.17. Given modules $M_R, {}_R N$ and an abelian group G , a function $\theta : M \times N \rightarrow G$ is said to be *R -bilinear* if

- a) $\theta(m_1 + m_2, n) = \theta(m_1, n) + \theta(m_2, n)$
- b) $\theta(m, n_1 + n_2) = \theta(m, n_1) + \theta(m, n_2)$
- c) $\theta(mr, n) = \theta(m, rn) = r\theta(m, n)$

for all $m, m_1, m_2 \in M, n, n_1, n_2 \in N$ and $r \in R$.

The following is a general definition of tensor product via a universal property.

Definition D.1.18. A tensor product of the modules $M_R, {}_R N$ is an abelian group denoted $M \otimes_R N$, together with an R -bilinear map $\tau : M \times N \rightarrow M \otimes_R N$ such that for any abelian group G and any R -bilinear map $\theta : M \times N \rightarrow G$ there exists a unique group homomorphism $f : M \otimes_R N \rightarrow G$ such that $f\tau = \theta$.

For $m \in M$ and $n \in N$ we denote $\tau(m, n)$ by $m \otimes n$.

Although seemingly quite abstract this definition essentially states that elements of $M \otimes_R N$ satisfy a bilinear property and that arguments about R -bilinear maps can be worked out by considering just group homomorphism from tensor products.

Proposition D.1.19. The tensor product $M \otimes_R N$ exists for any modules M_R and ${}_R N$.

Proposition D.1.20. For modules M_R and ${}_R N$ the tensor product $M \otimes_R N$ is unique up to isomorphism.

Proposition D.1.21. *Let M be a right R -module and $\{N_\alpha\}_{\alpha \in I}$ be a collection of left R -modules, then*

$$M \otimes_R \left(\bigoplus_{\alpha \in I} N_\alpha \right) \cong \bigoplus_{\alpha \in I} M \otimes_R N_\alpha$$

(as groups)

Refer to [Bea99, sec 2.6] for the proofs of the above propositions. A result similar to proposition D.1.21 also holds when the direct sum is on the right side, that is for N a left R -module and $\{M_\alpha\}_{\alpha \in I}$ a collection of right R -modules

$$\left(\bigoplus_{\alpha \in I} M_\alpha \right) \otimes_R N \cong \bigoplus_{\alpha \in I} M_\alpha \otimes_R N$$

If M_α , $\alpha \in I$ and N are vector spaces over a field $\mathbb{F}$ then because $\mathbb{F}$ is commutative M_α , $\alpha \in I$ and N are all $\mathbb{F}$ -modules on both sides, that is they are both left and right $\mathbb{F}$ modules (refer to our discussion succeeding definition 4.1.42), thus for $m \in M_\alpha$, $n \in N$ and $r \in \mathbb{F}$

$$rm \otimes_{\mathbb{F}} n = mr \otimes_{\mathbb{F}} n = m \otimes_{\mathbb{F}} rn = m \otimes_{\mathbb{F}} nr$$

Hence we can define

$$r(m \otimes_{\mathbb{F}} n) = (rm \otimes_{\mathbb{F}} n)$$

thus $M_\alpha \otimes_R N$ is vector space over $\mathbb{F}$. With this the preceding proposition states that $M \otimes_R \left(\bigoplus_{\alpha \in I} N_\alpha \right)$ and $\bigoplus_{\alpha \in I} M \otimes_R N_\alpha$ are isomorphic as vector spaces.

D.1.4 Semisimple modules

We've seen before in D.1.16 the definition of a completely reducible module, here we'll see that completely reducible modules have a very interesting structure.

Definition D.1.22. *An R -module is said to be semisimple if it's isomorphic to a direct sum of simple modules.*

Theorem D.1.23. *Any submodule of a semisimple module is semisimple and has a complement that is a direct sum of simple submodules, thus also semisimple.*

As a consequence of this theorem we have that a semisimple module is completely reducible, as it turns out the converse is also true.

Corollary D.1.24. *An R -module M is semisimple if and only if it is completely reducible.*

Again we refer the reader to [Bea99, sec 2.3] for the proofs. As the homomorphic image of a simple module must be simple one may easily verify the following result.

Proposition D.1.25. *The homomorphic image of a semisimple module is semisimple.*

D.2. Algebras

Algebras play an important role in representation theory, as many result about the structure of representations can be ascertained from the structure of related algebras. Here we present some basic definitions.

Definition D.2.1. *Let A be a vector space over a field $\mathbb{F}$. If additionally A is endowed with a multiplication operations such that A is a ring under (vector space) addition and multiplication, and for $\alpha \in \mathbb{F}$ and $a, b \in A$*

$$\alpha(ab) = (\alpha a)b = a(\alpha b)$$

then A is said to be an algebra over $\mathbb{F}$.

An algebra A is said to be *commutative* if the underlying ring in A is commutative. If A is finite dimensional as a vector space over $\mathbb{F}$ then we say A is a *finite dimensional algebra*.

Definition D.2.2. *Let A be an algebra, if B is a subspace of A that is closed for the multiplication operation in A then B is a subalgebra of A .*

Considering the above definition it's easy to verify that a subalgebra is in particular an algebra by itself, as expected.

Definition D.2.3. *Let A and B be algebras. A function $f : A \rightarrow B$ is said to be an algebra homomorphism if it's both a vector space homomorphism and a ring homomorphism.*

The kernel of an algebra homomorphism coincides with the kernel of the underlying group homomorphism, as was the case for modules, and the isomorphism theorem for algebras is analogous to the one for rings (which is analogous to the one for groups), it's a simple exercise to demonstrate it for algebras.

Definition D.2.4. *Let $\{A_\alpha\}_{\alpha \in I}$ be a collection of algebras, their direct sum/direct product is defined analogously to the definition for modules in D.1.11, with the additional operation of component wise multiplication, that is for $a, b \in \prod_{\alpha \in I} A_\alpha$, $(ab)_\alpha = a_\alpha b_\alpha$.*

It's straightforward to prove the direct sum/direct product of algebras is an algebra, as expected.

Definition D.2.5. An algebra is said to be simple if the underlying ring is simple, that is it has no non trivial ideals.

Definition D.2.6. Given an algebra A , it's center, denoted $Z(A)$, is defined as the subalgebra

$$Z(A) = \{a \in A \mid \forall_{b \in A} ab = ba\}$$

From the above definition it's clear that $Z(A)$ is a commutative algebra.

Definition D.2.7. Let A be an algebra and B a subset of A , we define the centralizer of B in A as the subalgebra

$$C_A(B) = \{a \in A \mid \forall_{b \in B} ab = ba\}$$

D.2.1 Involutione algebras

Let $\mathbb{F} = \mathbb{C}$ the field of complex numbers and recall the complex conjugate map.

Definition D.2.8. Let A be an algebra over $\mathbb{C}$, an involution on A is a bijective map $a \rightarrow a^*$ for $a \in A$ such that, for $\alpha, \beta \in \mathbb{C}$ and $a, b \in A$

- $(\alpha a + \beta b)^* = \overline{\alpha} a^* + \overline{\beta} b^*$
- $(ab)^* = b^* a^*$
- $(a^*)^* = a$

The element a^* is called the adjoint of a , and a is said to be normal if it commutes with its adjoint, that is $aa^* = a^*a$, moreover if $a^* = a$ then a is said to be self-adjoint or hermitian.

Definition D.2.9. An algebra is said to be involutive if it's equipped with an involution, it's then usually called a $*$ -algebra.

Example D.2.10. If we consider an inner product space V over $\mathbb{C}$ then $\text{Hom}(V, V)$ is an algebra under addition, composition of maps and scalar product by $\mathbb{C}$. The adjoint map that send $T \in \text{Hom}(V, V)$ to its adjoint T^* as in definition C.0.6 is an involution as can be easily verified using proposition C.0.8.

D.2.2 Complexification

The following construction will be useful to prove when certain algebras are commutative.

Definition D.2.11. Let A be an involutive algebra over $\mathbb{R}$ the field of real numbers, then the $*$ -complexification of A is the algebra $A \times A$ where elements $(a, b) \in A \times A$ are usually written as $a + ib$, with the operations of addition, product and scalar multiplication by $\mathbb{R}$ as in the field $\mathbb{C}$ of complex numbers, and component wise involution, that is

$$(a + ib)^* = a^* + ib^*$$

A real element of the $*$ -complexification of A is an element of the form $a + i0$, for $a \in A$.

Theorem D.2.12. Let A be an involutive algebra over $\mathbb{C}$.

- a) An element $a \in A$ is normal if and only if $a = x + iy$ for some self-adjoint $x, y \in A$ that commute with each other.
- b) The algebra A is commutative if and only if every element of A is normal.
- c) If A is a $*$ -complexification of a real involutive algebra then A is commutative if every real element of A is self-adjoint.

The proof of the above theorem is not hard and can be seen in [Meh16, 2.21].

