August 2022

The Earth Mover's Distance Through the Lens of Algebraic Combinatorics

William Quentin Erickson
University of Wisconsin-Milwaukee

THE EARTH MOVER'S DISTANCE THROUGH THE LENS OF ALGEBRAIC COMBINATORICS

by

William Q. Erickson

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Mathematics

at

The University of Wisconsin–Milwaukee

August 2022

ABSTRACT

THE EARTH MOVER'S DISTANCE THROUGH THE LENS OF ALGEBRAIC COMBINATORICS

by

William Q. Erickson

The University of Wisconsin–Milwaukee, 2022
Under the Supervision of Professor Jeb F. Willenbring

The earth mover's distance (EMD) is a metric for comparing two histograms, with burgeoning applications in image retrieval, computer vision, optimal transport, physics, cosmology, political science, epidemiology, and many other fields. In this thesis, however, we approach the EMD from three distinct viewpoints in algebraic combinatorics. First, by regarding the EMD as the symmetric difference of two Young diagrams, we use combinatorial arguments to answer statistical questions about histogram pairs. Second, we adopt as a natural model for the EMD a certain infinite-dimensional module, known as the first Wallach representation of the Lie algebra $\mathfrak{su}(p,q)$, which arises in the Howe duality setting in Type A; in this setting, we show how the second fundamental theorem of invariant theory generalizes the "northwest corner rule" from optimal transport theory, yielding a simple interpretation of the "partial matching" case of the EMD via separation into invariants and harmonics. Third, we reapproach partial matching in the context of crystal bases of Types A, B, and C, which leads us to introduce a variation of the EMD in terms of distance on a crystal graph. Having exploited these three approaches, we generalize all of our EMD results to an arbitrary number of histograms rather than only two at a time. In the final chapter, we observe a combinatorial connection between generalized BGG resolutions arising in Type-A Howe duality and certain non-holomorphic discrete series representations of the group $\mathrm{SU}(p,q)$.

To my parents,

for suggesting that perhaps I might like to add a major in mathematics.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

$\mathbb{N}$	$\{0, 1, 2, 3, \ldots\}$
$[n]$	$\{1, \ldots, n\}$
λ	a histogram $(\lambda_1, \ldots, \lambda_n) \in \mathbb{N}^n$ with bins $1, \ldots, n$
λ_i	number of data points in the ith bin of λ
$\mathcal{H}(m, n)$	the set of histograms λ with mass $m = \sum_i \lambda_i$ and with n bins
C_{ij}	the "cost" $\lvert i - j \rvert$
$\mathrm{w}(\lambda)$	the word $\underbrace{\boxed{1}\cdots\boxed{1}}_{\lambda_1 \text{ times}}\underbrace{\boxed{2}\cdots\boxed{2}}_{\lambda_2 \text{ times}}\cdots\underbrace{\boxed{n}\cdots\boxed{n}}_{\lambda_n \text{ times}}$
$W_{\lambda,\mu}$	the $2 \times m$ matrix $\begin{bmatrix} \mathrm{w}(\lambda) \\ \mathrm{w}(\mu) \end{bmatrix}$
i'	$n - i$, for $i \in [n]$
$t(\lambda)$	the weighted total $\sum_i i' \cdot \lambda_i$
$\mathbf{D}(\lambda, \mu)$	the weighted difference $t(\lambda) - t(\mu)$; generalized in Chapter 6
$\mathbf{EMD}(\lambda, \mu)$	the earth mover's distance between histograms λ and μ
$\mathrm{w}'(\lambda)$	the complementary word $(\underbrace{1', \ldots, 1'}_{\lambda_1}, \ldots, \underbrace{n', \ldots, n'}_{\lambda_n})$; a partition of $t(\lambda)$
$\mathrm{Y}(\lambda)$	the Young diagram associated with $\mathrm{w}'(\lambda)$
$\mathcal{Y}(m, n-1)$	the set (or lattice) of Young diagrams fitting inside an $m \times (n-1)$ box
ε_i	$(0, \ldots, 0, \underbrace{1,}_{i} 0, \ldots, 0)$
F_k^ξ	irreducible rational representation of GL_k with highest weight ξ
$C(\mathbf{x})$	generalized cost, i.e., ℓ_1-distance between $\mathbf{x} \in \mathbb{N}^d$ and the line $\mathbb{N}(1, \ldots, 1)$
$\boldsymbol{\lambda}$	$(\lambda^1, \ldots, \lambda^d)$ with each $\lambda^i \in \mathcal{H}(m, n)$
$\mathbf{EMC}(\boldsymbol{\lambda})$	the earth mover's coefficient of the d-tuple $\boldsymbol{\lambda}$
$\mathbf{D}(\boldsymbol{\lambda})$	the generalized weighted difference $[t(\lambda^1), \ldots, t(\lambda^d)] \in \mathbb{Z}^d/\mathbb{Z}(1, \ldots, 1)$

ACKNOWLEDGMENTS

My deepest thanks go to my extraordinary advisor, Jeb Willenbring. Jeb seems always to have the right advice, both mathematical and pragmatic, and is deeply generous in making time for his graduate students. He has constantly shared problems for me to think about, while guiding me through the background literature and explaining examples at his office chalkboard. Just as importantly, he is also kind and thoughtful and funny and encouraging, and as a result, each of my years in graduate school became successively more fun and productive. I cannot imagine a better advisor, and I will always be supremely grateful.

I would like to thank all of my professors and course coordinators at UW–Milwaukee as well. In particular, I am indebted to Eric Key, whose night class in linear algebra sparked my path toward graduate school, and who spent many Saturday mornings in the City Market Café chatting with me about the harder problems in the Hoffman and Kunze textbook. I should also thank Mark Colarusso for the way he taught Modern Algebra in my first year, which cemented my interest in algebra research. Likewise, Allen Bell's second-year Modern Algebra course was immensely valuable to me, particularly his knack for assigning just the right exercises for us to struggle with and ultimately learn from.

Of course, I would like to thank the members of my doctoral defense committee — Allen Bell, Chris Hruska, Kevin McLeod, and Yi Ming Zou — for their careful reading and feedback on this thesis.

I have been fortunate to collaborate on research projects not only with Jeb, but also with Mark Colarusso, and separately with Rebecca Bourn. Both Mark and Rebecca have taught me much through our conversations and our crafting papers together.

I would like to acknowledge Paul Terwilliger, who (in conversation with Jeb Willenbring) realized that the earth mover's distance is equivalent to the graph distance on the

Type-A crystal $\mathcal{B}_{(m)}$; see Chapter 5.

I thank my parents for their constant interest and support; my brothers Steve and Christian for always encouraging me to keep trying to discover that four-sided triangle; Jeff, for being Jeff; and Grace, for making everything more fun, and who is far better at math than she wants to be. Also a certain domestic shorthair named Penguin who made several contributions to this thesis, which I hope I have deleted in the proofing process; and finally, my favorite uncle Mike, for being such an avid supporter and an all-around good buddy.

1 INTRODUCTION

The earth mover's distance (EMD) is a measure of the closeness of two histograms, a measure with a long history and with ever-widening applications in the present. Its story begins nearly 250 years ago, with French geometer GASPARD MONGE. Monge wrote a paper in 1781 bearing the French title *Mémoire sur la théorie des déblais et des remblais*: in approximate English, "Treatise on the theory of excavating and refilling." (Although the "EMD" moniker was still two centuries away, this title vindicates Monge as the original earth mover.) In the paper, Monge posed the following problem: given any two regions of land, how can the "earth" be dug out of the first region and transported to cover the second region, while expending the minimum amount of work? (A unit of work corresponds to moving, say, one acre's worth of earth by one mile.) This became known as the **Monge problem**, and the minimum amount of required work would eventually become known as the earth mover's distance between the two regions. Throughout the following centuries, the Monge problem proved fertile ground for many generalizations and became the launch pad for optimal transport theory; consult, for instance, the opening of Villani's monumental reference [Vil09].

Figure 1: Gaspard Monge (1746–1818).

1.1 Development of the EMD

In the late 1930's, the Monge problem was unearthed by Leonid Kantorovich [Kan42], who recast the solution as the distance between two probability distributions. Now widely called the Monge–Kantorovich problem, this adaptation has been well-studied in the decades since; in the context of probability distributions, the minimum work required is also called the **1-Wasserstein distance**, or the **Mallows distance**.

The Monge problem gained prominence in the 1990s, thanks to the paper [RTG98] advocating for its application to image retrieval. This paper was also the first to introduce the newly-coined nickname, the "earth mover's distance." (Evidently the name was suggested in a 1994 letter written by Jorge Stolfi, who used the phrase in reference, not to Monge, but to a CAD program for road design that involved transporting dirt.) The authors of [RTG98] generalized the "match distance" proposed in [WPR85], which had required that the two histograms have the same mass; on the contrary, the EMD definition in [RTG98] allowed for *partial matching* when the masses are different. Moreover, [RTG98] emphasized a generalization of a histogram called a *signature* (see Section 3 therein). In recent years, the EMD has been adopted as a tool not only in computer vision, but also in physics [KMT19], cosmology [FMM02], political science [LSW17], epidemiology [MKV$^+$20], and many other fields.

The solution to the Monge problem — called a "flow" or a "transport plan" — depends not only on the values in the histograms, but also on the "cost" (or "ground distance") between each of the histogram bins. In this thesis, we study the special case in which a histogram is regarded as n piles of earth lying one unit apart on the number line. We share a maximally easy example to provide some early intuition.

Example 1.1. Consider the following grade distributions:

Letter grade	A	B	C	D	F
Class X	3	1	5	0	1
Class Y	2	3	4	0	1

Suppose that moving one student up or down by one letter grade equals one unit of work. Then the EMD between the distributions is just the minimum amount of work required to transform one into the other. In this case, we can quickly compute this distance by inspection: in Class X, for example, move one student from A to B, and move another student from C to B, requiring two units of work in total. None of the other ways of transforming X into Y (or vice versa) is possible with fewer than two units of work, and so the EMD between X and Y is 2.

1.2 Preview of results

We now present an overview of the thesis, in particular highlighting the new results and contributions we have included.

Chapter 2: The EMD as a statistic

This chapter is purely preliminary in nature, defining the EMD from the original perspective of transport theory. We do, however, introduce another statistic on histogram pairs, which we call the **weighted difference** and denote by **D**; the interplay between the EMD and the weighted difference will be a recurring theme of the thesis. The main takeaway from this chapter is the existence of a greedy algorithm known as the **northwest corner rule**, for which (assuming a sufficiently nice cost function between the histogram bins) the input is a pair of histograms, and the output is the optimal flow matrix solving the Monge problem, and hence determining the EMD.

Chapter 3: A combinatorial approach to the EMD

Our primary innovation in this chapter is our viewing histograms as **Young diagrams**, in such a way that the weighted difference **D** is simply the difference between the sizes of the corresponding diagrams. We thereby use the language of Young diagrams to take

a purely combinatorial approach to the EMD. The main observation is **Proposition 3.4**, which identifies the EMD as the symmetric difference of the corresponding Young diagrams. (The underlying content of this result is well known, but our expression of it in terms of Young diagrams is a new perspective that allows for combinatorial arguments throughout the chapter.) It is somewhat surprising that the symmetric difference of Young diagrams has not been studied in the existing combinatorics literature; as a first step toward filling the gap, we use our Proposition 3.4 to reinterpret a known EMD generating function in terms of Young diagrams. In the opposite direction, we also answer two statistical questions using combinatorial arguments: in **Theorem 3.10** we find the probability that $\mathbf{EMD} = |\mathbf{D}|$, along with a nice asymptotic result in **Corollary 3.11**, while in **Theorem 3.12** we compute the expected value of $|\mathbf{D}|$. Finally, in **Proposition 3.13** we express the EMD directly in terms of $\mathbf{D}$, using the structure of Young's lattice.

Chapter 4: The EMD and the first Wallach representation

Now we introduce representation theory into our treatment of the EMD. Front and center is the Lie group $\mathrm{SU}(p,q)$ and a certain infinite-dimensional module for its Lie algebra, known as the **first Wallach representation**. The Wallach representations are significant because they were constructed in [Wal79] as analytic continuations of discrete series representations. (More on the discrete series at the end of this introduction.) The front end of this chapter is fairly technical by nature, but it all culminates neatly in Section 4.6. Specifically, we exhibit two models of the first Wallach representation of $\mathfrak{su}(p,q)$: (1) a certain algebra of invariant polynomials, and (2) the coordinate ring of a certain determinantal variety. The second fundamental theorem of classical invariant theory furnishes an algebra isomorphism φ between these two models: in **Theorem 4.15**, the highlight of the chapter, we show that φ is actually the representation-theoretic analogue of the northwest corner rule from Chapter 2. This new perspective then allows us to compute the EMD completely within the framework of representation theory (**Corollary 4.16**). Even

more, we show (**Theorem 4.19**) how the case of *partial matching* (i.e., unequal histogram masses) can be elegantly understood through representation theory — whereas until now, we had avoided this case because it was not amenable to our combinatorial approach.

Chapter 5: The EMD as distance in crystal graphs

Crystal bases and their corresponding graphs arise from the study of quantum groups. As deep as the representation theory of quantum groups is, the combinatorics of their crystal bases is remarkably simple. In this chapter, we show how certain crystal graphs, in each of the root system types A, B, and C, are perhaps the most natural possible setting for the EMD, and in fact serve to clarify some results in previous chapters. In particular, the EMD is nothing other than the graph distance in the Type-A crystal $\mathcal{B}_{(m)}$ (**Proposition 5.1**). Even more strikingly, by essentially following our noses in the analogous crystals of Types B and C (**Definitions 5.2** and **5.6**), we propose a variation on the EMD in order to better address the case of partial matching, which remedies some of the statistical drawbacks (described in [BL01]) of the original EMD definition.

Chapter 6: Generalization to more than two histograms

The project in this chapter is to generalize all of our previous combinatorial tools and results so that we can compare arbitrarily many histograms at a time, rather than only two. Most of the results here have appeared in the author's paper [Eri21] or preprint [Eria]. This generalization of the Monge problem has been studied in the literature, but our main result is a recursive formula (**Theorem 6.12**) for the expected value of this generalized EMD (which we now call an earth mover's *coefficient*, since "distance" is no longer accurate). We also propose a generalization of **D** as an element of the root lattice of the Lie algebra $\mathfrak{sl}_d$ (**Definition 6.13**), and we write down the distribution of **D**-values as a virtual $\mathfrak{sl}_d$-character (**Theorem 6.14**).

Chapter 7: Enright resolutions and Blattner's formula in Type A

In this final chapter, we present an unexpected connection that we observed between two settings: (1) the infinite-dimensional $\mathfrak{su}(p,q)$-modules in Chapter 4 (which included the first Wallach representation), and (2) certain non-holomorphic discrete series representations of $\mathfrak{su}(k,p+q)$. Specifically, in **Theorem 7.5**, we interpret the coefficients of a generating function (in the second setting) as the well-known generalized BGG resolutions of the $\mathfrak{su}(p,q)$-modules (in the first setting). This chapter is a light reworking of the author's preprint [Erib]; it is necessarily quite technical and assumes some familiarity with the representation theory of real reductive groups. Since this final chapter leaves the EMD behind, it can be regarded as an appendix to the thesis; but we believe that the result is noteworthy and relevant since it occurs in the precise setting where the important map φ arose in Chapter 4.

The discrete series in harmonic analysis

Since discrete series are the theme of Chapter 7, we provide a little background here. Despite our algebraic perspective in that chapter, the original motivation is overwhelmingly analytical. Although this angle will not appear in the main body of the thesis, we sketch the main idea below for those readers who are interested. We follow the exposition from the paper [Her91], which we recommend highly as a concise and excellent synopsis of Harish–Chandra's legacy in pioneering the harmonic analysis of semisimple Lie groups.

The **harmonic analysis** of a semisimple Lie group G refers to the decomposition of its regular representation $L^2(G)$ into irreducible unitary representations. In case G is compact, the **Plancherel formula** takes the following form [Her91, p. 6]:

$$f = \sum_{\pi \in \widehat{G}} d_\pi (\Theta_\pi * f)$$

for $f \in L^2(G)$, where $\widehat{G}$ is the **unitary dual**, i.e., the set of equivalence classes of irreducible unitary representations of G (all finite-dimensional in the compact case); the coefficient d_π is just the dimension of the representation space of π, and Θ_π is the **character** of π, defined as sending $g \longmapsto \operatorname{tr} \pi(g)$. In this case, the **Plancherel measure** is the measure on $\widehat{G}$ sending $\pi \longmapsto d_\pi$. Hence f decomposes into a sum of elementary functions parametrized by $\widehat{G}$, where the Plancherel measure of each $\pi \in \widehat{G}$ determines its contribution to the decomposition. (This generalizes the Fourier series, for which G is the compact abelian group $\mathbb{R}/\mathbb{Z} \cong S^1$, the Plancherel formula is $f(x) = \sum_{n \in \mathbb{Z}} \hat{f}(n) e^{2\pi i n x}$, and the Plancherel measure is just the counting measure.) Hence the two basic problems of harmonic analysis are (1) to determine $\widehat{G}$, and (2) to determine the Plancherel measure on $\widehat{G}$.

In general, for G noncompact, a direct-sum decomposition of $L^2(G)$ is not possible as it is above; instead, a direct integral is required, and the Plancherel formula becomes

$$f(1) = \int_{\widehat{G}} \Theta_\pi(f)\, d\mu(\pi),$$

with notation as follows:

- $f \in C_c^\infty(G)$.
- Θ_π is now the *distribution* on G defined by $\Theta_\pi(f) = \operatorname{tr} \int_G f(g)\pi(g)\, dg$.
- μ is the Plancherel measure on $\widehat{G}$.

A **discrete series** representation $\pi \in \widehat{G}$ has positive point mass with respect to the Plancherel measure μ. Therefore, upon restricting $\widehat{G}$ to the discrete series, the Plancherel formula truly is a discrete sum again. As Varadarajan writes, "the discrete series representations constitute a striking generalization of the finite-dimensional representations of compact groups ... it is hardly possible to have a more thoroughgoing generalization of the compact theory." [Var89, p. 188]

2 THE EMD AS A STATISTIC

This chapter is devoted to defining the EMD in terms of the solution to the Monge problem. For now we assume the original perspective of statistics and optimal transport theory, before taking up a combinatorial approach in the following chapter. The material here is all preliminary in nature.

2.1 Histograms and the weighted difference D

In this thesis, we restrict our attention to discrete histograms on a one-dimensional feature space, in which the bins are the elements of the set $[n] \coloneqq \{1, \ldots, n\}$. Therefore we will use the term **histogram** to refer to an element of $\mathbb{N}^n$ (with the convention $0 \in \mathbb{N}$). We will use a Greek letter to denote a histogram in the form

$$\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbb{N}^n,$$

where λ_i is the number of data points in bin i. By the **mass** of a histogram, we mean the total number of data points $\sum_i \lambda_i$, which we will typically denote by m. For fixed $m, n \in \mathbb{N}$, we define the set

$$\mathcal{H}(m, n) \coloneqq \Big\{\text{histograms with mass } m \text{ and with } n \text{ bins}\Big\}.$$

The grade distributions in Example 1.1 are both elements of $\mathcal{H}(10, 5)$, if we identify the grades A, B, C, D, F with bins $1, \ldots, 5$. We consider the cost (i.e., ground distance) between two bins to be their distance on the number line. To be precise, we define an $n \times n$ **cost matrix** $C = [C_{ij}]$, where

$$C_{ij} \coloneqq |i - j|. \tag{2.1}$$

For example, if $n = 5$, then

$$C = \begin{bmatrix} 0 & 1 & 2 & 3 & 4 \\ 1 & 0 & 1 & 2 & 3 \\ 2 & 1 & 0 & 1 & 2 \\ 3 & 2 & 1 & 0 & 1 \\ 4 & 3 & 2 & 1 & 0 \end{bmatrix}. \tag{2.2}$$

It will be most natural to regard bin 1 as the "best/highest" and bin n as the "worst/lowest." Therefore, for $i \in [n]$, we write $i' := n - i$. Then we define the **weighted total** of a histogram λ as follows:

$$t(\lambda) := \sum_i i' \cdot \lambda_i. \tag{2.3}$$

In words, each data point in bin i contributes i' to the weighted total. Now it is natural to define the **weighted difference**, denoted by $\mathbf{D}$, between two histograms $\lambda, \mu \in \mathcal{H}(m, n)$:

$$\mathbf{D}(\lambda, \mu) := t(\lambda) - t(\mu),$$

i.e., the (signed) difference between their weighted totals.

Example 2.1. If $n = 5$, and we regard the letter grades A, ..., F as the bins $1, \ldots, 5$, then $t(\lambda)$ equals the total grade points earned, under the familiar system whereby A = 4, B = 3, C = 2, D = 1, and F = 0. Dividing by m then obtains the class GPA. Given two grade distributions $\lambda, \mu \in \mathcal{H}(m, 5)$, we are often interested in $\frac{\mathbf{D}(\lambda,\mu)}{m}$, the difference between the two GPA's. We will compute the expectation of the absolute value of $\mathbf{D}$ combinatorially in Theorem 3.12.

Example 2.2. Consider m voters participating in a ranked-choice election between n candidates. Then for each candidate, the results are represented by a histogram $\lambda \in \mathcal{H}(m, n)$, where λ_i is the number of ith-place votes received. In this context, $t(\lambda)$ equals the *Borda count* score for that candidate. In the Borda system, a candidate λ wins the election if and only if $\mathbf{D}(\lambda, \mu) > 0$ for all other candidates μ.

Remark. The weighted difference $\mathbf{D}$ may at first seem unconnected to the EMD. After all, consider the histograms $\lambda = (2, 2, 2)$, $\mu = (3, 0, 3)$, and $\nu = (0, 6, 0)$ in $\mathcal{H}(6, 3)$. On one

hand, we can see immediately that the EMD between each pair of histograms is nonzero, since *some* work must be done in order to equalize them. On the other hand, we find that $t(\lambda) = t(\mu) = t(\nu) = 6$, and therefore $\mathbf{D}(\lambda, \mu) = \mathbf{D}(\lambda, \nu) = \mathbf{D}(\mu, \nu) = 0$. We will study the interesting interplay between the EMD and $\mathbf{D}$, and in particular we will demonstrate a way in which the former can be reduced to the latter. Moreover, when we compare an arbitrary number d of histograms in Chapter 6, we will redefine $\mathbf{D}$ as a weight of the Lie algebra $\mathfrak{sl}_d$.

2.2 Definition of EMD

Consider two histograms with n bins, and with possibly unequal masses: $\lambda \in \mathcal{H}(\ell, n)$ and $\mu \in \mathcal{H}(m, n)$. Their EMD is defined in [RTG98] by the solution to a certain linear programming problem (traditionally named after Gaspard Monge, or various combinations of Monge, Kantorovich, Hitchcock, and Koopman). Recall that in our case, the $n \times n$ cost matrix $C = [C_{ij}]$ is given by $C_{ij} = |i - j|$. We want to find a real $n \times n$ **flow matrix** $F = [F_{ij}]$ in order to solve the following **Monge problem**:

$$\text{Minimize} \quad \sum_{i,j=1}^{n} C_{ij} F_{ij}, \tag{2.4}$$

$$\text{subject to} \quad F_{ij} \geqslant 0 \qquad \text{for all } 1 \leqslant i, j \leqslant n, \tag{2.5}$$

$$\text{and} \quad \sum_{j=1}^{n} F_{ij} \leqslant \lambda_i \qquad \text{for each } 1 \leqslant i \leqslant n, \tag{2.6}$$

$$\text{and} \quad \sum_{i=1}^{n} F_{ij} \leqslant \mu_j \qquad \text{for each } 1 \leqslant j \leqslant n, \tag{2.7}$$

$$\text{and} \quad \sum_{i,j=1}^{n} F_{ij} = \min\{\ell, m\}. \tag{2.8}$$

The matrix F is also sometimes called a "transport plan" in the literature. The idea is that any such F encodes a way to transform λ into μ (or as closely as possible, when $\ell \neq m$): namely, F_{ij} equals the amount of mass inside λ that must "flow" from bin i to bin j. Hence any candidate F must have nonnegative entries (2.5); the ith row sum (i.e., amount of

mass flowing out of bin i) of course cannot exceed the mass in bin i of λ (2.6); likewise, the jth column sum (the amount of mass flowing into bin j) cannot exceed the mass in bin j of μ (2.7); and as a final result of the flow, the smaller histogram must be completely contained in the larger (2.8). We define $\mathcal{F}_{\lambda,\mu}$ to be the set of all flow matrices F satisfying constraints (2.5) – (2.8).

Once we have found an optimal solution F to the Monge problem, the **earth mover's distance (EMD)** is defined to be the objective quantity in (2.4); in other words,

$$\mathbf{EMD}(\lambda, \mu) \coloneqq \inf_{F \in \mathcal{F}_{\lambda,\mu}} \left\{ \sum_{i,j} C_{ij} F_{ij} \right\}. \tag{2.9}$$

As a first example, consider the trivial pair (λ, λ) of two equal distributions. Then the diagonal matrix $F = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ lies in $\mathcal{F}_{\lambda,\lambda}$, and we have $F_{ij} = 0$ whenever $i \neq j$. Since $C_{ii} = 0$, the total work in (2.4) equals $\sum_i C_{ii}\lambda_i = 0$, and so $\mathbf{EMD}(\lambda, \lambda) = 0$, just as we would expect.

Remark (Normalization)**.** In [RTG98] and in subsequent sources, the EMD is actually defined to be the normalization of our EMD, via division by the mass $\min\{\ell, m\}$ of the smaller histogram. In this thesis we retain the definition (2.9) in order to keep the combinatorics front and center, although we will occasionally proceed to normalize when computing expected values in Section 6.4.

Remark (Complete vs. partial matching)**.** From now on, unless otherwise stated, we will consider only histogram pairs of equal mass m. In this case the flow determined by F is called a *complete matching* in the literature; this is the assumption, for example, in the application to texture matching in [PBRT99]. Under this assumption, $\min\{\ell, m\}$ can be replaced by m in (2.8), which in turn forces equality in both (2.6) and (2.7). Nevertheless, we will describe an elegant way to understand the case of unequal masses (*partial matching*) in Section 4.7, in the framework of representation theory. (See also Sections 5.3–5.4, where we introduce an alternative approach derived from the combinatorics of

crystal bases.) As explained in [BL01], the allowance for unequal masses is precisely what distinguishes the EMD from the 1-Wasserstein (or Mallows) distance.

Remark (Discrete vs. continuous)**.** The EMD definition (2.9) certainly remains valid when the histogram components λ_i are allowed to be nonnegative *real* numbers, as they are in certain applications. In this thesis we study discrete histograms (which are also prevalent in applications) in order to leverage tools from algebraic combinatorics, but we often can extend results to the continuous setting via normalizing by $1/m$ and then taking limits as $m \to \infty$.

Remark (Taxicab distance)**.** A recurring motif throughout our study of the EMD, in several different contexts, is the ℓ_1-norm (commonly called the "taxicab," "Manhattan," or "rectilinear" distance). We recall that given a vector $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$, its ℓ_1-norm is defined to be

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{d} |x_i|.$$

We then have the taxicab distance defined by

$$d_1(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_1 = \sum_{i=1}^{d} |x_i - y_i|.$$

Note that the cost matrix entry C_{ij} could alternatively be defined as the taxicab distance from (i, j) to the line $\mathbb{N}(1, 1)$, i.e., the main diagonal of the matrix. This fact will guide us in generalizing the EMD for more than two histograms in Chapter 6.

2.3 The Monge property and the northwest corner rule

By definition, computing the EMD depends upon first solving the Monge problem (2.4) – (2.8). The difficulty of this problem varies with the choice of cost matrix C. In our case, where we have defined $C_{ij} = |i - j|$, our cost matrix C has a certain property that allows the Monge problem to be solved by a greedy algorithm known as the northwest corner rule. In this section, rather than dwell upon the algorithm itself, we simply aim

to establish that there is an optimal $F \in \mathcal{F}_{\lambda,\mu}$ whose support is a chain under the product order.

Definition 2.3. A matrix $A = [A_{ij}]$ has the **Monge property** if, for all $i < I$ and $j < J$,

$$A_{ij} + A_{IJ} \leqslant A_{Ij} + A_{iJ}.$$

That is, choose any two distinct row indices i, I and two distinct column indices j, J; these two rows and two columns intersect at four entries arranged in a rectangle. The Monge property guarantees that the sum of the northwest and southeast entries is at most the sum of the northeast and southwest entries. From a moment's glance at the example (2.2), it is not hard to see that our cost matrix C has this property.

Lemma 2.4. *The cost matrix C given by $C_{ij} = |i - j|$ has the Monge property.*

Proof. Suppose the Monge property holds in the special case that $I = i + 1$ and $J = j + 1$. Then we have

$$\begin{aligned} C_{ij} + C_{i+1,j+1} &\leqslant C_{i+1,j} + C_{i,j+1}, \\ C_{i,j+1} + C_{i+1,j+2} &\leqslant C_{i+1,j+1} + C_{i,j+2}, \end{aligned} \tag{2.10}$$

which after adding and simplifying gives us

$$C_{ij} + C_{i+1,j+2} \leqslant C_{i+1,j} + C_{i,j+2},$$

and continuing in this fashion we arrive at all possible $J > j$; symmetrically we can do the same to arrive at all $I > i$. Hence it suffices to prove (2.10).

The left-hand side of (2.10) equals $2|i - j|$ for any i, j. In the case that $i = j$, the left-hand side equals 0 and the right-hand side equals 2. In the case that $i > j$, the left-hand side equals $2(i - j)$ and the right-hand side also equals $(i - j) + 1 + (i - j) - 1 = 2(i - j)$. Finally, in the case that $i < j$, the left-hand side is $2(j - i)$ and the right-hand side is also $(j - i) - 1 + (j - i) + 1 = 2(j - i)$. Hence in the first case we obtain a strict inequality in (2.10), and in the other two cases we obtain equality. □

Suppose $\lambda, \mu \in \mathcal{H}(m,n)$. The significance of the Monge property lies in a result of Hoffman in [Hof63]: namely, the Monge problem (2.4) – (2.8) can be solved via an $O(2n)$-time greedy algorithm (the **"northwest corner rule"**), if and only if the associated cost matrix C has the Monge property. The northwest corner rule solves the Monge problem by constructing an optimal flow matrix $F \in \mathcal{F}_{\lambda,\mu}$: beginning in the upper-left, the algorithm sets $F_{11} = \min\{\lambda_1, \mu_1\}$, and then modifies both λ_1 and μ_1 by subtracting this same value (so at least one of them becomes 0). This allows us to fill the remainder of either the first row or the first column with 0's; then we proceed either south or east, and repeat the process until we have filled the entire matrix F, which is the solution to the Monge problem. It is immediate from this construction that F lies in $\mathcal{F}_{\lambda,\mu}$. Furthermore, since $\lambda, \mu \in \mathcal{H}(m,n)$ have bin masses in $\mathbb{N}$, it is clear that the entries of F are natural numbers as well. Finally, and most importantly, we observe that the nonzero entries of this F lie in a single path traveling either south or east at each step; in other words, for any $i < I$ and $j < J$, we never have both F_{iJ} and F_{Ij} nonzero. This makes sense intuitively: if we want to minimize work, we would not move mass both from bin i to bin J and also from bin I to bin j. Instead, we move from i to j, then either from i to J or from I to j, and finally from I to J. (This scheme of minimizing work will reappear in the context of standard monomial theory in Section 4.3.)

To make the matter more precise, we recall the **product order** $\preceq$ on the poset $[n] \times [n]$, whereby $(i,j) \preceq (k,\ell)$ if and only if $i \leqslant k$ and $j \leqslant \ell$. Recall also that a **chain** in a poset is a totally-ordered subset, i.e., a subset in which every pair of elements is comparable. We now record the following crucial consequence of Lemma 2.4 and Hoffman's result.

Proposition 2.5. *There exists an optimal $F \in \mathcal{F}_{\lambda,\mu}$ with entries in $\mathbb{N}$, whose support is a chain in $[n] \times [n]$.*

(Incidentally, the existence of this algorithm shows that the infimum in the definition (2.9) is actually a true minimum.) We rather emphatically do *not* end this section with an example of the northwest corner rule, because in the next chapter we develop a purely

combinatorial alternative which we will use in its place.

3 A COMBINATORIAL APPROACH TO THE EMD

In [BW20], Bourn and Willenbring introduced an alternative to the northwest corner rule for solving the Monge problem: in particular, they observed that the nonzero entries of the optimal flow matrix F fall out of the well-known Robinson–Schensted–Knuth (RSK) correspondence, a ubiquitous combinatorial phenomenon with applications in many fields of mathematics. In this chapter, we build upon these insights by associating each histogram to a combinatorial object called a *Young diagram*, which provides an elegant framework to answer statistical questions about the EMD and weighted difference. Much of the work in this chapter has appeared in the author's preprint [Eria].

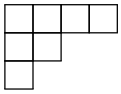
3.1 Realizing histograms as Young diagrams

The set $\mathcal{H}(m,n)$ is equivalent (in combinatorial language) to the set of **weak integer compositions** of m into n parts, i.e., the subset of $\mathbb{N}^n$ consisting of n-tuples whose components sum to m. We will continue to use the term "histogram" and the symbol $\mathcal{H}$ as a nod to the statistical motivation for the EMD, but the reader is welcome to substitute the term "composition" whenever it seems more natural. It is well known that $\#\mathcal{H}(m,n) = \binom{m+n-1}{m} = \binom{m+n-1}{n-1}$, which we can see at once by imagining a histogram as m indistinguishable data points along with $n-1$ dividers between the n bins: choosing m locations for the data points (or equivalently $n-1$ locations for the dividers) uniquely determines a histogram.

A **partition** ξ of $t \in \mathbb{N}$ is a weakly decreasing tuple $(\xi_1, \ldots, \xi_k)$ of positive integers (**parts**) that sum to t. Hence the notion of *partition* is stricter than that of *histogram*, due to the weakly-decreasing condition. We call t the **size** of ξ, which we write either as $|\xi| = t$ or $\xi \vdash t$. The number of parts is called the **length**, written $\ell(\xi)$. When convenient, we pad with 0's on the right, so that, for example, the partition $(5,3)$ is considered identical

to $(5, 3, 0, 0, 0)$, although in both cases the length is 2. Partitions will play their largest role in the representation theory of Chapter 4, but they are also part of this early picture, via their visual avatars known as Young diagrams.

The **Young diagram** of a partition ξ is an arrangement of **cells** (boxes), justified along the left and top edges, whose ith row (from the top) contains ξ_i cells. For example, if $\xi = (4, 2, 1)$, then its associated Young diagram is shown below:



Terminology carries over from partitions: the **size** $|Y|$ is the number of cells in a Young diagram Y, and its **length** is the number of rows. We also say that the **width** is the number of columns, which equals the first part ξ_1 of the associated partition.

In algebraic combinatorics, we often consider Young diagrams in which each cell is filled with a value (usually a positive integer) taken from a certain **alphabet**. We call these **Young tableaux**; moreover, if the values of a Young tableau weakly increase along rows and strictly increase down columns, then it is said to be a **semistandard Young tableau (SSYT)**. The partition defining the underlying Young diagram is called the **shape** of the tableau. For example, below are all SSYT's with the shape $(2, 1)$ and the alphabet $[3] = \{1, 2, 3\}$:

$$\begin{array}{|c|c|}\hline 1&1\\\hline 2\\\cline{1-1}\end{array}\quad \begin{array}{|c|c|}\hline 1&1\\\hline 3\\\cline{1-1}\end{array}\quad \begin{array}{|c|c|}\hline 1&2\\\hline 2\\\cline{1-1}\end{array}\quad \begin{array}{|c|c|}\hline 1&2\\\hline 3\\\cline{1-1}\end{array}\quad \begin{array}{|c|c|}\hline 1&3\\\hline 2\\\cline{1-1}\end{array}\quad \begin{array}{|c|c|}\hline 1&3\\\hline 3\\\cline{1-1}\end{array}\quad \begin{array}{|c|c|}\hline 2&2\\\hline 3\\\cline{1-1}\end{array}\quad \begin{array}{|c|c|}\hline 2&3\\\hline 3\\\cline{1-1}\end{array}$$

Given an SSYT in the alphabet $[n]$, its **content** is the n-tuple

$$(\text{number of 1's}, \ldots, \text{number of } n\text{'s}). \tag{3.1}$$

For example, the first SSYT displayed above has content $(2, 1, 0)$, while the second has content $(2, 0, 1)$.

We return to Young *diagrams*. Let $\mathcal{Y}(m,\, n-1)$ denote the set of Young diagrams with length at most m and width at most $n-1$; in other words,

$$\mathcal{Y}(m,\, n-1) = \big\{\text{Young diagrams that fit inside an } m \times (n-1) \text{ rectangle}\big\}.$$

We are especially interested in this set because of the bijection

$$\mathcal{H}(m,n) \longleftrightarrow \mathcal{Y}(m,\, n-1), \tag{3.2}$$

which we obtain by letting each data point in bin i contribute a row of length $i' := n - i$. By taking **conjugates** of Young diagrams (i.e., reflection across the main diagonal), we have these additional bijective correspondences:

$$\begin{array}{ccc} \mathcal{H}(m,n) & \longleftrightarrow & \mathcal{Y}(m,\, n-1) \\ \updownarrow & & \updownarrow \\ \mathcal{H}(n-1,\, m+1) & \longleftrightarrow & \mathcal{Y}(n-1,\, m) \end{array} \tag{3.3}$$

This duality between the values (m,n) and $(n-1,\, m+1)$ will appear throughout our results.

To make the correspondence (3.2) more precise, we define the **word** of a histogram. For $\lambda \in \mathcal{H}(m,n)$, define

$$\mathrm{w}(\lambda) := \underbrace{\boxed{1}\cdots\boxed{1}}_{\lambda_1 \text{ times}}\underbrace{\boxed{2}\cdots\boxed{2}}_{\lambda_2 \text{ times}}\cdots\underbrace{\boxed{n}\cdots\boxed{n}}_{\lambda_n \text{ times}},$$

namely, the unique SSYT of shape (m) with content λ. Recalling our abbreviation $i' := n - i$, we also define the "complementary" word

$$\mathrm{w}'(\lambda) := (\underbrace{1',\ldots,1'}_{\lambda_1 \text{ times}},\underbrace{2',\ldots,2'}_{\lambda_2 \text{ times}},\ldots,\underbrace{n',\ldots,n'}_{\lambda_n \text{ times}}),$$

which is a partition whose size, by (2.3), equals the weighted total $t(\lambda)$. Since $\mathrm{w}'(\lambda)$ is a partition, it has an associated Young diagram, which we denote by $\mathrm{Y}(\lambda)$. Clearly $\left|\mathrm{Y}(\lambda)\right| = t(\lambda)$. Moreover, $\mathrm{w}'(\lambda)$ has length at most m, and each of its parts is at most $1' = n-1$, so we have $\mathrm{Y}(\lambda) \in \mathcal{Y}(m,\, n-1)$. In this way, we can now easily write down the explicit correspondence in (3.2) as $\lambda \longleftrightarrow \mathrm{Y}(\lambda)$. Whenever convenient, therefore, we will identify a histogram λ with the Young diagram $\mathrm{Y}(\lambda)$.

Remark. We have defined $\mathrm{w}'(\lambda)$ only as a means to an end, that end being the Young diagram $\mathrm{Y}(\lambda)$. It is worth observing that $\mathrm{Y}(\lambda)$ is a cumulative histogram of λ, in the following sense: the length of the ith column of $\mathrm{Y}(\lambda)$ equals the mass contained in the first i' bins of λ.

Example 3.1. Let $\lambda = (2, 1, 0, 3, 1) \in \mathcal{H}(7, 5)$. Then we have

$$\mathrm{w}(\lambda) = \begin{array}{|c|c|c|c|c|c|c|}\hline 1 & 1 & 2 & 4 & 4 & 4 & 5 \\ \hline \end{array},$$

$$\mathrm{w}'(\lambda) = (4, 4, 3, 1, 1, 1, 0),$$

$$\mathrm{Y}(\lambda) = \begin{array}{|c|c|c|c|}\hline & & & \\ \hline & & & \\ \cline{1-4} & & \\ \cline{1-3} \\ \cline{1-1} \\ \cline{1-1} \\ \cline{1-1} \end{array}.$$

We observe that $t(\lambda) = \left|\mathrm{Y}(\lambda)\right| = 14$, and clearly $\mathrm{Y}(\lambda) \in \mathcal{Y}(7, 4)$.

3.2 The RSK correspondence

This section presents a previous result of [BW20], adapted to our notation. Its subject is the **Robinson–Schensted–Knuth (RSK) correspondence**, which furnishes a bijection

$$\left\{ \begin{array}{c} \text{ordered pairs of SSYT's} \\ \text{with the same shape,} \\ \text{in the alphabet } [n] \end{array} \right\} \longleftrightarrow \left\{ \begin{array}{c} n \times n \text{ matrices} \\ \text{with entries in } \mathbb{N} \end{array} \right\}. \tag{3.4}$$

We will require only a very special case; see [Ful97, Chapter 4] or [Sta99, Sections 7.11–7.14] for a more complete treatment of this remarkable correspondence. The connection with the EMD arises from the following idea in [BW20]: given two histograms $\lambda, \mu \in \mathcal{H}(m, n)$, their words $\mathrm{w}(\lambda)$ and $\mathrm{w}(\mu)$ are both SSYT's with shape (m) in the alphabet $[n]$. Therefore this pair $(\mathrm{w}(\lambda), \mathrm{w}(\mu))$ is an element of the left-hand side of (3.4). It turns out that the matrix corresponding to this pair (via RSK) is an optimal flow matrix F that solves the Monge problem and thus determines $\mathbf{EMD}(\lambda, \mu)$.

We now describe how this works. In the case of a pair of one-row SSYT's, the RSK correspondence is especially easy to implement; this is the only case we will need, since words of histograms are one-row SSYT's:

1. Regard $\mathrm{w}(\lambda)$ and $\mathrm{w}(\mu)$ as the rows of a $2 \times m$ matrix $W_{\lambda,\mu} := \begin{bmatrix} \mathrm{w}(\lambda) \\ \mathrm{w}(\mu) \end{bmatrix}$.
2. Construct the $n \times n$ matrix F so that F_{ij} equals the number of occurrences of the column $\begin{bmatrix} i \\ j \end{bmatrix}$ in $W_{\lambda,\mu}$.
3. This F is the RSK image of $(\mathrm{w}(\lambda),\ \mathrm{w}(\mu))$.

We claim that this F is the *unique* optimal matrix F described in Proposition 2.5. To prove this, we must first show that $F \in \mathcal{F}_{\lambda,\mu}$ with entries in $\mathbb{N}$. But this is obvious from the RSK construction above: the row-sum and column-sum vectors of F are (respectively) λ and μ, and hence the sum of all the F_{ij} equals m, with each $F_{ij} \in \mathbb{N}$. Finally, we must show that the support of F is a chain in $[n] \times [n]$. But since both $\mathrm{w}(\lambda)$ and $\mathrm{w}(\mu)$ are weakly increasing, the columns of $W_{\lambda,\mu}$ are also weakly increasing when regarded as elements of $[n] \times [n]$, and hence form a chain. Since this chain is precisely the support of F, and since the RSK correspondence is bijective, we are done.

Example 3.2. Let $\lambda = (1,7,0,1,1)$ and $\mu = (4,0,1,3,2)$. Then $\lambda, \mu \in \mathcal{H}(10,5)$. We first write down their words

$$\mathrm{w}(\lambda) = \boxed{1}\boxed{2}\boxed{2}\boxed{2}\boxed{2}\boxed{2}\boxed{2}\boxed{2}\boxed{4}\boxed{5},$$
$$\mathrm{w}(\mu) = \boxed{1}\boxed{1}\boxed{1}\boxed{1}\boxed{3}\boxed{4}\boxed{4}\boxed{4}\boxed{5}\boxed{5},$$

which gives us

$$W_{\lambda,\mu} = \begin{bmatrix} 1 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 4 & 5 \\ 1 & 1 & 1 & 1 & 3 & 4 & 4 & 4 & 5 & 5 \end{bmatrix}.$$

Now regarding each of these 10 columns $\begin{bmatrix} i \\ j \end{bmatrix}$ as the (i,j) position in F, we obtain

$$F = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 3 & 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Observe that the row and column sums of F recover λ and μ respectively, and the support of F is indeed a chain. Now by inspecting F and ignoring all but its nonzero entries, we conclude that

$$\begin{aligned}\mathbf{EMD}(\lambda,\mu) = \sum_{i,j} C_{ij}F_{ij} &= C_{11}\cdot 1 + C_{21}\cdot 3 + C_{23}\cdot 1 + C_{24}\cdot 3 + C_{45}\cdot 1 + C_{55}\cdot 1 \\ &= 0(1) + 1(3) + 1(1) + 2(3) + 1(1) + 0(1) \\ &= 11.\end{aligned}$$

As the example illustrates, the full matrix F is actually superfluous for the EMD computation: in practice, rather than constructing the sparse $n \times n$ matrix F, we can stop once we have the $2 \times m$ matrix $W_{\lambda,\mu}$, since the EMD is just the sum of the costs C_{ij} of the columns $\left[\begin{smallmatrix} i \\ j \end{smallmatrix}\right]$ of $W_{\lambda,\mu}$. Equivalently, if we regard words as m-dimensional vectors, then the EMD is just the taxicab distance between $\mathrm{w}(\lambda)$ and $\mathrm{w}(\mu)$. We record these facts for our use in the following section:

Lemma 3.3. *Let $\lambda, \mu \in \mathcal{H}(m,n)$, with $W_{\lambda,\mu} = \left[\begin{smallmatrix} \mathrm{w}(\lambda) \\ \mathrm{w}(\mu) \end{smallmatrix}\right] = \left[\begin{smallmatrix} i_1 & \cdots & i_m \\ j_1 & \cdots & j_m \end{smallmatrix}\right]$ as above. Then we have*

$$\mathbf{EMD}(\lambda,\mu) = \sum_{k=1}^{m} C_{i_k,j_k} = d_1\Big(\mathrm{w}(\lambda),\, \mathrm{w}(\mu)\Big).$$

3.3 EMD as a symmetric difference

Recall that the **symmetric difference** $\triangle$ of two sets X and Y is defined as

$$X \triangle Y := (X \cup Y) \backslash (X \cap Y),$$

the set difference between their union and their intersection. If we regard each Young diagram in $\mathcal{Y}(m,\, n-1)$ as a subset of the full $m \times (n-1)$ rectangle of cells, then it makes sense to say that a given cell is an element of several different Young diagrams; for example, the upper-left cell is an element of every Young diagram except for the empty diagram. Therefore we can speak of the symmetric difference of two Young diagrams.

The following proposition is the focal point of this chapter, and is essentially a combinatorial formulation of the fact (see [RDG11]) that the EMD of two histograms equals the ℓ_1-distance between their cumulative histograms.

Proposition 3.4. *Let $\lambda, \mu \in \mathcal{H}(m,n)$. Then* $\mathbf{EMD}(\lambda,\mu) = \left|\, \mathrm{Y}(\lambda) \triangle \mathrm{Y}(\mu) \,\right|$.

Proof. By Lemma 3.3, we have $\mathbf{EMD}(\lambda,\mu) = \sum_{k=1}^{m} C_{i_k,j_k}$, where $\left[\begin{smallmatrix} i_k \\ j_k \end{smallmatrix}\right]$ is the kth column of $W_{\lambda,\mu}$. Now, we observe that the cost matrix C is invariant under our "complement" operation $i \mapsto i' := n - i$:

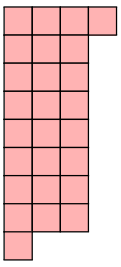
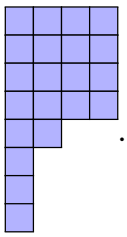
$$C_{i'j'} = |i' - j'| = |(n-i) - (n-j)| = |j - i| = |i - j| = C_{ij},$$

and thus we can rewrite Lemma 3.3 as $\mathbf{EMD}(\lambda,\mu) = \sum_k C_{i'_k,j'_k}$. But i'_k (respectively j'_k) is the kth component of $\mathrm{w}'(\lambda)$ (resp. $\mathrm{w}'(\mu)$), and therefore is the length of the kth row of $\mathrm{Y}(\lambda)$ (resp. $\mathrm{Y}(\mu)$). Hence $C_{i'_k,j'_k}$ is just the size of the symmetric difference of the kth row of $\mathrm{Y}(\lambda)$ and the kth row of $\mathrm{Y}(\mu)$. Summing over all m rows of the two Young diagrams, we conclude that $\mathbf{EMD}(\lambda,\mu) = \left|\, \mathrm{Y}(\lambda) \triangle \mathrm{Y}(\mu) \,\right|$. □

Example 3.5. We return to the histograms from Example 3.2, namely $\lambda = (1,7,0,1,1)$ and $\mu = (4,0,1,3,2)$. We have

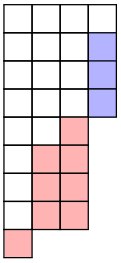
$$\mathrm{w}'(\lambda) = (4,3,3,3,3,3,3,3,1,0) \qquad \text{and} \qquad \mathrm{w}'(\mu) = (4,4,4,4,2,1,1,1,0,0),$$

and therefore

$\mathrm{Y}(\lambda) =$  and $\mathrm{Y}(\mu) =$  .

Incidentally, we see that $\mathbf{D}(\lambda,\mu) = 26 - 21 = 5$. We have shaded the diagrams red and blue respectively, to help visualize their symmetric difference, which we represent by the

set of shaded cells below:

$\mathrm{Y}(\lambda) \triangle \mathrm{Y}(\mu) =$ 

Counting the shaded cells, we find that $\left|\mathrm{Y}(\lambda) \triangle \mathrm{Y}(\mu)\right| = 11$, which agrees with our result for $\mathbf{EMD}(\lambda, \mu)$ from Example 3.2.

In light of Proposition 3.4, we can now describe not only the weighted difference, but also the EMD of two histograms, entirely in the language of Young diagrams:

$$\mathbf{D}(\lambda, \mu) = \left|\mathrm{Y}(\lambda)\right| - \left|\mathrm{Y}(\mu)\right| \qquad \text{and} \qquad \mathbf{EMD}(\lambda, \mu) = \left|\mathrm{Y}(\lambda) \triangle \mathrm{Y}(\mu)\right|.$$

3.4 A generating function

A **generating function** is a formal power series, in one or more indeterminates, whose coefficients encode desired information. Generating functions are a powerful combinatorial tool, especially when the closed formula for a desired sequence is impossible or impractical to write down. In [Kre20], a recursive definition is given for a generating function that keeps track of the number of histogram pairs with given EMD and given weighted totals. (This is a refinement of the generating function for EMD alone, defined in [BW20].) In light of the previous sections of this thesis, we can reinterpret these data combinatorially, as the symmetric difference and sizes of certain Young diagram pairs. It is somewhat surprising that symmetric differences of Young diagrams have not been studied in the combinatorics literature, and we believe this lends some interest to our reinterpretation of the generating function from [Kre20].

Because of the recursive nature of the definition in [Kre20], we must formally consider histogram pairs with unequal numbers of bins (n_1 and n_2). As noted in [BW20], the EMD is still defined on $\mathcal{H}(m, n_1) \times \mathcal{H}(m, n_2)$ since we can append zeros so that all histograms

are tuples of length $\max\{n_1,\ n_2\}$. We do not consider the statistical meaning of the EMD in such a scenario; the unequal bin numbers are just a means to obtain a recursion.

Proposition 3.6. *Formally, define*

$$H_{n_1,n_2} = H_{n_1,n_2}(q,x,y,z) \coloneqq \sum_{m=0}^{\infty} \left(\sum_{\substack{\lambda\in\mathcal{H}(m,n_1),\\ \mu\in\mathcal{H}(m,n_2)}} q^{\mathbf{EMD}(\lambda,\mu)} x^{t(\lambda)} y^{t(\mu)} \right) z^m. \tag{3.5}$$

Then we have the recursion

$$H_{n_1,n_2} = \frac{H_{n_1-1,n_2} + H_{n_1,n_2-1} - H_{n_1-1,n_2-1}}{1 - q^{C_{n_1,n_2}} x^{n_1-1} y^{n_2-1} z} \tag{3.6}$$

where $H_{1,1} = \frac{1}{1-z}$, *and* $H_{0,n_2} = H_{n_1,0} = 0$.

The proof given in [Kre20] is a specialization of the inclusion–exclusion principle used in [BW20]. Here, however, we give a short proof entirely in terms of Young diagrams.

Proof. First we reinterpret the generating function (3.5) in terms of Young diagrams, setting $X = \mathrm{Y}(\lambda)$ and $Y = \mathrm{Y}(\mu)$:

$$H_{n_1,n_2}(q,x,y,z) = \sum_{m=0}^{\infty} \left(\sum_{\substack{X\in\mathcal{Y}(m,\,n_1-1),\\ Y\in\mathcal{Y}(m,\,n_2-1)}} q^{|X\triangle Y|} x^{|X|} y^{|Y|} \right) z^m.$$

For the base case $H_{1,1}$ it is clear that $\mathcal{Y}(m,0)$ contains only the empty diagram, so the coefficient of z^m has only one term; since the empty diagram has size 0, the exponents of x and y in this term are both 0. Moreover, the symmetric difference of two empty diagrams is also empty, so the exponent of q is also 0. Hence the coefficient of z^m in $H_{1,1}$ is $q^0x^0y^0 = 1$, so $H_{1,1} = 1 + z + z^2 + \cdots = \frac{1}{1-z}$. For the second base case, in which n_1 or n_2 is 0, we have $\mathcal{Y}(m,\,-1) = \varnothing$, since a Young diagram cannot have a negative number of columns.

Now we consider H_{n_1,n_2} in general. For any pair (X,Y) in the inside sum, the diagram X has a certain number r_X of maximal rows (length $n_1 - 1$) at the top, and likewise Y has

a certain number r_Y of maximal rows (length $n_2 - 1$) at the top. Hence we can remove $r := \min\{r_X, r_Y\}$ maximal rows from each diagram. Since z tracks the number of rows in each diagram, each removed row-pair had contributed 1 to the exponent of z. Note that the size of the symmetric difference of a maximal row from X and a maximal row from Y is exactly $|n_1 - n_2| = C_{n_1,n_2}$, so each of the removed row-pairs had contributed C_{n_1,n_2} to the exponent of q. Moreover, each removed row-pair had contributed the size $n_1 - 1$ to the exponent of x, and $n_2 - 1$ to the exponent of y. Therefore we see that from each term in H_{n_1,n_2} we can factor out $(q^{C_{n_1,n_2}} x^{n_1-1} y^{n_2-1} z)^r$ for some r, leaving us with a pair of diagrams *at least one of which contains no maximal rows*; that is, at least one of them can fit inside a rectangle with one less column than before. Hence, starting with H_{n_1,n_2} and factoring out the reciprocal of the denominator in (3.6), the remaining terms are precisely those appearing in H_{n_1-1,n_2} or in H_{n_1,n_2-1}. After correcting for double-counting the terms appearing in both (namely H_{n_1-1,n_2-1}), we obtain the numerator of (3.6), and so we are done. □

Example 3.7. We first examine the specific case $H_2 := H_{2,2}$, which encodes information about pairs of diagrams in $\mathcal{Y}(m, 1)$, i.e., diagrams which are single columns of length m. Either by hand or by software, we use the recursion (3.6) to find

$$H_2 = \frac{1 - q^2 xyz^2}{(1-z)(1-xyz)(1-qxz)(1-qyz).}$$

We now interpret this in terms of Young diagram pairs. In the denominator, each of the four factors is the denominator of a geometric series, and corresponds to one of the ways to build a diagram pair by adding a row to each diagram X and Y: namely, $(1 - z)$ adds an empty row to both diagrams, $(1 - xyz)$ adds a 1-cell row to both diagrams, $(1 - qxz)$ adds a 1-cell row to X only (hence contributing 1 to the symmetric difference), and $(1 - qyz)$ adds a 1-cell row to Y only (again contributing 1 to the symmetric difference). But we cannot choose just any combination of these four row-pairs and still obtain a pair of Young diagrams: specifically, we cannot choose row-pairs from *both* the third and fourth factors,

since it is impossible for a Young diagram to have a nonempty row below an empty row. This relation necessitates the numerator term $-q^2xyz^2$, which effectively kills any term in the series containing the product $(qxz)(qyz)$.

Example 3.8. For larger values of n, the closed form of $H_n := H_{n,n}$ spills over into several lines, or pages, of text; an explicit (as opposed to recursive) formula involves an enormous alternating sum of least common multiples, ranging over the power set of a certain set of monomials (see the partial order we define below, and consider the set of noncomparable pairwise products). As an illustrative example, we take $n = 3$, corresponding to Young diagrams in $\mathcal{Y}(m, 2)$, and we inspect the denominator of H_3:

$$(1-t)(1-xyz)(1-x^2y^2z)(1-qxz)(1-qyz)(1-qx^2yz)(1-qxy^2z)(1-q^2x^2z)(1-q^2y^2z)$$

Again, each of these factors describes one way to choose a row in each diagram simultaneously, where the rows may now contain 0, 1, or 2 cells; notice that the exponent of q is always the positive difference between those of x and y. But again, we cannot choose just any combination of row-pairs to build a pair of Young diagrams. This time, the "forbidden" combinations seem more daunting to describe; to resolve this, we put a partial order on monomials (ignoring the exponents of q and z), such that $x^iy^j \preceq x^ky^\ell$ if and only if $i \leqslant k$ and $j \leqslant \ell$. We now observe that two given row-pairs can coexist in X and Y if and only if their monomials are comparable under our partial order. (This is a combinatorial translation of the "matrix straightening" from [BW20] via modding out by a determinantal ideal.) Hence the "forbidden" combinations of factors are precisely those which contain the product of two noncomparable monomials from the denominator. This necessitates an alternating sum of terms in the numerator, which (despite many cancellations) quickly becomes intractable as n increases, making the recursive definition the only viable computational approach.

The coefficient of z^2 in the expansion of H_3 is

$$\begin{aligned}
&1 + xy + 2x^2y^2 + x^3y^3 + x^4y^4 \\
&+ q(x + y + 2x^2y + 2xy^2 + 2x^3y^2 + 2x^2y^3 + x^4y^3 + x^3y^4) \\
&+ q^2(2x^2 + x^3y + 2y^2 + 2x^2y^2 + 2x^4y^2 + xy^3 + 2x^2y^4) \\
&+ q^3(x^3 + x^4y + y^3 + xy^4) \\
&+ q^4(x^4 + y^4).
\end{aligned}$$

The reader can verify that these terms indeed encode all 36 ordered pairs of Young diagrams fitting inside a 2×2 rectangle, along with the sizes of the diagrams and the symmetric difference between them. (Also compare this with the top-center frequency plot in Figure 2, where $m = 2$ and $n = 3$.)

3.5 Plotting D vs. EMD

Consider the generating function $H_n := H_{n,n}$ defined in (3.5), with the recursive formula given in (3.6). If we make the substitution $y \mapsto x^{-1}$, then we obtain $x^{t(\lambda)}y^{t(\mu)} \mapsto x^{t(\lambda)-t(\mu)} = x^{\mathbf{D}(\lambda,\mu)}$, so that the exponent of x in each term now keeps track of the weighted difference **D**. Programming this modified generating function in Mathematica, we can then organize the coefficients into two-dimensional frequency plots, with **D** on the horizontal axis and **EMD** on the vertical axis. Specifically, for given values of m and n, the coefficient of z^m in H_n is itself a polynomial in x and q, in which the coefficient of $x^a q^b$ is the number of histogram pairs in $\mathcal{H}(m,n) \times \mathcal{H}(m,n)$ whose weighted difference equals a and whose EMD equals b. Our code then plots this coefficient of $x^a q^b$ at the point (a,b), giving us a visualization of the distribution of **D**-values and **EMD**-values among pairs of histograms from $\mathcal{H}(m,n)$. For example, the pair $(\lambda,\mu) \in \mathcal{H}(10,5) \times \mathcal{H}(10,5)$ from Example 3.2 had a weighted difference of 5 and an EMD of 11; therefore (λ,μ) contributes one to the point plotted at the coordinates $(5,11)$ on the lower-right plot in Figure 3.

We now take the reader on a brief guided tour of these plots, aided by Figures 2 and 3. We also include a three-dimensional rendering in Figure 4. The observations below provide a sort of dictionary between the statistics of histograms and the combinatorics of Young diagram pairs; ultimately, in light of Section 3.3, everything below is a statement about the interplay of two Young diagrams' symmetric difference with their difference in size.

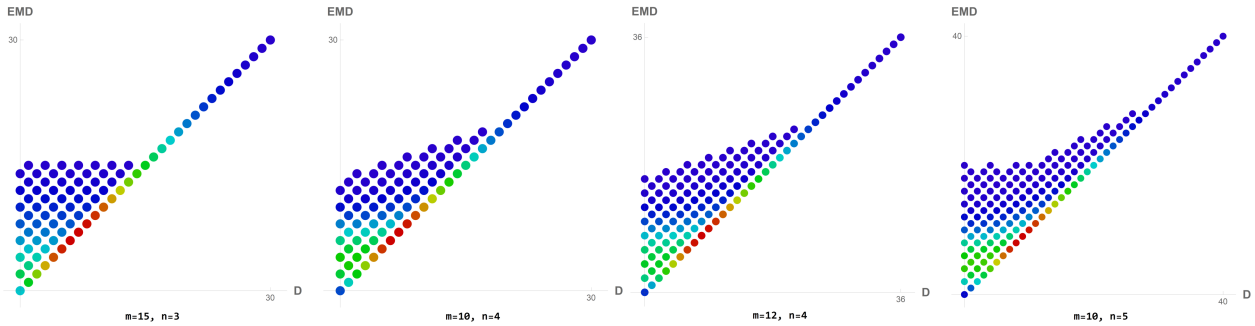


Figure 3: Frequency plots for larger values of m and n. Dark blue = minimum and red = maximum.

- By definition, histogram pairs (λ, μ) lying further to the right (higher positive **D**-values) are more "lopsided" (in terms of weighted total) in favor of λ, and vice versa. But since the plot is symmetric about the vertical axis, without loss of generality we can restrict our attention to the right-hand side, i.e., those pairs for which $t(\lambda) \geqslant t(\mu)$. We therefore refer only to the right-hand half of the plots in the remainder of this list.

- The plot for $\mathcal{H}(m, n)$ is identical to that for $\mathcal{H}(n-1,\ m+1)$. (Compare, for instance, the first two plots in the second row of Figure 2; this is the only redundancy we have included in the figures.) This duality is not at all obvious *a priori* from the statistical definitions of **EMD** and **D**, but it is immediate when we consider the conjugates of Young diagrams, as in (3.3): $\left|\, \mathrm{Y}(\lambda) \triangle \mathrm{Y}(\mu) \,\right|$ and $\left|\, \mathrm{Y}(\lambda) \,\right|$ and $\left|\, \mathrm{Y}(\mu) \,\right|$ are all invariant under conjugates, so the **D**-vs.-**EMD** plots are as well.

- In general, **EMD** $\geqslant$ **D** since $\mathrm{Y}(\lambda) \triangle \mathrm{Y}(\mu) \supseteq \mathrm{Y}(\lambda) \setminus \mathrm{Y}(\mu)$.

- The values of both **EMD** and **D** range from 0 to $m(n-1)$. This is obvious when we consider the $m \times (n-1)$ rectangle containing every diagram in $\mathcal{Y}(m,\, n-1)$.

- In each plot, the upper-rightmost data point corresponds to the unique pair (λ, μ) such that $\mathrm{Y}(\lambda)$ is the full $m \times (n-1)$ rectangle and $\mathrm{Y}(\mu)$ is the empty diagram. Starting from here, moving one unit to the left corresponds to either subtracting one cell from $\mathrm{Y}(\lambda)$ or adding one cell to $\mathrm{Y}(\mu)$, in such a way that both remain in $\mathcal{Y}(m,\, n-1)$.

- The plots make clear that for any histogram pair, **D** and **EMD** have the same parity as each other. This reflects the fact that adding or removing one cell in either diagram changes the symmetric difference by exactly 1.

- The "tail" along the diagonal on the far right reveals that after a certain threshold when **D** is sufficiently large, **D** and **EMD** coincide with each other. In the language of Young diagrams X and Y, this equality $|X \triangle Y| = |X| - |Y|$ implies $X \supseteq Y$: thus for any pair (λ, μ) on the diagonal, we have $\mathrm{Y}(\lambda) \supseteq \mathrm{Y}(\mu)$.

  We claim that this threshold is $\mathbf{D} = (m-1)(n-2)$. To see this, start from the far right at $\mathbf{D} = m(n-1)$, where $\mathrm{Y}(\lambda)$ is the full rectangle and $\mathrm{Y}(\mu)$ is empty. The most efficient way to make $\mathrm{Y}(\mu)$ spill outside $\mathrm{Y}(\lambda)$ is to remove the m cells from the rightmost column of $\mathrm{Y}(\lambda)$ and to add $n-1$ cells to the top row of $\mathrm{Y}(\mu)$; at this point, the cell in the upper-right will belong to $\mathrm{Y}(\mu)$ but not to $\mathrm{Y}(\lambda)$. Therefore, since moving left from $m(n-1)$ by $(m+n-1)$ steps takes us to the vertical line $\mathbf{D} = mn - 2m - n + 1$, the "tail" must begin just to the right of this **D**-value, namely $\mathbf{D} = mn - 2m - n + 2 = (m-1)(n-2)$.

- We immediately observe that the highest concentration of histogram pairs occurs relatively close to the origin, on the diagonal. These are pairs in which one Young diagram contains the other, with a size difference of only a "few" cells. This observation naturally raises two statistical questions: (1) What proportion of pairs lie along the diagonals of the plots? (2) Clearly the expected value of **D** is zero due to

the symmetry of the plots, but what is the expected value of its absolute value $|\mathbf{D}|$? (The expected value of **EMD** has been derived in [BW20].) In the next section, we answer both these questions using the combinatorics of Young diagrams.

3.6 Statistical answers via Young diagrams

In this section, we use the combinatorics of Young diagams to answer two statistical questions about histogram pairs:

1. With what probability does $\mathbf{EMD} = |\mathbf{D}|$?
2. What is the expected value of $|\mathbf{D}|$?

3.6.1 Preliminaries: plane partitions

To answer the first question, we will use a generalization of Young diagrams known as plane partitions. A **plane partition** is a Young tableau in which the entries weakly *decrease* along each row and column. A plane partition is visualized three-dimensionally by stacking cubes on a Young diagram such that the height above each cell is its integer entry; for this reason, a plane partition whose underlying Young diagram has at most x rows and y columns, with entries at most z, is said to fit inside an $x \times y \times z$ box. The number of such plane partitions is

$$\mathrm{PP}(x, y, z) = \prod_{i=1}^{x} \prod_{j=1}^{y} \prod_{k=1}^{z} \frac{i+j+k-1}{i+j+k-2.} \tag{3.7}$$

(This formula was proved by MacMahon in [Mac16]. See also [Sta99, Sections 7.20–22] for an expansive treatment of the combinatorics of plane partitions.) For our particular problem, we will appeal to the following specialization:

Lemma 3.9. *In the case $z = 2$, we have*

$$\mathrm{PP}(x, y, 2) = \frac{(x+1)\cdots(x+y)\cdot(x+2)\cdots(x+y+1)}{y!(y+1)!} \tag{3.8}$$

where the omitted factors increase by 1.

Proof. Setting $z = 2$, we separate (3.7) into two main factors (one for $k = 1$ and the other for $k = 2$) and cancel the repeated factors:

$$\mathrm{PP}(x, y, 2) = \left(\prod_{i,j} \frac{i+j}{i+j-1}\right)\left(\prod_{i,j} \frac{i+j+1}{i+j}\right) = \prod_{i,j} \frac{i+j+1}{i+j-1.}$$

Now we use induction on x and y to prove that this product is equivalent to (3.8). In the base case $x = y = 1$, the product above is just the single factor $\frac{1+1+1}{1+1-1} = 3$, and likewise the expression in (3.8) is $\frac{(1+1)(1+2)}{1!2!} = \frac{6}{2} = 3$. (Note that the second string of factors in the numerator of (3.8) is the empty product 1 in the case $x = y = 1$.)

To induct on x, assume that (3.8) holds for $x - 1$ and y. Then we have

$$\begin{aligned}
\mathrm{PP}(x, y, 2) &= \prod_{i=1}^{x}\prod_{j=1}^{y} \frac{i+j+1}{i+j-1} \\
&= \prod_{j=1}^{y} \frac{x+j+1}{x+j-1} \cdot \mathrm{PP}(x-1, y, 2) \\
&= \frac{(x+y)(x+y+1)}{(x)(x+1)} \cdot \frac{(x)\cdots(x+y-1)\cdot(x+1)\cdots(x+y)}{y!(y+1)!} \\
&= \frac{(x+1)\cdots(x+y)\cdot(x+2)\cdots(x+y+1)}{y!(y+1)!}
\end{aligned}$$

as in (3.8). The induction on y works out similarly. □

Remark. The expression (3.8) can be written more elegantly using the Pochhammer symbol $(a)_b := \frac{\Gamma(a+b)}{\Gamma(a)} = a(a+1)\cdots(a+b-1)$, as follows: $\mathrm{PP}(x, y, 2) = \frac{(x+1)_y (x+1)_{y+1}}{(x+1)y!(y+1)!}$.

3.6.2 Probability that **EMD** = |**D**|

As we observed at the end of Section 3.5, there is a relatively high concentration along the diagonals $y = |x|$ of our **D**-vs.-**EMD** plots. In other words, if we choose an element uniformly at random from $\mathcal{H}(m, n) \times \mathcal{H}(m, n)$, there is a (perhaps surprisingly) high probability that its EMD is nothing other than the absolute value of its weighted difference **D**. We now determine this probability combinatorially.

Theorem 3.10. *Let $(\lambda, \mu) \in \mathcal{H}(m,n) \times \mathcal{H}(m,n)$ be chosen uniformly at random. Then*

$$\mathbb{P}\Big(\mathbf{EMD}(\lambda,\mu) = |\mathbf{D}(\lambda,\mu)|\Big) = \frac{2(m+n)}{n(m+1)} - \frac{(n-1)!}{(m+1)\cdots(m+n-1).}$$

Proof. Translating into Young diagrams, we have $\mathbf{EMD}(\lambda,\mu) = |\mathbf{D}(\lambda,\mu)|$ if and only if one of the two Young diagrams $\mathrm{Y}(\lambda)$ and $\mathrm{Y}(\mu)$ contains the other. We claim that there is a bijective correspondence

$$\left\{ \begin{array}{c} \text{unordered pairs of Young} \\ \text{diagrams from } \mathcal{Y}(m,\ n-1) \\ \text{such that one diagram} \\ \text{contains the other} \end{array} \right\} \longleftrightarrow \left\{ \begin{array}{c} \text{plane partitions} \\ \text{fitting in an} \\ m \times (n-1) \times 2 \text{ box} \end{array} \right\}.$$

To see this, consider a pair of diagrams in the left-hand set and superimpose them so that their upper-left cells coincide. Filling any cell common to both diagrams with a "2," and all the others with a "1," we obtain a plane partition in the right-hand set. In the opposite direction, starting with a plane partition from the right-hand set, we produce a unique pair of Young diagrams: the larger diagram is just the underlying diagram of the plane partition, and the smaller diagram is the collection of cells with the entry "2."

Therefore the number of unordered histogram pairs such that $\mathbf{EMD} = |\mathbf{D}|$ is equal to $\mathrm{PP}(m,\ n-1,\ 2)$. By Lemma 3.9, this number is

$$\mathrm{PP}(m,\ n-1,\ 2) = \frac{(m+1)\cdots(m+n-1)\cdot(m+2)\cdots(m+n)}{n!(n-1)!}.$$

To find the number of such *ordered* pairs in $\mathcal{H}(m,n) \times \mathcal{H}(m,n)$, we multiply by 2 to account for both pairs (λ,μ) and (μ,λ), and then correct for the pairs (λ,λ) by subtracting $\#\mathcal{H}(m,n) = \binom{m+n-1}{n-1} = \frac{(m+1)\cdots(m+n-1)}{(n-1)!}$. Hence the number of ordered pairs such that $\mathbf{EMD} = |\mathbf{D}|$ is

$$2 \cdot \mathrm{PP}(m,\ n-1,\ 2) - \frac{(m+1)\cdots(m+n-1)}{(n-1)!}.$$

m \ n	2	3	4	5	6	7	8	9	10
10	1	0.773	0.633	0.544	0.485	0.441	0.409	0.384	0.364
20	1	0.726	0.571	0.476	0.413	0.367	0.333	0.307	0.286
30	1	0.708	0.548	0.452	0.387	0.341	0.306	0.280	0.258
40	1	0.698	0.537	0.439	0.374	0.328	0.293	0.266	0.244
50	1	0.692	0.529	0.431	0.366	0.319	0.284	0.257	0.235
60	1	0.688	0.525	0.426	0.361	0.314	0.279	0.251	0.230
70	1	0.685	0.521	0.423	0.357	0.310	0.275	0.247	0.225
80	1	0.683	0.519	0.420	0.354	0.307	0.272	0.244	0.222
90	1	0.681	0.516	0.418	0.352	0.305	0.269	0.242	0.220
100	1	0.680	0.515	0.416	0.350	0.303	0.267	0.240	0.218
10,000	1	0.667	0.500	0.400	0.333	0.286	0.250	0.222	0.200

Table 1: Values of $\mathbb{P}(\mathbf{EMD} = |\mathbf{D}|)$ on $\mathcal{H}(m,n) \times \mathcal{H}(m,n)$. Obtained in Mathematica using Theorem 3.10.

Finally, to obtain the probability, we divide by $|\mathcal{H}(m,n)|^2$, and after simplifying we obtain

$$\frac{2 \cdot \mathrm{PP}(m,\, n-1,\, 2) - \frac{(m+1)\cdots(m+n-1)}{(n-1)!}}{\left(\frac{(m+1)\cdots(m+n-1)}{(n-1)!}\right)^2}$$

$$= \frac{\frac{2(m+1)\cdots(m+n-1)\cdot(m+2)\cdots(m+n)}{n!(n-1)!}}{\left(\frac{(m+1)\cdots(m+n-1)}{(n-1)!}\right)^2} - \frac{\frac{(m+1)\cdots(m+n-1)}{(n-1)!}}{\left(\frac{(m+1)\cdots(m+n-1)}{(n-1)!}\right)^2}$$

$$= \frac{2(m+n)}{n(m+1)} - \frac{(n-1)!}{(m+1)\cdots(m+n-1).}$$ □

See Table 1, which records this probability for selected values of m and n. Moreover, taking the limit in Theorem 3.10 as $m \to \infty$, we immediately observe the following, which is confirmed by the bottom row of Table 1:

Corollary 3.11. *For fixed n, as $m \to \infty$, the probability in Theorem 3.10 is asymptotic to $2/n$.*

Remark. Suppose we define normalized versions of the EMD and weighted total, namely

$$\widetilde{\mathbf{EMD}}(\lambda,\mu) := \frac{\mathbf{EMD}(\lambda,\mu)}{m} \quad \text{and} \quad \widetilde{t}(\lambda) := \frac{t(\lambda)}{m},$$

so that both statistics take values in the interval $[0,\ n-1]$. (In this way, $\widetilde{t}$ is a weighted *average*; for example, a class GPA.) Then Corollary 3.11 becomes a statement about *probability distributions* (rather than discrete histograms) on the set $[n]$. Indeed, both $\widetilde{\mathbf{EMD}}$ and $\widetilde{t}$ (and therefore $\widetilde{\mathbf{D}}$) have natural analogues on probability distributions on $[n]$, obtained by letting $m \to \infty$. Then for a randomly chosen ordered pair (X, Y) of probability distributions on $[n]$, we have $\mathbb{P}\Big(\widetilde{\mathbf{EMD}}(X,Y) = |\widetilde{\mathbf{D}}(X,Y)|\Big) = 2/n$. We find it noteworthy that this probabilistic result is so easily obtained using only the combinatorics of plane partitions.

3.6.3 Expected value of $|\mathbf{D}|$

The second problem from this section was to find the expected value of $|\mathbf{D}(\lambda,\mu)|$, assuming the uniform distribution on $\mathcal{H}(m,n) \times \mathcal{H}(m,n)$. In order to express this result, we recall the q-analogue of the binomial coefficients $\binom{a}{b}$. First we define $[a]_q := 1 + q + \cdots + q^{a-1}$. Then writing $[a]_q! := [a]_q[a-1]_q \cdots [1]_q$, we define the **$q$-binomial coefficient**

$$\begin{bmatrix} a \\ b \end{bmatrix}_q := \frac{[a]_q!}{[b]_q![a-b]_q!}.$$

This is actually a polynomial (not just a rational function) in q, which encodes combinatorial information central to this thesis:

$$\begin{bmatrix} m+n-1 \\ m \end{bmatrix}_q = \sum_{Y \in \mathcal{Y}(m,\,n-1)} q^{|Y|} = \sum_{\lambda \in \mathcal{H}(m,n)} q^{t(\lambda)}. \tag{3.9}$$

The first equality in (3.9) is Proposition 1.7.3 in [Sta12], and the second equality follows from the fact that $\Big|\,\mathrm{Y}(\lambda)\,\Big| = t(\lambda)$. In other words, the coefficient of q^t in $\begin{bmatrix} m+n-1 \\ m \end{bmatrix}_q$ is the number of Young diagrams in $\mathcal{Y}(m,\ n-1)$ with size t, or equivalently, the number of histograms in $\mathcal{H}(m,n)$ with weighted total t.

We will borrow the "plussing" operator $[\quad]_+$ from the theory of economic modeling; this operator, also called a "linear annihilation operator" in [HS80], acts on a Laurent series in q by annihilating negative powers of q:

$$\left[\sum_{k=-\infty}^{\infty} c_k q^k\right]_+ := \sum_{k=0}^{\infty} c_k q^k.$$

(See [HS80, p. 14], and the original Wiener–Kolmogorov prediction formula dating back to [Wie49, Kol41].)

Theorem 3.12. *The expected value of* $|\mathbf{D}|$ *on* $\mathcal{H}(m,n) \times \mathcal{H}(m,n)$ *equals*

$$\frac{2}{\binom{m+n-1}{m}^2} \cdot \frac{d}{dq} \left[\begin{bmatrix} m+n-1 \\ m \end{bmatrix}_q \begin{bmatrix} m+n-1 \\ m \end{bmatrix}_{q^{-1}} \right]_+ \Bigg|_{q=1.}$$

Proof. We work inside–out, first using (3.9) to expand the product of the q-binomial coefficients, a Laurent polynomial which we abbreviate by $f(q)$:

$$\begin{aligned} f(q) = \begin{bmatrix} m+n-1 \\ m \end{bmatrix}_q \begin{bmatrix} m+n-1 \\ m \end{bmatrix}_{q^{-1}} &= \sum_{\lambda \in \mathcal{H}(m,n)} q^{t(\lambda)} \sum_{\mu \in \mathcal{H}(m,n)} q^{-t(\mu)} \\ &= \sum_{\lambda,\mu} q^{t(\lambda)-t(\mu)} \\ &= \sum_{\lambda,\mu} q^{\mathbf{D}(\lambda,\mu)}. \end{aligned}$$

(We observe in passing that $f(q)$ is the generating function for the column sums of our **D**-vs.-**EMD** plots in Section 3.5.) Now applying the plussing operator, we have

$$\begin{aligned} [f(q)]_+ &= \sum_{t(\lambda) \geqslant t(\mu)} q^{\mathbf{D}(\lambda,\mu)} \\ &= \sum_{k \geqslant 0} c_k \, q^k \end{aligned}$$

where $c_k = \#\{(\lambda,\mu) \mid \mathbf{D}(\lambda,\mu) = k\}$. Hence

$$\frac{d}{dq}[f(q)]_+ = \sum_{k \geqslant 0} (c_k \cdot k) q^{k-1}$$

and therefore, since by the symmetry $(\lambda,\mu) \leftrightarrow (\mu,\lambda)$ we must have $c_k = c_{-k}$, we obtain

$$\begin{aligned} \frac{d}{dq}[f(q)]_+ \Bigg|_{q=1} &= \sum_{k \geqslant 0} c_k \cdot k \\ &= \sum_{k \geqslant 0} c_{-k} \cdot k \\ &= \frac{1}{2} \sum_{k > 0} \#\Big\{ (\lambda,\mu) \;\Big|\; |\mathbf{D}(\lambda,\mu)| = k \Big\} \cdot k. \end{aligned}$$

m \ n	2	3	4	5	6	7	8	9	10
10	0.364	0.266	0.224	0.201	0.185	0.174	0.166	0.159	0.154
20	0.349	0.250	0.208	0.183	0.167	0.155	0.146	0.139	0.133
30	0.344	0.245	0.202	0.177	0.160	0.148	0.139	0.132	0.126
40	0.341	0.242	0.199	0.174	0.157	0.145	0.135	0.128	0.122
50	0.340	0.240	0.197	0.172	0.155	0.143	0.133	0.126	0.119
60	0.339	0.239	0.196	0.171	0.154	0.141	0.132	0.124	0.118
70	0.338	0.238	0.195	0.170	0.153	0.140	0.130	0.123	0.116
80	0.337	0.238	0.194	0.169	0.152	0.139	0.130	0.122	0.115
90	0.337	0.237	0.194	0.168	0.151	0.139	0.129	0.121	0.115
100	0.337	0.237	0.193	0.168	0.151	0.138	0.128	0.121	0.114

Table 2: Expected value of $|\mathbf{D}|$ on $\mathcal{H}(m,n) \times \mathcal{H}(m,n)$, unit normalized by dividing by $m(n-1)$. Obtained in Mathematica using the combinatorial formula in Theorem 3.12.

(Note that we omit the term corresponding to $k = 0$ since it contributes 0 to the sum; in this way, we avoid the need to correct for double-counting pairs (λ, λ).) Now if we multiply both sides by 2, then the right-hand side is just the expected value of $|\mathbf{D}(\lambda, \mu)|$ multiplied by $\#\mathcal{H}(m,n)^2$, so we are done. □

Remark. As before, we can normalize this result by dividing by m, or equivalently, by using $\sqrt[m]{q}$ in place of q. Then the asymptotic as $m \to \infty$ equals the expected value of $|\widetilde{\mathbf{D}}|$ on pairs of probability distributions on $[n]$.

Using the combinatorial formula in Theorem 3.12, we can quickly calculate expected values of $|\mathbf{D}|$ in Mathematica. See Table 2, in which we have *unit* normalized to obtain values between 0 and 1. For example, from Table 2, we see that the expected GPA difference for two classes of 30 students equals $4(0.177) = 0.708$. We have double-checked the values in both Tables 1 and 2 by brute force in Mathematica, for sufficiently small m and n. Of course, brute force quickly becomes impractical, since (for example) $\mathcal{H}(100,10) \times \mathcal{H}(100,10)$ contains approximately 1.8×10^{25} pairs; and yet the combinatorial approach allows us to compute both statistics over all these pairs in just a few seconds.

3.7 EMD and straightening on Young's lattice

In this section we express **EMD** directly in terms of weighted difference **D**, and thereby establish the recurring theme of "straightening" in this thesis. The first of several contexts for this notion is that of Young's lattice.

The set $\mathcal{Y}$ of all Young diagrams admits a partial order, defined by inclusion. Equipped with this partial order, the poset $\mathcal{Y}$ has the structure of a distributive lattice, known as **Young's lattice**. This lattice $\mathcal{Y}$ is **ranked** with respect to the size of Young diagrams, so that $\operatorname{rank} X = |X|$ for $X \in \mathcal{Y}$. As usual, for two lattice elements x, y, we define their **join** $x \vee y$ to be their least upper bound, and their **meet** $x \wedge y$ to be their greatest lower bound. See [Sta99, Chapter 7]; we provide a truncated picture of the Hasse diagram of $\mathcal{Y}$ in Figure 5, up to rank 5, where the ith row from the bottom corresponds to elements of rank i.

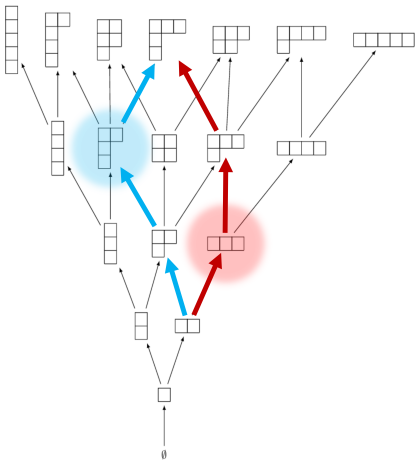


Figure 5: Hasse diagram of Young's lattice $\mathcal{Y}$, shown up to rank 5.

The set $\mathcal{Y}(m, n-1)$ forms a sublattice of $\mathcal{Y}$, namely the interval between the empty diagram and the $m \times (n-1)$ rectangular diagram. We can continue, as we have throughout this chapter, to regard diagrams in $\mathcal{Y}(m,\ n-1)$ as histograms in $\mathcal{H}(m, n)$, via the correspondence $\lambda \leftrightarrow \mathrm{Y}(\lambda)$. Under this identification, without risk of confusion, we can regard $\mathcal{H}(m, n)$ itself as a poset, with partial order induced by that of $\mathcal{Y}(m,\ n-1)$. Hence it

makes sense to write (for example) $\lambda \vee \mu$, which actually means the histogram $\nu \in \mathcal{H}(m,n)$ such that $\mathrm{Y}(\nu) = \mathrm{Y}(\lambda) \vee \mathrm{Y}(\mu)$. It is easy to translate $\mathbf{D}$ in terms of Young's lattice: for $\lambda, \mu \in \mathcal{H}(m,n)$, we have $\mathbf{D}(\lambda,\mu) = \operatorname{rank}\mathrm{Y}(\lambda) - \operatorname{rank}\mathrm{Y}(\mu)$, namely, the vertical "distance" between $\mathrm{Y}(\lambda)$ and $\mathrm{Y}(\mu)$ on the Hasse diagram of $\mathcal{Y}$. In fact, the lattice structure allows us to express the EMD entirely in terms of $\mathbf{D}$:

Proposition 3.13. *Let $\lambda, \mu \in \mathcal{H}(m,n)$. Then $\mathbf{EMD}(\lambda,\mu) = \mathbf{D}(\lambda \vee \mu,\ \lambda \wedge \mu)$.*

Proof. Let $X, Y \in \mathcal{Y}(m,\ n-1)$. Since the partial order on $\mathcal{Y}$ is defined by inclusion of Young diagrams, we have $X \vee Y = X \cup Y$ and $X \wedge Y = X \cap Y$. On the other hand, by definition of symmetric difference, we have $\left|X \triangle Y\right| = \left|(X \cup Y) \backslash (X \cap Y)\right| = |X \cup Y| - |X \cap Y|$, since $(X \cup Y) \supseteq (X \cap Y)$. Combining this with Proposition 3.4, we conclude that

$$\begin{aligned}
\mathbf{EMD}(\lambda,\mu) &= \left|\, \mathrm{Y}(\lambda) \triangle \mathrm{Y}(\mu) \,\right| \\
&= \left|\, \mathrm{Y}(\lambda) \cup \mathrm{Y}(\mu) \,\right| - \left|\, \mathrm{Y}(\lambda) \cap \mathrm{Y}(\mu) \,\right| \\
&= \operatorname{rank}(\lambda \vee \mu) - \operatorname{rank}(\lambda \wedge \mu).
\end{aligned}$$

□

As an example, see Figure 5, and assume $m \geqslant 3$ and $n \geqslant 4$. Suppose we have highlighted $\mathrm{Y}(\lambda)$ in blue and $\mathrm{Y}(\mu)$ in red. Since they are one row apart, we have $\mathbf{D}(\lambda,\mu) = 1$. We have added darker arrows to illustrate their meet and join; since the meet and join are three rows apart, we see that $\mathbf{EMD}(\lambda,\mu) = 3$.

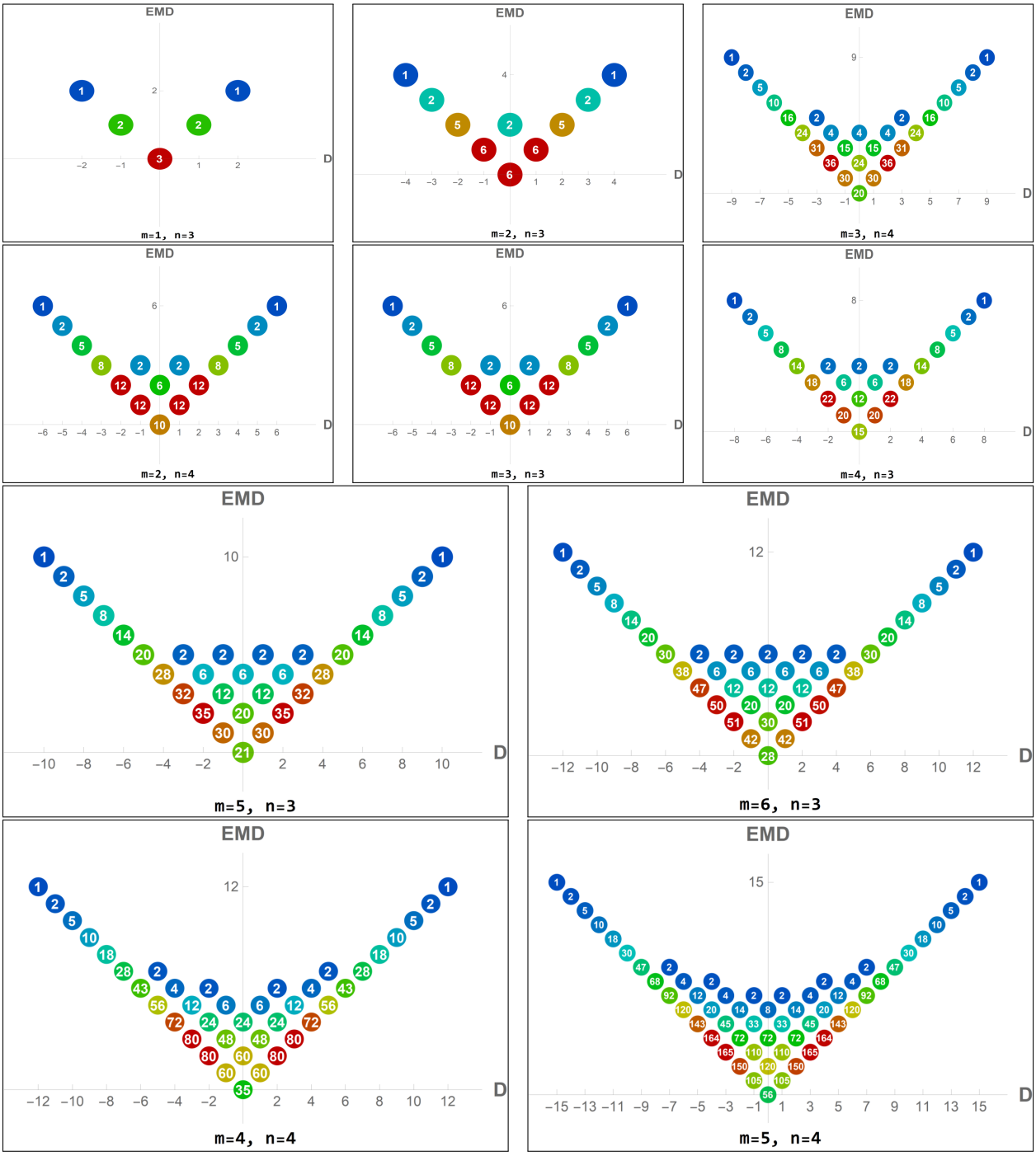


Figure 2: Frequency plots for **D** vs. **EMD**.

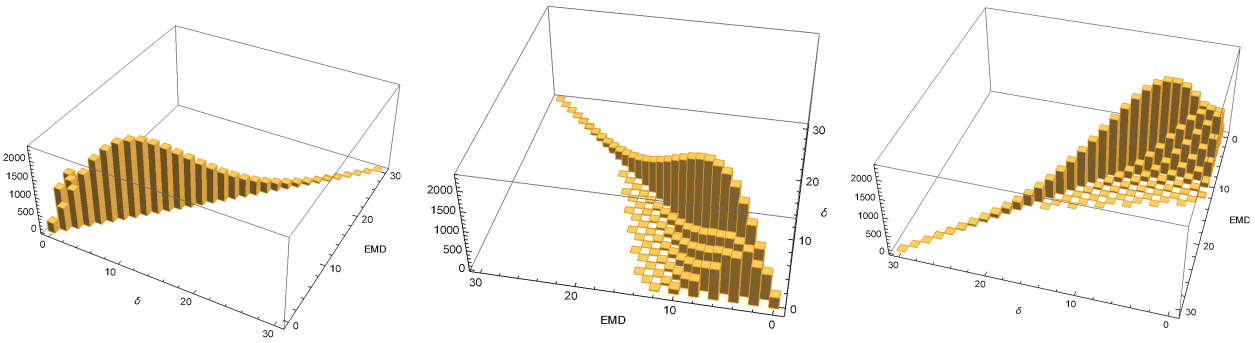


Figure 4: A three-dimensional rendering of (the right half of) the plot for $m = 10$ and $n = 4$ (dually, for $m = 3$ and $n = 11$), seen from three angles.

4 THE EMD AND THE FIRST WALLACH REPRESENTATION

In this chapter, we interpret the EMD in the context of the representation theory of the special indefinite unitary group $\mathrm{SU}(p,q)$. In particular we study a certain infinite-dimensional module known as the first Wallach representation. Our main result in this chapter is Theorem 4.15, which states that in the context of the first Wallach representation, the second fundamental theorem of classical invariant theory is an analogue of the northwest corner rule from Section 2.3. In addition, we simplify the issue of *partial matching* (i.e., the EMD for unequal histogram masses) by viewing the problem through the lens of Howe duality, and "separating variables" into invariant and harmonic components.

4.1 SU(*p*,*q*) and other classical groups

Let $\mathrm{M}_{n,k}$ denote the set of complex $n \times k$ matrices, and $\mathrm{M}_n := \mathrm{M}_{n,n}$. We begin by recalling the **general linear group** $\mathrm{GL}_n = \{g \in \mathrm{M}_n \mid \det g \neq 0\}$ of invertible $n \times n$ matrices under matrix multiplication. The **special linear group** $\mathrm{SL}_n = \{g \in \mathrm{GL}_n \mid \det g = 1\}$ is the subgroup of matrices with unit determinant. Both GL_n and SL_n are examples of **complex Lie groups**, meaning that each has the structure of a complex manifold on which the group multiplication and inversion maps are holomorphic.

Next we define the **unitary group** $\mathrm{U}_n = \mathrm{U}(n) = \{g \in \mathrm{GL}_n \mid g^* I g = I\}$, where $g^* = \overline{g}^T$ denotes the conjugate transpose. Hence elements of U_n preserve the standard Hermitian form on $\mathbb{C}^n$ given by $(v, w) = v * w$.

Now let I_p (respectively I_q) denote the $p \times p$ (respectively $q \times q$) identity matrix. Writing $I_{pq} = \begin{bmatrix} I_p & 0 \\ 0 & -I_q \end{bmatrix}$, we can define the **indefinite unitary group** of signature (p, q):

$$\mathrm{U}(p,q) = \left\{ g \in \mathrm{GL}_{p+q} \;\middle|\; g^*(I_{pq})g = I_{pq} \right\}.$$

More intuitively, elements of $\mathrm{U}(p,q)$ are linear transformations preserving the Hermitian

form on $\mathbb{C}^{p+q}$ given by $(v,w) = v^*(I_{pq})w$. Even more explicitly, we have

$$\mathrm{U}(p,q) = \left\{ g = \left[\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right] \in \mathrm{GL}_{p+q} \;\middle|\; g^{-1} = \left[\begin{smallmatrix} A^* & -C^* \\ -B^* & D^* \end{smallmatrix}\right] \right\}. \tag{4.1}$$

Finally, the **special indefinite unitary group** is $\mathrm{SU}(p,q) = \mathrm{U}(p,q) \cap \mathrm{SL}_{p+q}$.

The groups U_n, $\mathrm{U}(p,q)$, and $\mathrm{SU}(p,q)$ are examples of **real Lie groups**, meaning that each has the structure of a (real) differentiable manifold on which the group multiplication and inversion maps are differentiable. As a topological space, U_n is compact, while $\mathrm{SU}(p,q)$ is noncompact; nevertheless, $\mathrm{SU}(p,q)$ contains a maximal compact subgroup $\mathrm{S}(\mathrm{U}_p \times \mathrm{U}_q)$, where the "S" denotes the unit-determinant condition, realized block-diagonally as follows:

$$\mathrm{S}(\mathrm{U}_p \times \mathrm{U}_q) = \left\{ \left[\begin{smallmatrix} A & 0 \\ 0 & B \end{smallmatrix}\right] \;\middle|\; A \in \mathrm{U}_p,\ B \in \mathrm{U}_q,\ \det A \det B = 1 \right\}. \tag{4.2}$$

In general, given a real Lie group G_0, the maximal compact subgroup is typically denoted by K_0.

Example 4.1. Let $p = q = 1$. Then we see from (4.1) that

$$\mathrm{SU}(1,1) = \left\{ \left[\begin{smallmatrix} a & b \\ \bar{b} & \bar{a} \end{smallmatrix}\right] \;\middle|\; a, b \in \mathbb{C} \text{ with } |a|^2 - |b|^2 = 1 \right\}.$$

The group $G_0 = \mathrm{SU}(1,1)$ is the isometry group of the Poincaré disk $\mathbb{D} = \left\{ z \in \mathbb{C} \;\middle|\; |z| < 1 \right\}$, acting transitively by Möbius transformations. Moreover,

$$K_0 = \mathrm{S}(\mathrm{U}_1 \times \mathrm{U}_1) = \left\{ \left[\begin{smallmatrix} a & 0 \\ 0 & \bar{a} \end{smallmatrix}\right] \;\middle|\; |a| = 1 \right\} \cong \mathrm{U}_1$$

is the stabilizer of the point 0, so we have $G_0/K_0 \cong \mathbb{D}$. In general, G_0/K_0 is called a **symmetric space**.

Given a Lie group G, its **Lie algebra** $\mathrm{Lie}(G)$ is the vector space of left-invariant vector fields on G, equipped with the commutator bracket. The Lie algebra of a real Lie group is a real vector space, while that of a complex Lie group is complex. Since a left-invariant vector field is determined by its value at the identity element of G, we can identify the Lie algebra with the tangent space at the identity.

We briefly recall the Lie algebras of the groups above. In each case, the Lie algebra is a space of matrices equipped with the commutator bracket $[X, Y] = XY - YX$.

- $\mathrm{Lie}(\mathrm{GL}_n) = \mathfrak{gl}_n = \{X \in \mathrm{M}_n\}$.
- $\mathrm{Lie}(\mathrm{SL}_n) = \mathfrak{sl}_n = \{X \in \mathrm{M}_n \mid \mathrm{tr}(X) = 0\}$.
- $\mathrm{Lie}(\mathrm{U}_n) = \mathfrak{u}_n = \{X \in \mathfrak{gl}_n \mid X^* + X = 0\}$.
- $\mathrm{Lie}(\mathrm{U}(p,q)) = \mathfrak{u}(p,q) = \left\{ \left[\begin{smallmatrix} A & B \\ B^* & D \end{smallmatrix}\right] \in \mathfrak{sl}_{p+q} \;\middle|\; A \in \mathfrak{u}_p,\ D \in \mathfrak{u}_q \right\}$.
- $\mathrm{Lie}(\mathrm{SU}(p,q)) = \mathfrak{su}(p,q) = \mathfrak{u}(p,q) \cap \mathfrak{sl}_{p+q}$.
- $\mathrm{Lie}(\mathrm{S}(\mathrm{U}_p \times \mathrm{U}_q)) = \mathfrak{s}(\mathfrak{u}_p \oplus \mathfrak{u}_q) = \left\{ \left[\begin{smallmatrix} A & 0 \\ 0 & B \end{smallmatrix}\right] \;\middle|\; A \in \mathfrak{u}_p,\ B \in \mathfrak{u}_q,\ \mathrm{tr}(A) + \mathrm{tr}(B) = 0 \right\}$.

The Lie algebras $\mathfrak{gl}_n$ and $\mathfrak{sl}_n$ are **complex** (i.e., as vector spaces), whereas the other four are **real**. Since we will treat complex representations of these Lie algebras, it is most convenient to deal with their complexified versions. Specifically, given a real Lie algebra $\mathfrak{g}_0$, its **complexification** is defined to be $\mathfrak{g}_0^{\mathbb{C}} = \mathbb{C} \otimes_{\mathbb{R}} \mathfrak{g}_0 \cong \mathfrak{g}_0 \oplus i\mathfrak{g}_0$. We then have the following relationships:

- $\mathfrak{u}_n^{\mathbb{C}} = \mathfrak{gl}_n$.
- $\mathfrak{u}(p,q)^{\mathbb{C}} = \mathfrak{gl}_{p+q}$.
- $\mathfrak{su}(p,q)^{\mathbb{C}} = \mathfrak{sl}_{p+q}$.
- $\mathfrak{s}(\mathfrak{u}_p \oplus \mathfrak{u}_q)^{\mathbb{C}} = \mathfrak{s}(\mathfrak{gl}_p \oplus \mathfrak{gl}_q)$, embedded block-diagonally as above.

If $\mathrm{Lie}(G_0)^{\mathbb{C}} = \mathrm{Lie}(G)$, then we also say that the complex group G is the complexification of the real group G_0.

4.2 Preliminaries in representation theory

This section collects the notions from representation theory which we will need in the rest of the thesis; readers familiar with the subject can proceed directly to Section 4.3. The material in this section can be found in any standard reference on representation theory, such as [GW09], [FH91], or [Hum72], among the scores of others.

4.2.1 Roots and weights

Let G be a simple linear algebraic group over $\mathbb{C}$, i.e., a subgroup of GL_n defined by polynomial equations. Let $H \cong (\mathbb{C}^\times)^r \subset G$ be a maximal algebraic torus; we call r the **rank** of G or of its Lie algebra $\mathfrak{g}$. Let B be a choice of a Borel subgroup (i.e., a maximal connected solvable algebraic subgroup), with $T \subset B \subset G$. In turn, we let $\mathfrak{g}$, $\mathfrak{h}$, and $\mathfrak{b}$ denote the Lie algebras of G, H, and B, respectively; this makes $\mathfrak{h}$ a choice of a **Cartan subalgebra** of $\mathfrak{g}$.

Example 4.2. Let $G = \mathrm{SL}_n$, defined by the polynomial equation $\det = 1$. In this case, we choose H to be the subgroup of unit-determinant diagonal matrices, so that $r = n - 1$. We choose B to be the subgroup of upper-triangular matrices. Then $\mathfrak{h} \subset \mathfrak{g}$ is the subalgebra of trace-zero diagonal matrices, and $\mathfrak{b}$ the subalgebra of trace-zero upper-triangular matrices.

Let $\mathfrak{g}$ be a semisimple complex Lie algebra (i.e., a direct sum of simple Lie algebras). A **weight** is an element of $\mathfrak{h}^*$, i.e., a linear functional on the Cartan subalgebra $\mathfrak{h}$. Defining $\mathrm{ad} : \mathfrak{g} \longrightarrow \mathrm{End}(\mathfrak{g})$ by $(\mathrm{ad}\, x)(y) = [x, y]$ for $x, y \in \mathfrak{g}$, we have the standard fact that the operators in $\mathrm{ad}(\mathfrak{h})$ are simultaneously diagonalizable, and hence there is a direct sum decomposition $\mathfrak{g} = \bigoplus_{\alpha \in \mathfrak{h}^*} \mathfrak{g}_\alpha$, where $\mathfrak{g}_\alpha = \{x \in \mathfrak{g} \mid [h, x] = \alpha(h)x \text{ for all } h \in \mathfrak{h}\}$. The space $\mathfrak{g}_\alpha$ is nonzero for only finitely many $\alpha \in \mathfrak{h}^*$. If $\mathfrak{g}_\alpha \neq \{0\}$ and $\alpha \neq 0$, then α is called a **root**, and $\mathfrak{g}_\alpha$ the corresponding **root space**. Each root space is one-dimensional (since $\mathfrak{h}$ is abelian). We denote by $\Phi := \Phi(\mathfrak{g}, \mathfrak{h})$ the set of roots (or the **root system**) for the pair

$(\mathfrak{g}, \mathfrak{h})$. (It is a fundamental result, due mainly to the work of Wilhelm Killing and Elie Cartan, that the root system Φ completely characterizes the semisimple Lie algebra $\mathfrak{g}$.) Since $\mathfrak{h} = \mathfrak{g}_0$, we obtain $\mathfrak{g} = \mathfrak{h} \oplus \bigoplus_{\alpha \in \Phi} \mathfrak{g}_\alpha$. Furthermore, the choice of B induces a set partition $\Phi = \Phi^+ \sqcup \Phi^-$ such that $\bigoplus_{\alpha \in \Phi^+} \mathfrak{g}_\alpha \subset \mathfrak{b}$, where $\Phi^- = -\Phi^+$. We call Φ^+ and Φ^- the sets of **positive** and **negative** roots, respectively. A set of **simple roots** is a subset $\Pi = \{\alpha_1, \ldots, \alpha_r\} \subset \Phi^+$ such that every positive root can be written as a unique $\mathbb{N}$-linear combination of the α_i. Then Π is in fact a basis for $\mathfrak{h}^*$. We denote by Λ_R the **root lattice** $\mathbb{Z}\Pi$ generated additively by the simple roots.

For each $\alpha \in \Phi^+$, there exist unique $e_\alpha \in \mathfrak{g}_\alpha$ and $f_\alpha \in \mathfrak{g}_{-\alpha}$ such that $h_\alpha := [e_\alpha, f_\alpha]$ satisfies the equation $\alpha(h_\alpha) = 2$; this h_α is called the **coroot** of α. The set $\{h_{\alpha_1}, \ldots, h_{\alpha_{n-1}}\}$ is a basis for $\mathfrak{h}$, and its dual basis for $\mathfrak{h}^*$ is the set $\{\varpi_1, \ldots, \varpi_{n-1}\}$ of **fundamental weights**: hence $\varpi_i(h_{\alpha_j}) = \delta_{ij}$. Finally, we denote by $P(\mathfrak{g})$ the set of **integral weights**, namely, $\mathbb{Z}$-linear combinations of the fundamental weights ϖ_i; likewise, $P_+(\mathfrak{g})$ denotes the subset of **dominant integral weights**, namely, $\mathbb{N}$-linear combinations of the ϖ_i. (In the literature, $P(\mathfrak{g})$ and $P_+(\mathfrak{g})$ are often denoted by Λ and Λ^+ as well.)

Example 4.3. We continue with the example $\mathfrak{g} = \mathfrak{sl}_n$, in which Φ is known as the **Type-A** root system, or more specifically $\Phi = A_{n-1}$ (since we have already seen that $r = n - 1$):

- We have $\mathfrak{h} \cong \{h \in \mathbb{C}^n \mid h_1 + \cdots + h_n = 0\}$, via $\operatorname{diag}(h_1, \ldots, h_n) \mapsto h$.
- We define the weight $\varepsilon_i : \operatorname{diag}(h_1, \ldots, h_n) \longmapsto h_i$, for $i = 1, \ldots, n$.
- We thus write elements of $\mathfrak{h}^*$ as n-tuples in the standard (or "epsilon") coordinates. Hence $(a_1, \ldots, a_n) = a_1\varepsilon_1 + \cdots + a_n\varepsilon_n$. Note that $\varepsilon_1 + \cdots + \varepsilon_n = 0 \in \mathfrak{h}^*$, and so elements of $\mathfrak{h}^*$ are actually equivalence classes of n-tuples. For example, $(1, 1, 0) = \left(\frac{1}{3}, \frac{1}{3}, -\frac{2}{3}\right)$.
- We have the positive roots $\Phi^+ = \{\varepsilon_i - \varepsilon_j \mid 1 \leqslant i < j \leqslant n\}$.

- We have the root space $\mathfrak{g}_{\varepsilon_i - \varepsilon_j} = \mathbb{C}E_{ij}$, where E_{ij} is the matrix with 1 in the (i,j) position and 0's elsewhere.
- We have the simple roots $\Pi = \{\alpha_1, \ldots, \alpha_{n-1}\}$, where $\alpha_i = \varepsilon_i - \varepsilon_{i+1}$.
- We have the coroots $h_{\alpha_i} = E_{ii} - E_{i+1,i+1}$, for $i = 1, \ldots, n-1$.
- We have the fundamental weights $\varpi_i = \sum_{k=1}^{i} \varepsilon_k = (\underbrace{1, \ldots, 1}_{i}, 0, \ldots, 0) \in \mathfrak{h}^*$.

Example 4.4. The root system $\Phi = A_{n-1}$ is associated with $\mathfrak{gl}_n$ as well as $\mathfrak{sl}_n$. Although $\mathfrak{gl}_n$ is not semisimple (having a nontrivial center, namely the ideal of scalar matrices), nonetheless it will be convenient for us (in lining up histograms with weights) to consider $P(\mathfrak{gl}_n)$ rather than $P(\mathfrak{sl}_n)$. This is because for $\mathfrak{gl}_n$ we do not have the relation $\varepsilon_1 + \cdots + \varepsilon_n = 0$, resulting in the following natural descriptions for weights:

- We have integral weights $P(\mathfrak{gl}_n) = \{a_1\varepsilon_1 + \cdots + a_n\varepsilon_n \mid a_i \in \mathbb{Z}\} \cong \mathbb{Z}^n$.
- We have dominant integral weights $P_+(\mathfrak{gl}_n) = \{a_1\varepsilon_1 + \cdots + a_n\varepsilon_n \mid a_1 \geqslant \cdots \geqslant a_n \in \mathbb{Z}\}$.

By the term "Type A," therefore, we refer to any of the following: the groups SL_n or GL_n, the Lie algebras $\mathfrak{sl}_n$ or $\mathfrak{gl}_n$, or the real forms SU_n, U_n, $\mathrm{SU}(p,q)$, $\mathrm{U}(p,q)$ and their Lie algebras. In the context of the EMD we will work almost exclusively in the Type-A root system, but in Sections 5.3–5.4 we will have reason to consider Types B and C, associated with the special orthogonal Lie algebra $\mathfrak{so}(2n+1,\mathbb{C})$ and the symplectic Lie algebra $\mathfrak{sp}(2n,\mathbb{C})$ respectively.

4.2.2 Representations

A **representation** of a group G on a complex vector space V is a homomorphism $\pi : G \longrightarrow \mathrm{Aut}(V)$. When G is a Lie group, a finite-dimensional representation π induces a Lie algebra homomorphism $\mathrm{d}\pi : \mathrm{Lie}(G) \longrightarrow \mathrm{End}(V)$, which is also called a **representation** of $\mathrm{Lie}(G)$ on V. Letting $\mathfrak{g} = \mathrm{Lie}(G)$, we then say that G (or $\mathfrak{g}$) "acts on" V, which

we therefore call a G-module (or a $\mathfrak{g}$-module); furthermore, we abbreviate $\pi(g)(v)$ and $\mathrm{d}\pi(x)(v)$ by the simpler notation $g \cdot v$ and $x \cdot v$, respectively. We also often speak of V itself as a representation of G or $\mathfrak{g}$, when the action via π is understood. We also sometimes say that V is a **model** for the representation π. A G-module (or $\mathfrak{g}$-module) V is called **irreducible** if $\{0\}$ and V are the only subspaces that are stable under the group (or Lie algebra) action. Two representations (π, V) and (π', V') are called **equivalent** if there is a linear isomorphism $V \longrightarrow V'$ that commutes with the G-action.

For the following discussion, it is most natural to work at the level of the Lie algebra, and then extend the results to the group level. Let V be a representation of $\mathfrak{g}$. Then for $\lambda \in \mathfrak{h}^*$, we define

$$V_\lambda := \left\{ v \in V \;\middle|\; h \cdot v = \lambda(h)v \text{ for all } h \in \mathfrak{h} \right\}.$$

If $V_\lambda \neq \{0\}$, then we say that λ is a weight of V, and we call V_λ the **weight space of** V **with weight** λ. When V is finite-dimensional, we have a weight space decomposition $V = \bigoplus_{\lambda \in \mathfrak{h}^*} V_\lambda$.

Remark. Our use of the letters λ and μ is no accident — one of the themes of this chapter (and even more so of Chapter 5) is viewing histograms in $\mathcal{H}(m, n)$ as elements of $P(\mathfrak{gl}_n)$.

The set of integral weights admits a partial order as follows: for $\lambda, \mu \in P(\mathfrak{g})$, we have $\lambda \geq \mu$ if and only if $\lambda - \mu$ is an $\mathbb{N}$-linear combination of positive roots (equivalently, of simple roots). We say that ξ is the **highest weight** of a $\mathfrak{g}$-module V if $\xi > \lambda$ for any other weights λ of V. On the other hand, a **highest-weight representation** is defined to be a $\mathfrak{g}$-module that is generated by a weight vector $v \neq 0$, such that v is killed by the action of all positive root spaces in $\mathfrak{g}$. These two seemingly disparate notions are tied together by the theorem of the highest weight:

- For any semisimple Lie algebra $\mathfrak{g}$, and $\xi \in P_+(\mathfrak{g})$, there is a unique (up to equivalence) irreducible finite-dimensional representation of $\mathfrak{g}$ with highest weight ξ.

- Conversely, every irreducible finite-dimensional representation of $\mathfrak{g}$ is a highest-weight representation with highest weight in $P_+(\mathfrak{g})$.

Therefore, the irreducible finite-dimensional representations of $\mathfrak{g}$ are indexed (up to equivalence) by their highest weights, i.e., by the set $P_+(\mathfrak{g})$.

Now consider the ring $\mathbb{Z}[P(\mathfrak{g})]$ of formal $\mathbb{Z}$-linear combinations of integral weights:

$$\mathbb{Z}[P(\mathfrak{g})] = \left\{\text{finite sums } \sum_{\lambda \in P(\mathfrak{g})} m_\lambda e^\lambda \;\middle|\; m_\lambda \in \mathbb{Z}\right\}.$$

Note that we write a weight λ using the formal symbol e^λ; this is (in part) to avoid confusion between the natural additive structure of $P(\mathfrak{g})$ and that of $\mathbb{Z}[P(\mathfrak{g})]$. For a finite-dimensional representation V of $\mathfrak{g}$, we define the **character** as follows:

$$\operatorname{ch}_{\mathfrak{g}} V := \sum_{\lambda \in P(\mathfrak{g})} (\dim V_\lambda)\, e^\lambda,$$

where $\dim V_\lambda$ is also called the **multiplicity** of the weight λ in V. (We often suppress the subscript $\mathfrak{g}$ from $\operatorname{ch}_{\mathfrak{g}}$ if the Lie algebra action is understood from context.) The exponential notation is especially motivated by the way the ring operations mirror the direct sum and tensor product of representations:

$$\operatorname{ch}(V \oplus W) = \operatorname{ch} V + \operatorname{ch} W, \qquad \operatorname{ch}(V \otimes W) = \operatorname{ch} V \operatorname{ch} W.$$

4.2.3 Important examples in Type A

Example 4.5. Let $\mathfrak{g} = \mathfrak{sl}_2$. Then $\mathfrak{g}$ has just one fundamental weight $\varpi_1 = \varepsilon_1$, and so $P_+(\mathfrak{g}) = \{a\varepsilon_i \mid a \in \mathbb{N}\}$. Hence there is one irreducible $\mathfrak{g}$-module for each natural number. As it turns out, the representation with highest weight $n-1 := (n-1)\varepsilon_1$ is a vector space of dimension n, which decomposes into one-dimensional weight spaces $V_{n-1} \oplus V_{n-3} \oplus \cdots \oplus V_{3-n} \oplus V_{1-n}$. There is a combinatorial device to express the character of this irreducible $\mathfrak{sl}_2$-module, namely another guise of the q-binomial coefficients we encountered in Section 3.6.3: specifically, we define $(n)_q := q^{n-1} + q^{n-3} + \cdots + q^{3-n} + q^{1-n}$. Then by

identifying q with e^{ε_1}, we see that $(n)_q$ equals the character of the n-dimensional irreducible representation of $\mathfrak{sl}_2$. (We develop these q-binomial coefficients further in Example 4.10.)

Example 4.6. Let $\mathfrak{g} = \mathfrak{gl}_n$. Then we write F_n^ξ to denote the irreducible representation of $\mathfrak{gl}_n$ with highest weight ξ, where ξ is a weakly-decreasing n-tuple of integers. In general, F_n^ξ is a **rational representation** of the group GL_n, and if the ξ_i are all nonnegative (i.e., if ξ is a partition), then F_n^ξ is a **polynomial representation** of GL_n. ("Rational" and "polynomial" refer to the matrix entry functions in the image of the representation, in terms of the coordinate functions on GL_n.) Note that any weakly decreasing n-tuple ξ can be written as a pair $[\xi^+, \xi^-]$ of partitions ξ^+ and ξ^-, of lengths a and b respectively (where $a + b \leqslant n$), by setting

$$\xi = [\xi^+, \xi^-] := (\xi_1^+, \ldots, \xi_a^+, 0, \ldots, 0, -\xi_b^-, \ldots, -\xi_1^-). \tag{4.3}$$

For example, if $\xi = (5, 2, 0, 0, -1, -3, -4)$, then $\xi^+ = (5, 2)$ and $\xi^- = (4, 3, 1)$.

Example 4.7. Again let $G = \mathrm{GL}_n$ and $\mathfrak{g} = \mathfrak{gl}_n$. The **trivial representation** is the one-dimensional space $\mathbb{C}$ where $g \cdot v = v$ (respectively, $x \cdot v = 0$) for $g \in G$ (respectively, for $x \in \mathfrak{g}$), and all $v \in \mathbb{C}$. Clearly the highest weight is $\mathbf{0} = (0, \ldots, 0)$. We therefore denote the trivial representation either as $\mathbb{1}$ or as $F_n^{\mathbf{0}}$. Moreover, we have $\operatorname{ch} \mathbb{1} = e^{\mathbf{0}} = 1$.

Example 4.8. Again $G = \mathrm{GL}_n$ and $\mathfrak{g} = \mathfrak{gl}_n$. Then $\mathbb{C}^n$ is the **defining representation**, on which G and $\mathfrak{g}$ each act by matrix multiplication. Note that the standard basis vector $e_i \in \mathbb{C}^n$ is a vector of weight ε_i, and hence the highest weight is ε_1. Therefore $\mathbb{C}^n = F_n^{(1,0,\ldots,0)}$, and $\operatorname{ch} \mathbb{C}^n = \sum_{i=1}^n e^{\varepsilon_i}$.

Example 4.9. This example will be central in Sections 4.6, 4.7, and 5.2, each for different reasons. Let $(m) := (m, 0, \ldots, 0) \in P_+(\mathfrak{gl}_n)$. Then $F_n^{(m)} \cong S^m(\mathbb{C}^n)$, the mth symmetric power of the defining representation. Moreover, the character of $F_n^{(m)}$ is the **complete homogeneous symmetric polynomial of degree m in n variables**:

$$\operatorname{ch} F_n^{(m)} = h_m(\mathbf{x}) := \sum_{\lambda_1 + \cdots + \lambda_n = m} x_1^{\lambda_1} \cdots x_n^{\lambda_n}, \tag{4.4}$$

where $x_i = e^{\varepsilon_i}$. From this, we can see that $F_n^{(m)} \cong S^m(\mathbb{C}^n)$ has a basis of distinct weight vectors, in natural bijective correspondence with the set $\mathcal{H}(m,n)$ of histograms. In this way, $F_n^{(m)}$ is the representation-theoretic analogue of $\mathcal{H}(m,n)$, a fact that underlies much of the next two chapters.

Example 4.10. As a final example, we return to $\mathfrak{g} = \mathfrak{sl}_2$. Let $F_2^{n-1} \cong \mathbb{C}^n$ denote the n-dimensional irreducible representation of $\mathfrak{g}$ from Example 4.5. Now we consider $S^m(F_2^{n-1})$, on which $\mathfrak{sl}_2$ also acts irreducibly, and we seek to write down its character. Recalling the second guise of q-binomial coefficients from Example 4.5, we can write $(a)_q! := (a)_q \cdots (1)_q$. Then we define

$$\binom{a}{b}_q := \frac{(a)_q!}{(b)_q!(a-b)_q!}.$$

This is a Laurent polynomial, invariant under $q \longleftrightarrow q^{-1}$. More to the point, we have the following identity:

$$\mathrm{ch}_{\mathfrak{sl}_2} S^m(F_2^{n-1}) = \binom{m+n-1}{m}_q, \tag{4.5}$$

where again $q = e^{\varepsilon_1}$. (See Theorem 4.1.20 in [GW09].) In Chapter 6 we will need the following conversion between the two guises of q-binomial coefficients:

$$\binom{m+n-1}{m}_q = q^{-m(n-1)} \begin{bmatrix} m+n-1 \\ m \end{bmatrix}_{q^2}. \tag{4.6}$$

4.3 The first Wallach representation of $\mathfrak{su}(p,q)$

The purpose of this section is to explain the representation-theoretic interest in a certain infinite-dimensional module that naturally models histogram pairs. The discussion is fairly technical, but the reader may skip to the next section without serious loss of continuity.

Real form	Complexification
$G_0 = \mathrm{SU}(p,q)$	$G = \mathrm{SL}_{p+q}$
$\mathfrak{g}_0 = \mathfrak{su}(p,q)$	$\mathfrak{g} = \mathfrak{sl}_{p+q}$
$K_0 = \mathrm{S}(\mathrm{U}_p \times \mathrm{U}_q)$	$K = \mathrm{S}(\mathrm{GL}_p \times \mathrm{GL}_q)$
$\mathfrak{k}_0 = \mathfrak{s}(\mathfrak{u}_p \oplus \mathfrak{u}_q)$	$\mathfrak{k} = \mathfrak{s}(\mathfrak{gl}_p \oplus \mathfrak{gl}_q)$

Table 3: Notation for the groups and Lie algebras associated with Hermitian symmetric pairs in Type A.

4.3.1 Hermitian symmetric pairs: Type A

Throughout this section, we appeal to the general theory of Hermitian symmetric pairs. (See [EHP14] for a thorough exposition of the theory, which we adopt as our perspective here.) Given a real simple noncompact group G_0 and subgroup K_0, the pair (G_0, K_0) is called an **irreducible Hermitian symmetric pair of noncompact type** if G_0/K_0 is a noncompact symmetric space admitting a complex structure on which G_0 acts by biholomorphic transformations. There are seven families of such pairs; in this chapter we will single out the Type-A case, where $G_0 = \mathrm{SU}(p,q)$, with $K_0 = \mathrm{S}(\mathrm{U}_p \times \mathrm{U}_q) \subset G_0$ embedded block-diagonally as in (4.2). In Example 4.1, where $p = q = 1$, we already observed that G_0/K_0 was a noncompact symmetric space (namely the Poincaré disk $\mathbb{D}$) admitting the requisite G_0-action. In Table 3, we recall the notation for the various groups and Lie algebras associated with Hermitian symmetric pairs in Type A.

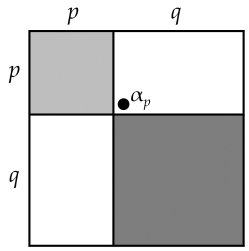


Figure 6: The embedding $\mathfrak{k} = \mathfrak{s}(\mathfrak{gl}_p \oplus \mathfrak{gl}_q) \subset \mathfrak{g} = \mathfrak{sl}_{p+q}$, with unique noncompact simple root α_p.

Let $\mathfrak{h} \subset \mathfrak{k} \subset \mathfrak{g}$ be the usual choice of Cartan subalgebra consisting of diagonal matrices.

We observe from Figure 6 the Type-A Cartan decomposition

$$\begin{array}{rcccccc} \mathfrak{g} & = & \mathfrak{k} & \oplus & \mathfrak{p}^+ & \oplus & \mathfrak{p}^- \\ \mathfrak{sl}_{p+q} & = & \mathfrak{s}(\mathfrak{gl}_p \oplus \mathfrak{gl}_q) & \oplus & \mathrm{M}_{p,q} & \oplus & \mathrm{M}_{q,p} \end{array} \tag{4.7}$$

where $\mathfrak{p}^+$ and $\mathfrak{p}^-$ are the upper-right and lower-left block-embeddings, respectively. For each root $\alpha \in \Phi$, we call α a **compact root** if $\mathfrak{g}_\alpha \subset \mathfrak{k}$, and a **noncompact root** if $\mathfrak{g}_\alpha \subset \mathfrak{p}$. In Type A, since $\mathfrak{k}$ embeds block-diagonally, we see from Figure 6 that α_p is the unique noncompact simple root. (In fact, the existence of a unique noncompact simple root can be taken as an equivalent definition of a Hermitian symmetric pair.) We see further that the compact roots in Type A are

$$\left\{\pm(\varepsilon_i - \varepsilon_j) \mid 1 \leqslant i < j \leqslant p \text{ or } p < i < j \leqslant p+q\right\}. \tag{4.8}$$

4.3.2 Definition of the Wallach representations

The exposition below closely follows that in [Wal79, EW04, EH04]. The distinguished weight ζ is defined to be the fundamental weight of $\mathfrak{g}$ that is orthogonal to all compact roots. Inspecting (4.8), we see that in Type A,

$$\zeta = (\underbrace{1, \ldots, 1}_{p}, \underbrace{0, \ldots, 0}_{q}) = \varpi_p.$$

In the general theory, to each family of Hermitian symmetric pairs is associated a certain constant c; in Type A, we have $c = 1$. We let r denote the rank of the symmetric space G_0/K_0; in Type A, we have $r = \min\{p, q\}$.

For $\xi \in P_+(\mathfrak{k})$, we can define the **generalized** (or **parabolic**) **Verma module**

$$\begin{aligned} M_\xi &:= U(\mathfrak{g}) \otimes_{U(\mathfrak{q})} V_\xi \\ &\cong S(\mathfrak{p}^-) \otimes V_\xi, \end{aligned} \tag{4.9}$$

where V_ξ is the irreducible finite-dimensional $\mathfrak{k}$-module with highest weight ξ. Then M_ξ is a $U(\mathfrak{g})$-module, and L_ξ is defined to be the unique irreducible quotient of M_ξ.

Definition 4.11. Let (G_0, K_0) be an irreducible Hermitian symmetric pair. For $1 \leqslant k < r$, the ***k*th Wallach representation** of $\mathfrak{g}_0$ is the $\mathfrak{g}$-module $W_k := L_{-kc\zeta}$. Specifically, in Type A, the kth Wallach representation of $\mathfrak{su}(p,q)$ is

$$W_k = L_{(-k,\ldots,-k,0,\ldots,0)}.$$

Remark. The Wallach representations are significant in that they were constructed in [Wal79] as analytic continuations of holomorphic discrete series representations; see the end of the introduction to this thesis for the analytical motivation behind the discrete series.

The complex group K acts on $\mathfrak{p}^+$ with finitely many orbits, whose closures form a chain of varieties

$$0 = Y_0 \subset Y_1 \subset \cdots \subset Y_r = \mathfrak{p}^+.$$

The key fact (see [DES91, Jos92, Wal79]) is that the ideal

$$I(Y_k) := \left\{ f \in \mathbb{C}[\mathfrak{p}^+] \;\middle|\; f(X) = 0 \text{ for all } X \in Y_k \right\},$$

when viewed as an $S(\mathfrak{p}^-)$-module, is the annihilator of W_k. It follows that as $\mathfrak{g}$-modules, we have $W_k \cong \mathbb{C}[Y_k]$. In Type A, the closed orbits Y_k are the **determinantal varieties**

$$Y_k = \mathrm{M}_{p,q}^{\leqslant k} := \left\{ X \in \mathrm{M}_{p,q} \mid \operatorname{rank} X \leqslant k \right\}.$$

We therefore have the following realization of the Wallach representations in Type A. (We will encounter a second in Section 4.6.)

Proposition 4.12. *As $\mathfrak{su}(p,q)$-modules, we have $W_k \cong \mathbb{C}[\mathrm{M}_{p,q}^{\leqslant k}]$.*

In the context of the EMD, we will restrict our attention to the $k = 1$ case; then the first Wallach representation W_1 of $\mathfrak{su}(p,q)$ is isomorphic to $\mathbb{C}[\mathrm{M}_{p,q}^{\leqslant 1}]$. Letting z_{ij} denote the natural coordinate functions on $\mathrm{M}_{p,q}$, we conclude that $I(\mathrm{M}_{p,q}^{\leqslant 1})$ is the **determinantal ideal** generated by all 2×2 minors:

$$I(\mathrm{M}_{p,q}^{\leqslant 1}) = \left\langle z_{ij} z_{IJ} - z_{Ij} z_{iJ} \right\rangle_{i<I,\, j<J.}$$

Therefore we now have

$$W_1 \cong \mathbb{C}[\mathrm{M}_{p,q}^{\leqslant 1}] = \mathbb{C}[z_{ij}] \Big/ \langle z_{ij}z_{IJ} - z_{Ij}z_{iJ} \rangle.$$

As a result, in $\mathbb{C}[\mathrm{M}_{p,q}^{\leqslant 1}]$, we can replace any products of the form $z_{Ij}z_{iJ}$ by $z_{ij}z_{IJ}$. Thus we say that $\mathbb{C}[\mathrm{M}_{p,q}^{\leqslant 1}]$ has a standard monomial theory, i.e., a monomial basis given by

$$\left\{ z_{i_1,j_1} \cdots z_{i_m,j_m} \;\middle|\; m \in \mathbb{N},\ (i_1,j_1) \preceq \cdots \preceq (i_m,j_m) \text{ w.r.t. product order on } [p] \times [q] \right\}. \quad (4.10)$$

Remark. For fixed m, the basis above should remind the reader of Proposition 2.5, which characterized the optimal flow matrix for the Monge problem as a matrix whose entries sum to m, with support in a chain. Indeed, the relations $z_{ij}z_{IJ} - z_{Ij}z_{iJ}$ cutting out the determinantal variety are a ring-theoretic analogue of "minimizing work" as in Section 2.3: just as an optimal flow matrix F can be constructed such that F_{Ij} and F_{iJ} are never both nonzero, so here a monomial in $\mathbb{C}[\mathrm{M}_{p,q}^{\leqslant 1}]$ can always be written such that the exponent of z_{Ij} and the exponent of z_{iJ} are never both nonzero. We will pursue this connection in Section 4.6.

4.4 Howe duality in Type A

We have just seen that $\mathbb{C}[\mathrm{M}_{p,q}^{\leqslant k}]$ affords a model for the kth Wallach representation W_k in Type A. In this section, we describe the Type-A Howe duality setting in classical invariant theory, which is the "natural habitat" of a *second* model for W_k. These two realizations of W_k will ultimately allow us (in Theorem 4.15) to recast the northwest corner rule from Section 2.3 in the language of representation theory. The main result in Chapter 7 will also depend heavily upon this section. We largely follow the exposition in [HTW05, Section 3].

Fix $k, p, q \in \mathbb{N}$, and let

$$V = \mathrm{M}_{k,p} \oplus \mathrm{M}_{k,q}\,. \quad (4.11)$$

(For reasons that will become clear, we purposely reuse the "k" that parametrized the kth Wallach representation in the previous section.) The pair $(\mathrm{U}_k, \mathrm{U}(p,q))$ is called a **compact**

dual pair, since the first element is a compact group, and since the two groups centralize each other as subgroups of the symplectic group $\mathrm{Sp}(2k(p+q), \mathbb{R})$. We are interested in the action of the dual pair $(\mathrm{U}_k, \mathrm{U}(p,q))$ on the coordinate ring $\mathbb{C}[V]$. It will be convenient to consider instead the complexification GL_k of the first group, and the complexified Lie algebra $\mathfrak{gl}_{p+q}$ of the second group. This will allow us to avoid technicalities involving a covering group of $\mathrm{U}(p,q)$. Hence, we consider the pair $(\mathrm{GL}_k, \mathfrak{gl}_{p+q})$ and its action on $\mathbb{C}[V]$. We will denote an element of V by an ordered pair (X, Y), and when we have need of coordinates, we will let x_{ij} denote the natural coordinate functions on $\mathrm{M}_{k,p}$, and y_{ij} those on $\mathrm{M}_{k,q}$.

The group GL_k acts by matrix multiplication on the second summand $\mathrm{M}_{k,q}$ of V, and by the inverse transpose (denoted by the superscript $-T$) on the first summand $\mathrm{M}_{k,p}$:

$$g \cdot (X, Y) = (g^{-T} X, gY). \tag{4.12}$$

In classical invariant theory, this is the natural action obtained by regarding $\mathrm{M}_{k,q}$ as the sum of q copies of the defining representation $\mathbb{C}^k$, and $\mathrm{M}_{k,p}$ as the sum of p copies of the dual representation $(\mathbb{C}^k)^*$. The action (4.12) induces the usual action on $\mathbb{C}[V]$ via

$$g \cdot f(X, Y) = f(g^T X, g^{-1} Y). \tag{4.13}$$

This classical action of GL_k on $\mathbb{C}[V]$ extends to the Weyl algebra $\mathbb{D}(V)$ of polynomial-coefficient differential operators on $\mathbb{C}[V]$; in particular, GL_k normalizes $\mathbb{D}(V)$ inside $\mathrm{End}(\mathbb{C}[V])$. Let $\mathbb{D}(V)^{\mathrm{GL}_k}$ be the GL_k-invariant subalgebra, i.e., the subalgebra containing those operators that commute with the GL_k-action on $\mathbb{C}[V]$. Then $\mathbb{D}(V)^{\mathrm{GL}_k}$ is generated (as an algebra) by a subset which is isomorphic to $\mathfrak{gl}_{p+q}$ as a Lie algebra.

For future reference, we now explicitly describe this $\mathfrak{gl}_{p+q}$-action on $\mathbb{C}[V]$. (This was written out in [CEW22, Section 2].) Let x_{ij} and y_{ij} denote the coordinate functions on a pair of matrices $(X, Y) \in V$, and consider a $(p+q) \times (p+q)$ matrix D of differential

operators

$$D = \left[\begin{array}{c|c} -\mathfrak{k}_x(j,i) & \Delta(i,j) \\ \hline -h(j,i) & \mathfrak{k}_y(i,j) \end{array}\right], \tag{4.14}$$

written in 2×2 block form, where the indexing begins at $(1,1)$ within each individual block, and where

- $\mathfrak{k}_x(i,j) = \sum_{\ell=1}^{k} x_{\ell i} \frac{\partial}{\partial x_{\ell j}} + k\delta_{ij}$ (shifted Euler operators), for $1 \leqslant i,j \leqslant p$;
- $\mathfrak{k}_y(i,j) = \sum_{\ell=1}^{k} y_{\ell i} \frac{\partial}{\partial y_{\ell j}}$ (Euler operators), for $1 \leqslant i,j \leqslant q$;
- $\Delta(i,j) = \sum_{\ell=1}^{k} \frac{\partial^2}{\partial x_{\ell i} \partial y_{\ell j}}$ (raising operators), for $1 \leqslant i \leqslant p$ and $1 \leqslant j \leqslant q$;
- $h(i,j) = \sum_{\ell=1}^{k} x_{\ell i} y_{\ell j}$ (lowering operators), for $1 \leqslant i \leqslant q$ and $1 \leqslant j \leqslant p$.

(For the original analytic picture, see [KV78, p. 33].) The $(p+q)^2$ operators in D generate $\mathbb{D}(V)^{\mathrm{GL}_k}$, and they form a basis for a Lie algebra isomorphic to $\mathfrak{gl}_{p+q}$ as follows. Let $E_{ij} \in \mathfrak{gl}_{p+q}$ denote the matrix with 1 in the (i,j) position and 0's elsewhere. Then the map $E_{ij} \mapsto D_{ij}$ extends to a homomorphism of Lie algebras $\mathfrak{gl}_{p+q} \to \mathbb{D}(V)^{\mathrm{GL}_k}$, and $\mathfrak{gl}_{p+q}$ acts on $\mathbb{C}[V]$ via its image in $\mathbb{D}(V)^{\mathrm{GL}_k}$.

The fact that $\mathfrak{gl}_{p+q}$ generates $\mathbb{D}(V)^{\mathrm{GL}_k}$ implies the following **Howe duality** decomposition as a module for $\mathrm{GL}_k \times \mathfrak{gl}_{p+q}$:

$$\mathbb{C}[V] \cong \bigoplus_{\xi} F_k^{\xi} \otimes \widetilde{F}_{p,q}^{\xi} \tag{4.15}$$

where

- ξ ranges over a certain subset of the highest weights (which we will clarify below) indexing the rational representations of GL_k.
- F_k^{ξ} is the rational representation of GL_k with highest weight ξ.
- $\widetilde{F}_{p,q}^{\xi}$ is an infinite-dimensional $\mathfrak{gl}_{p+q}$-module uniquely determined by ξ.

To begin to connect this picture with the Wallach representations from Section 4.3, we consider the restriction of the action of $\mathrm{U}(p,q)$ to $G_0 = \mathrm{SU}(p,q)$, and therefore of $\mathfrak{u}(p,q)$ to $\mathfrak{g}_0 = \mathfrak{su}(p,q)$, and of $\mathfrak{gl}_{p+q}$ to $\mathfrak{g} = \mathfrak{sl}_{p+q}$ (with notation as in Section 4.3). In this way, $\widetilde{F}^{\xi}_{p,q}$ is a $\mathfrak{g}$-module.

4.5 Classical invariant theory: FFT and SFT for GL_k

With $\mathbb{C}[V]$ as in the previous section, we denote the GL_k**-invariant subalgebra** by

$$\mathbb{C}[V]^{\mathrm{GL}_k} := \Big\{ f \in \mathbb{C}[V] \,\Big|\, g \cdot f(X,Y) = f(X,Y), \text{ for all } g \in \mathrm{GL}_k \text{ and } (X,Y) \in V \Big\}.$$

This invariant subalgebra has an especially nice description, thanks to Hermann Weyl's first and second fundamental theorems (FFT and SFT) of classical invariant theory [Wey39].

Theorem 4.13 (FFT for GL_k)**.** *The invariant subalgebra $\mathbb{C}[V]^{\mathrm{GL}_k}$ is generated by the quadratic functions*

$$r_{ij} := \sum_{\ell=1}^{k} x_{\ell i} y_{\ell j} \tag{4.16}$$

for $i \in [p]$ and $j \in [q]$.

We can easily see this invariance by recognizing r_{ij} as the dot product of the two column vectors $\mathbf{x}_i$ and $\mathbf{y}_j$: namely, $r_{ij} = (\mathbf{x}_i, \mathbf{y}_j) = \mathbf{x}_i^T \mathbf{y}_j$. Under the action of $g \in \mathrm{GL}_k$ in (4.13), we have $g \cdot r_{ij} = (g^T \mathbf{x}_i,\, g^{-1}\mathbf{y}_j) = (g^T \mathbf{x}_i)^T g^{-1} \mathbf{y}_j = \mathbf{x}_i^T g g^{-1} \mathbf{y}_j = \mathbf{x}_i^T \mathbf{y}_j = (\mathbf{x}_i, \mathbf{y}_j) = r_{ij}$, and so the invariance of the r_{ij} is clear. See [GW09, Theorem 5.2.1] for the full proof, including the fact that the r_{ij} indeed generate $\mathbb{C}[V]^{\mathrm{GL}_k}$. While the FFT describes the generators of $\mathbb{C}[V]^{\mathrm{GL}_k}$, the SFT describes the relations among them, which are precisely the determinants of all $(k+1) \times (k+1)$ minors, if we regard the r_{ij} as entries of a $p \times q$ matrix.

Theorem 4.14 (SFT for GL_k)**.** *Let z_{ij} be the natural coordinate functions on $\mathrm{M}_{p,q}$. We have an algebra isomorphism $\varphi : \mathbb{C}[V]^{\mathrm{GL}_k} \longrightarrow \mathbb{C}[\mathrm{M}_{p,q}^{\leqslant k}]$, given by $\varphi : r_{ij} \longmapsto z_{ij}$.*

(See [GW09, Theorem 5.2.2].) In light of Proposition 4.12, the isomorphism φ maps $\mathbb{C}[V]^{\mathrm{GL}_k}$ to a model for the kth Wallach representation W_k. But in fact, from the perspective

of the Howe duality decomposition (4.15), we can make the identification $\mathbb{C}[V]^{\mathrm{GL}_k} = \widetilde{F}^{\mathbf{0}}_{p,q}$, since the corresponding GL_k-module $F^{\mathbf{0}}_k$ is the trivial representation. Hence we can restate the SFT as an algebra isomorphism between *two* different models of W_k:

$$\begin{aligned} \varphi : \widetilde{F}^{\mathbf{0}}_{p,q} &\longrightarrow \mathbb{C}[\mathrm{M}^{\leqslant k}_{p,q}] \\ r_{ij} &\longmapsto z_{ij}. \end{aligned} \tag{4.17}$$

In the next section, we will see that when $k = 1$, this isomorphism φ is actually the representation-theoretic analogue of the northwest corner rule from Section 2.3, which started with a pair of histograms and output an optimal flow matrix F to solve the Monge problem and hence determine the EMD.

4.6 The first Wallach representation and the northwest corner rule

To uncover the connection to the EMD, we adopt the following specializations for the remainder of the chapter:

- We set $k = 1$; hence $\mathrm{GL}_k = \mathrm{GL}_1 = \mathbb{C}^\times$, so $\xi \in \mathbb{Z}$. In particular, we now write the zero weight for GL_1 as 0 rather than $\mathbf{0}$.
- We set $p = q$, and rename both as the single parameter n. (This is meant to echo the "n" that denoted the number of histogram bins in previous chapters.)

Having set the Howe duality parameters $(k, p, q) = (1, n, n)$, we now have

$$V = \mathrm{M}_{1,n} \oplus \mathrm{M}_{1,n}$$

and thus in $\mathbb{C}[V]$ we can suppress the first index in our variables:

$$\mathbb{C}[V] = \mathbb{C}[\mathbf{x}, \mathbf{y}] \coloneqq \mathbb{C}[x_1, \ldots, x_n, y_1, \ldots, y_n],$$

where the GL_1-action (4.13) on the coordinate functions is given by $g \cdot x_i = g x_i$ and $g \cdot y_i = g^{-1} y_i$, for $g \in \mathbb{C}^\times$. The definition of the invariant generators r_{ij} from (4.16) can thus be

simplified as

$$r_{ij} = x_i y_j, \qquad i, j \in [n]. \tag{4.18}$$

Crucially, we can now identify a pair of n-bin histograms with an exponent vector, and so we can write monomials in $\mathbb{C}[V]$ as follows:

$$\mathbf{x}^\lambda \mathbf{y}^\mu := x_1^{\lambda_1} \cdots x_n^{\lambda_n} y_1^{\mu_1} \cdots y_n^{\mu_n}.$$

In this way, we can decompose $\widetilde{F}^0_{n,n}$ into one-dimensional subspaces corresponding to histogram pairs of equal mass:[1]

$$\widetilde{F}^0_{n,n} = \mathbb{C}[x_i y_j] = \bigoplus_{m \in \mathbb{N}} \left(\bigoplus_{\lambda,\mu \in \mathcal{H}(m,n)} \mathbb{C}\mathbf{x}^\lambda \mathbf{y}^\mu \right).$$

We are now equipped to state the main result of this chapter, which reinterprets the SFT (Theorem 4.14) as an analogue of the northwest corner rule.

Theorem 4.15. *Set $k = 1$ in the Howe duality setting, let φ be the SFT map in* (4.17)*, and choose the monomial basis* (4.10) *for $\mathbb{C}[\mathrm{M}_n^{\leqslant 1}]$. Then for $\lambda, \mu \in \mathcal{H}(m,n)$, we have*

$$\varphi(\mathbf{x}^\lambda \mathbf{y}^\mu) = \prod_{i,j \in [n]} z_{ij}^{F_{ij}}, \tag{4.19}$$

where $F = [F_{ij}]$ is the output of the northwest corner rule given the input (λ, μ).

Proof. By (4.18), the monomial $\mathbf{x}^\lambda \mathbf{y}^\mu$ is in the domain of φ if and only if $\lambda, \mu \in \mathcal{H}(m,n)$ for some $m \in \mathbb{N}$; therefore the left-hand side of (4.19) is indeed defined. Now, we can rewrite as $\mathbf{x}^\lambda \mathbf{y}^\mu = r_{i_1,j_1} \cdots r_{i_m,j_m}$ such that $W_{\lambda,\mu} = \left[\begin{smallmatrix} i_1 & \cdots & i_m \\ j_1 & \cdots & j_m \end{smallmatrix}\right]$, where $W_{\lambda,\mu}$ is the matrix defined in Section 3.2. By the definition of φ in (4.17), we then have

$$\varphi(\mathbf{x}^\lambda \mathbf{y}^\mu) = \varphi(r_{i_1,j_1}) \cdots \varphi(r_{i_m,j_m}) = z_{i_1,j_1} \cdots z_{i_m,j_m}. \tag{4.20}$$

By the definition of $W_{\lambda,\mu}$, we have $(i_1, j_1) \preceq \cdots \preceq (i_m, j_m)$ in $[n] \times [n]$, and so $z_{i_1,j_1} \cdots z_{i_m,j_m}$ is already an element of the basis (4.10). Finally, by the RSK construction, the output of

[1] Of course, the notation $\mathbb{C}[x_i y_j]$ is not to be understood as the polynomial ring in the $x_i y_j$, due to the obvious syzygies $(x_i y_j)(x_I y_J) = (x_i y_J)(x_I y_j)$.

the northwest corner rule is the matrix F such that F_{ij} equals the number of occurrences of the column $\left[\begin{smallmatrix} i \\ j \end{smallmatrix}\right]$ in $W_{\lambda,\mu}$, which is the number of occurrences of z_{ij} in the right-hand side of (4.20). Hence (4.20) can be written as (4.19), which completes the proof. □

Next, in order to capture the EMD, we define a sort of q-analogue denoted by φ_q:

$$\varphi_q(-) := \varphi(-)\Big|_{z_{ij}=q^{|i-j|}}. \tag{4.21}$$

Corollary 4.16. *We have* $\varphi_q(\mathbf{x}^\lambda \mathbf{y}^\mu) = q^{\mathbf{EMD}(\lambda,\mu)}$.

Proof. Applying the evaluation $z_{ij} = q^{|i-j|}$ to the right-hand side of (4.19), we have

$$\prod_{i,j} q^{|i-j|\,F_{ij}} = q^{\sum_{i,j}|i-j|\,F_{ij}} = q^{\sum_{i,j} C_{ij}F_{ij}} = q^{\mathbf{EMD}(\lambda,\mu)}. \qquad \square$$

Example 4.17. We return to Example 3.2, in which $\lambda = (1,7,0,1,1)$ and $\mu = (4,0,1,3,2)$. (It is instructive to compare each step below to its analogue in Example 3.2.) We have $\varphi(\mathbf{x}^\lambda \mathbf{y}^\mu) = \varphi(x_1 x_2^7 x_4 x_5\, y_1^4 y_3 y_4^3 y_5^2) = \varphi(r_{11} r_{21}^3 r_{23} r_{24}^3 r_{45} r_{55}) = z_{11} z_{21}^3 z_{23} z_{24}^3 z_{45} z_{55}$. Upon making the substitutions $z_{ij} = q^{|i-j|}$, we obtain $\varphi_q(\mathbf{x}^\lambda \mathbf{y}^\mu) = q^0 (q^1)^3 q^1 (q^2)^3 q^1 q^0 = q^{11}$, which agrees with our earlier result $\mathbf{EMD}(\lambda,\mu) = 11$.

Recall the decomposition $\widetilde{F}^0_{n,n} = \bigoplus_{m\in\mathbb{N}} \left(\bigoplus_{\lambda,\mu\in\mathcal{H}(m,n)} \mathbb{C}\mathbf{x}^\lambda \mathbf{y}^\mu \right)$. Each of the one-dimensional subspaces in this decomposition is a weight space of weight (λ,μ) under the action of $K = \mathrm{S}(\mathrm{GL}_n \times \mathrm{GL}_n)$. Hence we have the following character as a $\mathfrak{k}$-module:

$$\mathrm{ch}_{\mathfrak{k}}\, \widetilde{F}^0_{n,n} = \sum_{m\in\mathbb{N}} h_m(\mathbf{x})\, h_m(\mathbf{y}),$$

where the h_m's are the polynomials defined in (4.4), and $x_i = e^{\varepsilon_i}$ in the first copy of $\mathfrak{gl}_n$, while $y_i = e^{\varepsilon_i}$ in the second copy of $\mathfrak{gl}_n$. This character implies the K-decomposition

$$\begin{aligned} \widetilde{F}^0_{n,n} &\cong \bigoplus_{m\in\mathbb{N}} F_n^{(m)} \otimes F_n^{(m)} \\ &= \bigoplus_m S^m(\mathbb{C}^n) \otimes S^m(\mathbb{C}^n). \end{aligned} \tag{4.22}$$

Each graded component of (4.22) is called a ***K*-type**. We observe (as in Example 4.9) that each of these K-types $S^m(\mathbb{C}^n) \otimes S^m(\mathbb{C}^n)$ is the representation-theoretic analogue of $\mathcal{H}(m,n) \times \mathcal{H}(m,n)$, since the former has a basis of distinct weight vectors, each of which corresponds naturally to a distinct ordered pair of histograms.

The complexified action of $\mathfrak{g}_0 = \mathfrak{su}(p,q)$ on the first Wallach representation can naturally be interpreted in terms of manipulating histogram pairs. Each K-type in $\widetilde{F}^0_{n,n}$ is invariant under the action of $\mathfrak{k} \subset \mathfrak{g}$, via the differential operators from (4.14), whose action on monomials $\mathbf{x}^\lambda \mathbf{y}^\mu$ can be interpreted in terms of histogram pairs:

- $\mathfrak{k}_x(i,j) = x_i \frac{\partial}{\partial x_j} + \delta_{ij}$ moves a data point in λ from bin j to bin i.
- $\mathfrak{k}_y(i,j) = y_i \frac{\partial}{\partial y_j}$ moves a data point in μ from bin j to bin i.

Likewise, the action of $\mathfrak{p} = \mathfrak{p}^+ \oplus \mathfrak{p}^- \subset \mathfrak{g}$ from (4.14) has a similar interpretation, although it no longer preserves each K-type:

- $\Delta(i,j) = \frac{\partial^2}{\partial x_i \partial y_j}$ removes a data point from bin i in λ and from bin j in μ.
- $h(i,j) = x_i y_j$ adds one data point to bin i in λ and to bin j in μ.

4.7 Partial matching via Howe duality

We are finally equipped to shed light on the case of histograms with unequal masses (i.e., *partial matching*). Recall that the original definition of the EMD, in terms of the Monge problem (2.4) – (2.8), assumes that the masses of λ and μ may indeed be different. This possibility is precisely what distinguishes the EMD from the 1-Wasserstein distance, as noted in [BL01]: the EMD compares histograms without first normalizing them, whereas the Wasserstein distance compares distributions of random variables. Until now we have deferred discussion of unequal masses because it is not amenable to our combinatorial approach: the northwest corner rule assumes equal masses, the RSK correspondence requires words of equal length, and the symmetric difference fails to embody an EMD if

the two Young diagrams are limited to different numbers of rows. Indeed, *a priori* it is a complicated combinatorial problem to find an optimal flow matrix F when the masses are different. We find it notable, therefore, that there is a direct route to the EMD in this complicated case, via representation theory. Our method is to separate polynomials into an invariant component (corresponding to the data points being matched) and a harmonic component (corresponding to the superfluous data points in the larger histogram).

In the Howe duality setting of this chapter, the GL_k-**harmonic polynomials** on V are defined as

$$\mathscr{H}(V) \coloneqq \left\{ h \in \mathbb{C}[V] \;\middle|\; \Delta(i,j) \cdot h = 0 \text{ for all } i,j \right\},$$

where $\Delta(i,j)$ are the differential operators from (4.14). When $k = 1$, with the first index dropped, we simply have $\Delta(i,j) = \frac{\partial^2}{\partial x_i \partial y_j}$ and hence

$$\begin{aligned}
\mathscr{H}(V) &= \operatorname{span}\Big(\{\mathbf{x}^\lambda \mid \lambda \in \mathbb{N}^n\} \cup \{\mathbf{y}^\mu \mid \mu \in \mathbb{N}^n\} \Big) \\
&= \bigoplus_{a \geqslant 0} \mathbb{C}[\mathbf{x}]_a \oplus \bigoplus_{a<0} \mathbb{C}[\mathbf{y}]_{-a} \\
&\cong \bigoplus_{a \in \mathbb{Z}} F_1^a \otimes \begin{cases} F_n^a \otimes \mathbb{1}, & a \geqslant 0, \\ \mathbb{1} \otimes F_n^{-a}, & a < 0 \end{cases} \\
&=: \bigoplus_{a \in \mathbb{Z}} \mathscr{H}_a(V)
\end{aligned}$$

as a $\mathrm{GL}_1 \times K$-module, where $\mathbb{C}[-]_a$ denotes the homogeneous polynomial functions of degree a. It is a standard fact of invariant theory that we have a surjection

$$\mathscr{P} \coloneqq \mathbb{C}[V]^{\mathrm{GL}_1} \otimes \mathscr{H}(V) \twoheadrightarrow \mathbb{C}[V]^{\mathrm{GL}_1} \cdot \mathscr{H}(V) = \mathbb{C}[V], \tag{4.23}$$

defined by pointwise multiplication of functions and extended by linearity. (We use the letter $\mathscr{P}$ to evoke the *product* of invariant and harmonic functions.) This "separation of variables" is a generalization of the familiar case of $\mathrm{SO}(3)$-harmonic functions in physics, namely those polynomial functions on $\mathbb{R}^3$ killed by the Laplace operator: in that case, because the map in (4.23) is an isomorphism, every polynomial function on $\mathbb{R}^3$ can be

written *uniquely* as a product of an SO(3)-invariant function (the radial component) and a harmonic function (the angular component). Upon restriction to the sphere, Laplace's spherical harmonics furnish an orthonormal basis for the space of the harmonic functions.

We will regard $\mathscr{P} = \widetilde{F}^0_{n,n} \otimes \mathscr{H}(V)$ as a K-module only; hence for the sake of concreteness, for each $a \in \mathbb{Z}$, we will write out the ath component as

$$\begin{aligned}\mathscr{P}_a &:= \widetilde{F}^0_{n,n} \otimes \mathscr{H}_a(V) \\ &= \mathbb{C}[x_i y_j] \otimes \begin{cases} \mathbb{C}[\mathbf{x}]_a \otimes \mathbb{1}, & a \geqslant 0, \\ \mathbb{1} \otimes \mathbb{C}[\mathbf{y}]_{-a}, & a < 0. \end{cases}\end{aligned}$$

Recall the functions φ and φ_q from (4.17) and (4.21); we will now naturally extend their domain from $\widetilde{F}^0_{n,n}$ to $\mathscr{P}$. In order to maintain a *linear* map, for a harmonic $h = h(\mathbf{x}) \otimes 1$ or $h = 1 \otimes h(\mathbf{y})$, we define the evaluation $h|_{\mathbf{1}} := h(1, \ldots, 1) \in \mathbb{C}$. Then for $f \otimes h \in \mathscr{P}$, we can define the following and extend by linearity:

$$\varphi(f \otimes h) := \varphi(f) \cdot h|_{\mathbf{1}}. \tag{4.24}$$

Example 4.18. Let $n = 2$, with $f = x_1^2 y_1 y_2 \in \mathbb{C}[x_i y_j]$, and $h = x_1^3 x_2 \otimes 1 \in \mathscr{H}_4(V)$. Then

$$\varphi(f \otimes h) = \varphi(x_1^2 y_1 y_2) \cdot (x_1^3 x_2)|_{\mathbf{1}} = z_{11} z_{12} \cdot 1 = z_{11} z_{12}.$$

Our motivation is that φ "sees" only the invariant factor f, and "ignores" the harmonic factor h: we will soon see that this is analogous to the EMD "ignoring" superfluous data points in the larger histogram.

Let $\lambda \in \mathcal{H}(\ell, n)$ and $\mu \in \mathcal{H}(m, n)$. Under the action of K, the pair (λ, μ) determines the weight space

$$\mathscr{P}_{\lambda,\mu} := \left\{ p \in \mathscr{P} \;\middle|\; \left(\begin{bmatrix} s_1 & & \\ & \ddots & \\ & & s_n \end{bmatrix}, \begin{bmatrix} t_1 & & \\ & \ddots & \\ & & t_n \end{bmatrix} \right) \cdot p = \mathbf{s}^\lambda \mathbf{t}^\mu \, p \right\}.$$

If we put $a = \ell - m$, then $\mathscr{P}_{\lambda,\mu} \subset \mathscr{P}_a$, and has the monomial basis

$$\mathscr{B}_{\lambda,\mu} = \begin{cases} \left\{ \mathbf{x}^{\lambda'} \mathbf{y}^\mu \otimes (\mathbf{x}^{\lambda - \lambda'} \otimes 1) \;\middle|\; \lambda - \lambda' \in \mathcal{H}(a, n) \right\}, & a \geqslant 0, \\ \left\{ \mathbf{x}^\lambda \mathbf{y}^{\mu'} \otimes (1 \otimes \mathbf{y}^{\mu - \mu'}) \;\middle|\; \mu - \mu' \in \mathcal{H}(-a, n) \right\}, & a < 0. \end{cases} \tag{4.25}$$

Clearly $\mathscr{B}_{\lambda,\mu}$ is finite, since there are only finitely many ways to remove a data points from a histogram. In the language of representation theory, $\#\mathscr{B}_{\lambda,\mu} = \dim \mathscr{P}_{\lambda,\mu}$ is called the **multiplicity** of the weight (λ, μ) in $\mathscr{P}$.

In the theorem below, by the **order** (ord) of a polynomial we will mean the *lowest* power with nonzero coefficient. (This agrees with the usual definition of the order of a power series.) The upshot of the theorem is that the EMD for unequal masses can be reduced to the equal-mass case, organized in terms of the weight space $\mathscr{P}_{\lambda,\mu}$ in the Howe duality setting.

Theorem 4.19. *Let $\lambda \in \mathcal{H}(\ell, n)$ and $\mu \in \mathcal{H}(m, n)$. Then*

$$\mathbf{EMD}(\lambda, \mu) = \operatorname{ord} \sum_{b \in \mathscr{B}_{\lambda,\mu}} \varphi_q(b).$$

Proof. Set $a = \ell - m$. Without loss of generality, assume $a \geqslant 0$ (if not, then just switch λ and μ). By the constraints (2.5) – (2.8) of the Monge problem, any flow matrix $F \in \mathcal{F}_{\lambda,\mu}$ must have column sums given by the smaller histogram μ, whereas the row sums are given by some histogram $\lambda' \in \mathcal{H}(m, n)$ such that $\lambda - \lambda' \in \mathcal{H}(a, n)$. In words, λ' must be obtained by removing a data points from λ. Hence in the general case we have

$$\mathbf{EMD}(\lambda, \mu) = \begin{cases} \min\{\mathbf{EMD}(\lambda', \mu) \mid \lambda - \lambda' \in \mathcal{H}(a, n)\}, & a \geqslant 0, \\ \min\{\mathbf{EMD}(\lambda, \mu') \mid \mu - \mu' \in \mathcal{H}(-a, n)\}, & a < 0. \end{cases} \tag{4.26}$$

In either case, $\mathbf{x}^{\lambda'}\mathbf{y}^{\mu}$ (respectively $\mathbf{x}^{\lambda}\mathbf{y}^{\mu'}$) is an element of $\mathbb{C}[x_i y_j]$, which we know how to relate to the EMD via the function φ_q. Indeed, comparing (4.21), (4.24), (4.25), and (4.26), we see that $\mathbf{EMD}(\lambda, \mu)$ equals the smallest i such that $q^i \in \{\varphi_q(b) \mid b \in \mathscr{B}_{\lambda,\mu}\}$. The theorem follows immediately. □

Example 4.20. When λ and μ do have the same mass, then $a = 0$ and $\mathscr{B}_{\lambda,\mu}$ is the singleton $\{\mathbf{x}^{\lambda}\mathbf{y}^{\mu} \otimes (1 \otimes 1)\}$. Hence Theorem 4.19 gives us $\mathbf{EMD}(\lambda, \mu) = \operatorname{ord} \varphi_q(\mathbf{x}^{\lambda}\mathbf{y}^{\mu} \otimes 1 \otimes 1) = \operatorname{ord} q^{\mathbf{EMD}(\lambda,\mu)} = \mathbf{EMD}(\lambda, \mu)$, which is just as trivial as we would expect.

Example 4.21. When λ **contains** μ, in the sense that $\lambda_i \geqslant \mu_i$ for all $i \in [n]$, then $\mathbf{EMD}(\lambda, \mu) = 0$. In fact, this is a tautology when we consider that in general the EMD is the minimum

amount of work required to make the larger histogram contain the other. But this is also clear from Theorem 4.19, since when λ contains μ we will have $\lambda - \mu \in \mathcal{H}(a, n)$, and so $\mathbf{x}^{\mu}\mathbf{y}^{\mu} \otimes (\mathbf{x}^{\lambda-\mu} \otimes 1) \in \mathscr{B}_{\lambda,\mu}$. Clearly $\varphi_q(\mathbf{x}^{\mu}\mathbf{y}^{\mu} \otimes \mathbf{x}^{\lambda-\mu} \otimes 1) = \varphi_q(\mathbf{x}^{\mu}\mathbf{y}^{\mu}) = q^0$, so we must have $\mathbf{EMD}(\lambda, \mu) = 0$.

Example 4.22. Here is an easy but no longer trivial example. Let $n = 3$, with $\lambda = (1, 1, 4) \in \mathcal{H}(6, 3)$ and $\mu = (2, 2, 0) \in \mathcal{H}(4, 3)$. Then $a = 6 - 4 = 2$, and we have

$$\mathscr{B}_{\lambda,\mu} = \left\{ \mathbf{x}^{\lambda'}\mathbf{y}^{\mu} \otimes \mathbf{x}^{\lambda-\lambda'} \otimes 1 \;\middle|\; \lambda' \in \{(1,1,2),\ (1,0,3),\ (0,1,3),\ (0,0,4)\} \right\}.$$

The reader can check that these four choices for λ' are the only ways to remove two data points from λ. We now have

$$\begin{aligned} \operatorname{ord} \sum_{b \in \mathscr{B}_{\lambda,\mu}} \varphi_q(b) &= \operatorname{ord}\left[\varphi_q\left(\mathbf{x}^{(1,1,2)}\mathbf{y}^{\mu}\right) + \varphi_q\left(\mathbf{x}^{(1,0,3)}\mathbf{y}^{\mu}\right) + \varphi_q\left(\mathbf{x}^{(0,1,3)}\mathbf{y}^{\mu}\right) + \varphi_q\left(\mathbf{x}^{(0,0,4)}\mathbf{y}^{\mu}\right) \right] \\ &= \operatorname{ord}(q^3 + q^4 + q^3 + q^6) \\ &= 3, \end{aligned}$$

and hence $\mathbf{EMD}(\lambda, \mu) = 3$.

As discussed in [BL01], the fact that the EMD is zero whenever one histogram contains the other (as in our Example 4.21) is likely a drawback in most *statistical* applications (although it may have some merit when applied to image retrieval, for example). In particular, if one histogram is much smaller than the other, then the EMD tends to be determined by the smaller histogram alone, losing all information about the larger one. In the next chapter, we introduce a variation of the EMD, inspired directly by the structure of crystal bases in Types B and C, which remedies this disadvantage.

5 THE EMD AS A METRIC ON CRYSTAL GRAPHS

As we will show, a natural setting for understanding the EMD combinatorially arises from the study of quantum groups, which are deformations of the universal enveloping algebras of Lie groups. At the heart of quantum group theory is the notion of a crystal base, a combinatorial object analogous to a representation. Crystal bases are typically depicted as oriented graphs; we show how on certain Type-A crystal graphs, upon identifying the words $\mathrm{w}(\lambda)$ of histograms with crystal base elements in a natural way, the EMD is precisely the graph distance. Moreover, the analogous crystal bases in Types B and C suggest a variation on the EMD which we believe improves upon the statistical drawbacks of partial matching (i.e., unequal histogram masses) discussed at the end of the previous chapter.

5.1 Crystal base preliminaries

Throughout this chapter, we follow the exposition from the book [BS17] by Bump and Schilling. In the context of the EMD, we will view crystal bases as purely combinatorial objects, but we should say a word about the depth of representation theory behind them. A **quantum group** $U_q(\mathfrak{g})$ is a Hopf algebra that is a sort of q-analogue, or "noncommutative" analogue, to the universal enveloping algebra $U(\mathfrak{g})$ of a Lie group. Modules over quantum groups are said to have "crystal bases" — the name is derived from the fact that these bases emerge from setting the parameter $q = 0$, analogous to the physical notion of a "perfect crystal" that exists only at the temperature absolute zero. The theory of crystal bases was pioneered in the 1990s by Kashiwara [Kas91], parallel with Lusztig [Lus90] and Littelmann [Lit94], and has applications in mathematical physics, statistical mechanics, and number theory, among other areas. Another major application is the representation theory of certain infinite-dimensional ("Kac–Moody") Lie algebras. A standard reference

on crystal bases from the quantum group perspective is [HK02].

A **(Kashiwara) crystal** is an object associated with a given root system Φ with index set I and weight lattice Λ. Keeping representation theory at the back of our minds, we can regard I as indexing the simple roots $\alpha_i \in \Phi$; in this thesis, we will consider root systems Φ of types A, B, and C. Axiomatically, a crystal is a nonempty set $\mathcal{B}$ along with maps

$$\begin{aligned} e_i, f_i &: \mathcal{B} \longrightarrow \mathcal{B} \sqcup \{0\}, \\ \varepsilon_i, \varphi_i &: \mathcal{B} \longrightarrow \mathbb{Z} \sqcup \{-\infty\}, \\ \mathrm{wt} &: \mathcal{B} \longrightarrow \Lambda, \end{aligned}$$

for $i \in I$, and 0 an auxiliary element extrinsic to $\mathcal{B}$; these maps satisfy the following:

- For $x, y \in \mathcal{B}$, we have $e_i(x) = y$ if and only if $f_i(y) = x$.
- In the previous bullet point, we then have $\mathrm{wt}(y) = \mathrm{wt}(x) + \alpha_i$. Intuitively, we should think of the **raising operator** e_i and the **lowering operator** f_i as corresponding to the simple root α_i, where $\mathrm{wt}(x)$ is truly a **weight** in the sense of Lie theory.
- Continuing from the first bullet point, $\varepsilon_i(y) = \varepsilon_i(x) - 1$ and $\varphi_i(y) = \varphi_i(x) + 1$. Hence ε_i decreases and φ_i increases as we apply the raising operator e_i successively.
- For all $x \in \mathcal{B}$ and $i \in I$, we must have $\varphi_i(x) = \langle \mathrm{wt}(x), \alpha_i^\vee \rangle + \varepsilon_i(x)$. Intuitively, we should think of ε_i and φ_i as the distance from the "higher" (respectively, "lower") end of a root string.

The operators e_i and f_i are sometimes called **Kashiwara operators**. An element $x \in \mathcal{B}$ such that $e_i(x) = 0$ for all $i \in I$ is called a **highest weight element** of $\mathcal{B}$, while its weight $\mathrm{wt}(x)$ is called a **highest weight**. Notice that the terminology is completely analogous to the highest weight of a representation. To a crystal $\mathcal{B}$ we associate the **crystal graph**, whose vertices are the elements of $\mathcal{B}$; whenever $f_i(x) = y$, we draw an edge from x to y labeled by i, as follows: $x \xrightarrow{i} y$.

To bring this all back down to earth (moving), we will focus on the special case of crystal graphs whose highest weight is a multiple of the first fundamental weight ϖ_1, where $\Phi = A_{n-1}$, B_n, or C_n, depending on the application.

5.2 Type A: EMD as graph distance

We start by taking $\Phi = A_{n-1}$, and we will use the GL_n weight lattice (rather than SL_n), so that $\Lambda = \mathbb{Z}^n$. We have $I = \{1, \ldots, n-1\}$, where $\alpha_i = \varepsilon_i - \varepsilon_{i+1}$, and $\varpi_1 = \varepsilon_1$. (We will not refer again to the string lengths ε_i and φ_i from the previous section, and so we continue to write the weight ε_i without risk of confusion.) For each partition $\xi \in P_+(\mathfrak{gl}_n)$, we have an associated crystal $\mathcal{B}_\xi$ with highest weight ξ. The crystal graph of $\mathcal{B}_\xi$ is connected and is a combinatorial picture of the multiset of weights of the GL_n-module F_n^ξ; in the language of quantum groups, we say that $\mathcal{B}_\xi$ is the crystal of the $U_q(\mathfrak{gl}_n)$-module with highest weight ξ. Moreover, there is a bijective correspondence between $\mathcal{B}_\xi$ and the set of SSYT's of shape ξ; for this reason, Type-A crystals are often called "crystals of tableaux." The weight function simply sends an SSYT to its **content**, in the sense of (3.1).

In the special case where the highest weight is a multiple of the first fundamental weight, namely $m\varpi_1 = (m) := (m, 0, \ldots, 0)$, the SSYT's in $\mathcal{B}_{(m)}$ are in fact one-row tableaux, and so we are in familiar territory from Section 3.1. In particular, given a histogram $\lambda \in \mathcal{H}(m, n) \subset \Lambda$, we immediately have

$$\mathrm{w}(\lambda) \in \mathcal{B}_{(m),}$$
$$\mathrm{wt}(\mathrm{w}(\lambda)) = \lambda.$$

Hence we have a bijection $\mathcal{H}(m, n) \longleftrightarrow \mathcal{B}_{(m)}$. Although for general shapes ξ there are several equivalent combinatorial formulations for the effect of applying the raising and lowering operators e_i and f_i (see [BS17, Chapter 3]), the one-row case is particularly straightforward. On the SSYT's themselves, e_i changes the leftmost entry $i+1$ into an i, while f_i changes the rightmost i into an $i+1$ (and if these are not possible, then the operators

kill the SSYT by sending it to 0). Clearly the highest weight element in $\mathcal{B}_{(m)}$ is the row $\boxed{1}\cdots\boxed{1}$, whose weight is indeed the histogram in $\mathcal{H}(m,n)$ with the highest weighted total. We can also see that $\mathcal{B}_{(m)}$ can be regarded as a basis of weight vectors for the GL_n-module $F_n^{(m)} \cong S^m(\mathbb{C}^n)$, which we have already viewed as the free vector space on $\mathcal{H}(m,n)$ in the previous chapter.

As a visual example, in Figure 7 we show the crystal $\mathcal{B}_{(4)}$ for the $U_q(\mathfrak{gl}_3)$-module with highest weight (4). From the crystal perspective, each $x \in \mathcal{B}_{(4)}$ is depicted as a one-row SSYT, with $\mathrm{wt}(x)$ displayed beneath; from the histogram perspective, each $\lambda \in \mathcal{H}(4,3)$ is denoted by an ordered triple, with $\mathrm{w}(\lambda)$ displayed above. Because $n = 3$ and thus $\Phi = A_2$, we can easily depict the crystal in two dimensions, one for each simple root. To reduce the clutter, we suppress the labels on the edges, but it is easy to see that each horizontal (blue) arrow should be labeled "1," while each vertical (red) arrow should be labeled "2." The highest weight element is in the upper-right corner.

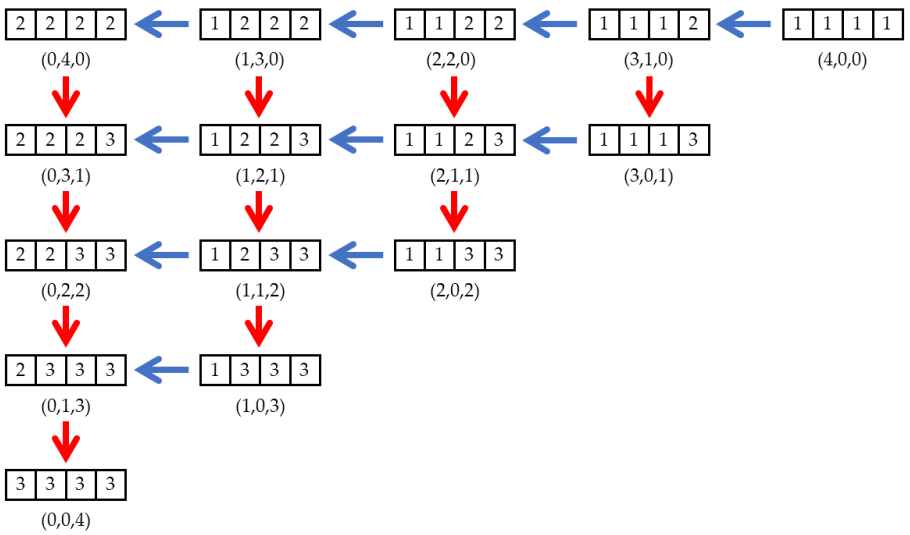


Figure 7: The crystal $\mathcal{B}_{(4)}$ for type $\Phi = A_2$. Each element corresponds (via its weight, displayed just below) to a histogram in $\mathcal{H}(4,3)$.

We now apply all this to the EMD. Applying e_i to $\mathrm{w}(\lambda)$, and then taking the weight of the result, results in the histogram $\lambda + \alpha_i$: in other words, identifying histograms with weights, e_i moves a data point from bin $i+1$ to bin i if possible, and sends $\mathrm{w}(\lambda)$ to 0

otherwise (i.e., if $\lambda_i = 0$, meaning that there are no data points in bin $i+1$ to be moved). Likewise, f_i moves a data point from bin i to bin $i+1$ if possible. In other words, the operators e_i and f_i each contribute one unit of "work" in the EMD sense; furthermore, in the language of this thesis, the raising operator e_i truly does raise the weighted total $t(\lambda)$ by one, while the lowering operator f_i lowers $t(\lambda)$ by one. As is standard in graph theory, the **distance** between two vertices in a crystal graph is the length of the shortest path (i.e., graph geodesic) connecting them; here we disregard the edge orientation in defining distance. From the preceding discussion, we obtain the following connection to the EMD.

Proposition 5.1. *Let $\lambda, \mu \in \mathcal{H}(m,n)$. Then* $\mathbf{EMD}(\lambda, \mu)$ *equals the distance between* $\mathrm{w}(\lambda)$ *and* $\mathrm{w}(\mu)$ *in the crystal graph of $\mathcal{B}_{(m)}$ for type $\Phi = A_{n-1}$.*

Remark. The authors of [BW20] recursively define a generating function whose partial derivative encodes the sum of EMD's over all histogram pairs in $\mathcal{H}(m,n) \times \mathcal{H}(m,n)$. In light of Proposition 5.1, we can reinterpret their result as a generating function for the Wiener index of the crystal graph $\mathcal{B}_{(m)}$. The **Wiener index** (also called the gross status or transmission) of a graph is the sum of all pairwise distances, and is an important topological invariant arising in mathematical chemistry and network theory. (See [Wie47].)

Remark. We can view this entire section as a cleaner perspective on the Young's sublattice $\mathcal{Y}(m,\, n-1)$ from Section 3.7. In fact, as a graph, the crystal $\mathcal{B}_{(m)}$ is isomorphic to the Hasse diagram of $\mathcal{Y}(m,\, n-1)$ endowed with edges $X \to Y$ whenever X covers Y, via the map $\mathrm{w}(\lambda) \longmapsto \mathrm{Y}(\lambda)$. But despite the two graphs being the same, nevertheless the crystal combinatorics certainly reveals a more elegant concept of distance than does Young's lattice in Proposition 3.13.

Remark. The **affine root system** $A_{n-1}^{(1)}$ is obtained by adjoining an additional simple root $\alpha_0 = \varepsilon_n - \varepsilon_1$, called the **affine root**. (In general, the affine root is the negative of the highest root; in Type A, on the Dynkin diagram, α_0 is joined to α_1 and to α_n by a single edge, forming a circle.) The affine crystal graph $\mathcal{B}^{1,m}$ has the same vertices as $\mathcal{B}_{(m)}$, but

with additional edges $x \xrightarrow{0} y$ whenever y is obtained from x by changing an n into a 1 (and moving that cell to the far left in order to maintain an SSYT). In terms of the EMD, this corresponds to moving a data point from bin n to bin 1. In our typical histograms, the cost of such a move is $n-1$ rather than only 1; hence the affine crystal graph corresponds to the EMD where the feature space consists of n bins in a circle, with bins 1 and n next to each other. The EMD between circular histograms is prominent in the field of computer vision; see [RDG08].

5.3 Type C: EMD as distance from the zero weight space

In these next two sections, we define a variation of the EMD purely in terms of crystals. Root system types B and C are nearly identical in their connection to the EMD, the only difference ultimately arising from the parity of the total mass of the histograms. Type C is slightly more straightforward, so we postpone Type B until the next section. Details can be found in [BS17, Chapter 6].

For $\Phi = C_n$, we take the weight lattice $\Lambda = \mathbb{Z}^n$ of the symplectic group $\mathrm{Sp}_{2n}(\mathbb{C})$. We have $I = [n]$, and the simple roots are the same $\alpha_1, \ldots, \alpha_{n-1}$ from Type A, with the addition of $\alpha_n = 2\varepsilon_n$. Just as in Type A, crystals in Type C can be realized as crystals of SSYT's sharing a common shape, but the notion of "semistandard" and the combinatorics of the operators e_i and f_i are now much more complicated. Nonetheless, because we will restrict our attention to those crystals whose highest weight is a multiple of the first fundamental weight $\varpi_1 = \varepsilon_1$, our tableaux will have only one row, which simplifies the situation considerably.

The alphabet for our tableaux is no longer $[n]$, but rather the totally ordered set consisting of the following $2n$ elements:

$$1 \prec 2 \prec \cdots \prec n \prec \overline{n} \prec \cdots \prec \overline{2} \prec \overline{1}.$$

It is still the case that a one-row tableau is semistandard if and only if its entries are

weakly increasing. We now show how a one-row Type-C crystal graph relates to the EMD between histograms. This will differ from Type A in two important ways. First, each SSYT will encode an ordered pair of histograms, rather than a single histogram, and thus our EMD will ultimately be defined for vertices rather than vertex pairs. Second, we allow the two histograms to have different masses; in Type C, we require that their combined mass be an even number. (Type B will cover the case when the combined mass is odd.) For $\lambda \in \mathcal{H}(\ell, n)$ and $\mu \in \mathcal{H}(m, n)$, such that $\ell + m$ is even, we define the joint word

$$\mathrm{w}(\lambda, \mu) = \underbrace{\boxed{1}\cdots\boxed{1}}_{\lambda_1}\underbrace{\boxed{2}\cdots\boxed{2}}_{\lambda_2}\cdots\underbrace{\boxed{n}\cdots\boxed{n}}_{\lambda_n}\underbrace{\boxed{\overline{n}}\cdots\boxed{\overline{n}}}_{\mu_n}\cdots\underbrace{\boxed{\overline{2}}\cdots\boxed{\overline{2}}}_{\mu_2}\underbrace{\boxed{\overline{1}}\cdots\boxed{\overline{1}}}_{\mu_1},$$

which by its construction is a one-row SSYT of length $\ell + m$. In this way, we use the alphabet of barred and unbarred elements to encode histogram pairs in a single row: we simply regard each entry i as a data point in bin i of λ, and each entry $\bar{\imath}$ as a data point in bin i of μ.

The raising and lowering operators e_i and f_i act as follows, for $1 \leqslant i < n$:

- If possible, e_i changes the leftmost $\bar{\imath}$ to $\overline{i+1}$; if not possible, then it changes the leftmost $i + 1$ to i; if still not possible, then it kills the SSYT.
- If possible, f_i changes the rightmost $\overline{i+1}$ to $\bar{\imath}$; if not possible, then it changes the rightmost i to $i + 1$; if still not possible, then it kills the SSYT.

In the language of histograms, note that each e_i above raises the weighted difference $\mathbf{D}(\lambda, \mu)$ by 1, since it either demotes a data point in μ or promotes a data point in λ. Likewise, each f_i above lowers $\mathbf{D}(\lambda, \mu)$ by 1.

The $i = n$ case is special:

- If possible, e_n changes the leftmost $\overline{n}$ to n; otherwise it kills the SSYT.
- If possible, f_n changes the rightmost n to $\overline{n}$; otherwise it kills the SSYT.

Hence e_n and f_n have no effect on $\mathbf{D}(\lambda, \mu)$; they simply transfer a data point between bin n of λ and bin n of μ.

The weight of an SSYT is defined as follows: each entry i contributes ε_i to the weight, and each entry $\bar{\imath}$ contributes $-\varepsilon_i$. In other words, the ith component of the weight equals the number of i's minus the number of $\bar{\imath}$'s. It follows that $\mathrm{wt}(x) = \mathbf{0}$ if and only if the SSYT x contains an equal number of i's and $\bar{\imath}$'s, for all $i \in [n]$. This has an important consequence: namely, $\mathrm{wt}(\mathrm{w}(\lambda, \mu)) = \mathbf{0}$ if and only if $\lambda = \mu$. In other words, the weight-zero condition is precisely the objective of the Monge problem that defines the EMD: to transform two histograms into the same histogram. This fact motivates us to define a variation of the EMD purely in terms of Type-C crystals, in which the relevant distance is measured from the nearest weight-zero vertex.

Definition 5.2. Let $\lambda \in \mathcal{H}(\ell, n)$ and $\mu \in \mathcal{H}(m, n)$, where $\ell + m$ is even. Then we define

$$\mathbf{EMD}_C(\lambda, \mu) := \text{distance from } \mathrm{w}(\lambda, \mu) \text{ to the zero weight space in } \mathcal{B}_{(\ell+m)} \text{ for } \Phi = C_n.$$

When the histograms have the same mass, this new definition coincides with the original EMD:

Proposition 5.3. *Let $\lambda, \mu \in \mathcal{H}(m, n)$. Then $\mathbf{EMD}_C(\lambda, \mu) = \mathbf{EMD}(\lambda, \mu)$.*

Proof. Any shortest path from $\mathrm{w}(\lambda, \mu)$ to a weight-zero vertex must not contain any edges labeled n, since e_n and f_n change the total number of barred and unbarred entries in an SSYT (and hence the masses of the histograms). Therefore a shortest path contains only edges labeled $1, \ldots, n-1$. But since the corresponding α_i are identical to the α_i in Type A, the corresponding operators e_i and f_i translate to one unit of work in the EMD sense. Hence the distance to a weight-zero vertex, which is the minimum number of e_i's and f_i's required to equalize λ and μ, is precisely the minimum amount of work required to transform both λ and μ into some $\nu \in \mathcal{H}(m, n)$. By definition this minimum equals $\mathbf{EMD}(\lambda, \mu)$. □

In the case of histograms with different masses, $\mathbf{EMD}_C$ no longer agrees with the original definition of **EMD** in terms of the Monge problem — but we believe that $\mathbf{EMD}_C$ has the advantage in certain applications. We illustrate with some examples.

Example 5.4. Recall from Example 4.21 that whenever λ contains μ (i.e., $\lambda_i \geqslant \mu_i$ for all i), we have $\mathbf{EMD}(\lambda, \mu) = 0$. In certain applications this property might be useful, but statistically it seems undesirable. For example, let $\lambda = (1,1,1,1)$ and $\mu = (1,0,0,1)$, so that $m + \ell = 6$ and $\mathbf{EMD}(\lambda, \mu) = 0$. Our newly defined $\mathbf{EMD}_C(\lambda, \mu)$ is nonzero, however, since $\mathrm{wt}(\mathrm{w}(\lambda, \mu)) = (0,1,1,0) \neq \mathbf{0}$. Indeed, the shortest path to a weight-zero vertex comprises four edges; one of several such paths is shown below:

$$
\begin{aligned}
\mathrm{w}(\lambda,\mu) = {} & \boxed{1}\boxed{2}\boxed{3}\boxed{4}\boxed{\overline{4}}\boxed{\overline{1}}, && \mathrm{wt} = (0,1,1,0),\\
\xrightarrow{f_4} {} & \boxed{1}\boxed{2}\boxed{3}\boxed{\overline{\mathbf{4}}}\boxed{\overline{4}}\boxed{\overline{1}}, && \mathrm{wt} = (0,1,1,-2),\\
\xrightarrow{f_3} {} & \boxed{1}\boxed{2}\boxed{3}\boxed{\overline{4}}\boxed{\overline{\mathbf{3}}}\boxed{\overline{1}}, && \mathrm{wt} = (0,1,0,-1),\\
\xrightarrow{f_3} {} & \boxed{1}\boxed{2}\boxed{3}\boxed{\overline{\mathbf{3}}}\boxed{\overline{3}}\boxed{\overline{1}}, && \mathrm{wt} = (0,1,-1,0),\\
\xrightarrow{f_2} {} & \boxed{1}\boxed{2}\boxed{3}\boxed{\overline{3}}\boxed{\overline{\mathbf{2}}}\boxed{\overline{1}}, && \mathrm{wt} = (0,0,0,0).
\end{aligned}
$$

Hence $\mathbf{EMD}_C(\lambda, \mu) = 4$. Note how each application of f_i corresponds to subtracting the simple root α_i from the weight. From the perspective of histograms, the four lowering operators first moved a data point from bin 4 of λ to bin 4 of μ; then from bin 4 to bin 3 of μ; then again from bin 4 to bin 3 of μ; then finally from bin 3 to bin 2 of μ. We also point out that the ordering among the lowering operators can be changed, and as long as the SSYT is not killed along the way, the final result will still be a (possibly different) weight-zero vertex.

Example 5.5. First let $\lambda = (1,0,0,0,3)$ and $\mu = (1,0,0,0,1)$, so that $\mathbf{EMD}(\lambda, \mu) = 0$. To determine $\mathbf{EMD}_C(\lambda, \mu)$, we observe that

$$
\mathrm{w}(\lambda,\mu) = \boxed{1}\boxed{5}\boxed{5}\boxed{5}\boxed{\overline{5}}\boxed{\overline{1}} \xrightarrow{f_5} \boxed{1}\boxed{5}\boxed{5}\boxed{\overline{\mathbf{5}}}\boxed{\overline{5}}\boxed{\overline{1}},
$$

whose weight is $\mathbf{0}$, and so $\mathbf{EMD}_C(\lambda, \mu) = 1$.

Now by contrast, let $\lambda = (3, 0, 0, 0, 1)$ and keep $\mu = (1, 0, 0, 0, 1)$. The EMD is still 0, but the $\mathbf{EMD}_C$ is now much greater than 1: as the reader can check, we must compose nine lowering operators in order to arrive at the weight $\mathbf{0}$, and so $\mathbf{EMD}(\lambda, \mu) = 9$.

As Example 5.5 illustrates, when using $\mathbf{EMD}_C$, it costs only one unit of work to transfer a data point between bin n of λ and bin n of μ, since this can be done by a single operator e_n or f_n. At the opposite extreme, it costs $2n - 1$ units of work to transfer a data point between bin 1 of λ and bin 1 of μ. This is because the only way to transfer data points from one histogram to the other is via bin n. While this asymmetry may seem to be a strange property, we actually believe that it is desirable for applications in which the weighted difference $\mathbf{D}$ is significant: after all, data points in bin n contribute nothing to the weighted totals, while those in bin 1 contribute maximally. Said another way, transferring an A student from one class to the other has a large effect upon the difference in GPA's, while transferring an F student has almost no effect. Therefore we believe that $\mathbf{EMD}_C$ is a natural candidate for a metric on histograms, and in many cases may be a better choice than the classical EMD or the Wasserstein distance.

5.4 Type B: *mutatis mutandis*

Our new metric $\mathbf{EMD}_C$ from the previous section can be thought of algorithmically as a two-stage process: first, repeatedly apply lowering operators f_i (or raising operators e_i, if $\ell < m$) in order to transfer data points from the larger to the smaller histogram via bin n until the masses both equal $\frac{\ell+m}{2}$; second, apply a combination of e_i and f_i for $1 \leqslant i < n$ in order to equalize the two resulting histograms; the EMD is the minimum number of Kashiwara operators required to carry out this process. A difficulty arises, however, when $\frac{\ell+m}{2}$ is not an integer, since then there is no way to equalize the two masses. This is the difficulty that is remedied by Type-B crystals, where the root system allows for annihilating or creating one data point.

For $\Phi = B_n$, we take $\Lambda = \mathbb{Z}^n$ to be the weight lattice of the special orthogonal group $\mathrm{SO}(2n+1, \mathbb{C})$. We have $I = [n]$, and the simple roots are the same $\alpha_1, \ldots, \alpha_{n-1}$ from Type A, along with $\alpha_n = \varepsilon_n$. Once again we will restrict our attention to crystals whose highest weight is a multiple of $\varpi_1 = \varepsilon_1$, and so the tableaux in our crystals will have one row of length $\ell + m$. The alphabet for our tableaux is the same as for Type C, with the addition of 0, as follows:

$$1 \prec 2 \prec \cdots \prec n \prec 0 \prec \overline{n} \prec \cdots \prec \overline{2} \prec \overline{1}.$$

As before, in order to be semistandard, a one-row tableau must have weakly increasing entries, with the added condition that no more than one "0" entry may appear. Once again, each entry i contributes ε_i to the weight, each $\bar{\imath}$ contributes $-\varepsilon_i$, and the entry 0 contributes $\mathbf{0}$.

The definition of the joint word $\mathrm{w}(\lambda, \mu)$ remains the same as in Type C, as do the actions of the operators e_i and f_i for $1 \leqslant i < n$. The only difference, in fact, is the action of e_n and f_n:

- If possible, e_n changes the leftmost $\overline{n}$ to 0; if not possible, then it changes a 0 into n; if still not possible, then it kills the SSYT.
- If possible, f_n changes a 0 into $\overline{n}$; if not possible, then it changes the rightmost n into 0; if still not possible, then it kills the SSYT.

Intuitively, e_n and f_n either annihilate or create a data point in bin n of one of the histograms. When the histograms' combined mass $\ell + m$ is odd, annihilating or creating a data point (and never more than one at a time, due to the condition that a semistandard row must contain no more than one "0" entry) changes that combined mass into an even number, so that it is possible to attain the weight $\mathbf{0}$, just as in Type C. Whereas in Type C we could expend a unit of work to transfer a data point from bin n of λ to bin n of μ, here in Type B we must expend a unit of work to annihilate a data point in bin n of λ, and then expend another unit of work to create a data point in bin n of μ. As before, we are led to

define a Type-B variant of the EMD:

Definition 5.6. Let $\lambda \in \mathcal{H}(\ell, n)$ and $\mu \in \mathcal{H}(m, n)$, where $\ell + m$ is odd. Then we define

$$\mathbf{EMD}_B(\lambda, \mu) := \text{distance from } \mathrm{w}(\lambda, \mu) \text{ to the zero weight space in } \mathcal{B}_{(\ell+m)} \text{ for } \Phi = B_n.$$

Example 5.7. Let $\lambda = (0, 0, 2, 1, 0)$ and $\mu = (0, 1, 1, 0, 0)$, so that $\ell + m = 3 + 2 = 5$. Then among the shortest paths to the zero weight space we have the following path of length three:

$$\begin{aligned}
\mathrm{w}(\lambda, \mu) = &\ \boxed{3}\boxed{3}\boxed{4}\boxed{\overline{3}}\boxed{\overline{2}}, && \mathrm{wt} = (0, -1, 1, 1, 0), \\
\xrightarrow{f_4} &\ \boxed{3}\boxed{3}\boxed{\mathbf{5}}\boxed{\overline{3}}\boxed{\overline{2}}, && \mathrm{wt} = (0, -1, 1, 0, 1), \\
\xrightarrow{f_5} &\ \boxed{3}\boxed{3}\boxed{\mathbf{0}}\boxed{\overline{3}}\boxed{\overline{2}}, && \mathrm{wt} = (0, -1, 1, 0, 0), \\
\xrightarrow{e_2} &\ \boxed{3}\boxed{3}\boxed{0}\boxed{\overline{3}}\boxed{\overline{\mathbf{3}}}, && \mathrm{wt} = (0, 0, 0, 0, 0).
\end{aligned}$$

Therefore $\mathbf{EMD}_B(\lambda, \mu) = 3$. This length-3 crystal path translates to the following: moving a data point in λ from bin 4 down to bin 5; then annihilating a data point in bin 5 of λ; then moving a data point in μ from bin 2 down to bin 3.

We find it striking that the combinatorics of crystal graphs are so amenable to capturing and generalizing the EMD, while also respecting the weighted difference $\mathbf{D}$ (even to the extent that raising and lowering operators literally raise and lower $\mathbf{D}$). An interesting question for further research is whether there is a natural application for the EMD analogue in Type D, where the root system seems to be less adaptable to the comparison of histograms.

6 GENERALIZATION TO MORE THAN TWO HISTOGRAMS

In this chapter, we generalize the EMD and our previous combinatorial approach in order to compare arbitrarily many histograms at a time. Then, extending the methods of [BW20], we devote much of the chapter to computing the expected value of this generalized EMD. We also introduce a natural analogue for the weighted difference **D** in this setting, and interpret the distribution of **D**-values as a character of a virtual representation, which we visualize using weight diagrams. Most of the results in this chapter have been published in the author's paper [Eri21] or written down in the preprint [Eria].

6.1 Definition of the EMC

Because our generalized EMD will compare more than two histograms, the term "distance" is no longer fitting; therefore from now on we will refer to the **earth mover's coefficient (EMC)**. Throughout this chapter, we let d denote the number of histograms under comparison. Since we will work with d-tuples of histograms, we need a way to write down elements of the d-fold Cartesian product $\mathcal{H}(m,n)^d$, and so we use a boldface Greek letter

$$\boldsymbol{\lambda} = (\lambda^1, \ldots, \lambda^d),$$

where each $\lambda^i = (\lambda^i_1, \ldots, \lambda^i_n) \in \mathcal{H}(m,n)$.

Next we need to generalize the cost matrix C from (2.1) to become a d-dimensional, $n \times \cdots \times n$ cost *array*. For an array position $\mathbf{x} = (x_1, \ldots, x_d) \in [n]^d$, we wish to define the cost entry $C(\mathbf{x})$ to be the taxicab distance between $\mathbf{x}$ and the line $\mathbb{N}(1, \ldots, 1)$, the main diagonal of the array. (We will justify this choice below, but for now we observe that in the $d = 2$ case this agrees with the original definition $C_{ij} = |i - j|$.) The formula for the d-dimensional taxicab distance from a point to a line is derived in [Col19]; taking the main

diagonal as the line in question, the formula can be reduced to

$$C(\mathbf{x}) = \min_{i \in [d]} \left\{ \sum_{j \in [d]} |x_i - x_j| \right\}. \tag{6.1}$$

We introduce, however, a more direct and convenient way to compute $C(\mathbf{x})$. Given $\mathbf{x} \in [n]^d$, let $x_{(i)}$ denote the ith **order statistic**, namely, the ith-smallest component of $\mathbf{x}$. For example, if $\mathbf{x} = (7, 1, 5, 3, 4)$, then $x_{(1)} = 1, x_{(2)} = 3, \ldots, x_{(5)} = 7$.

Proposition 6.1. *The cost in* (6.1) *can be computed as* $C(\mathbf{x}) = \sum_{i=1}^{\lfloor d/2 \rfloor} x_{(d-i+1)} - x_{(i)}$.

As an example before the proof, take $\mathbf{x} = (7, 1, 5, 3, 4)$ as above. Then by the proposition, to compute $C(\mathbf{x})$, we instead look at $\left(x_{(1)}, \ldots, x_{(d)}\right)$ and sum up the pairwise differences working outside-in (ignoring the median if d is odd, as it is here):

$$\begin{aligned} \left(x_{(1)}, \ldots, x_{(d)}\right) &= (1, \overbrace{3, 4, \overbrace{5}^{5-3=\mathbf{2}}}^{7-1=\mathbf{6}}, 7), \\ \text{therefore } C(\mathbf{x}) &= 6 + 2 \\ &= 8. \end{aligned}$$

Proof. For fixed $i \in [d]$, let i^* be such that $x_{(i^*)} = x_i$. Then we have

$$\begin{aligned} \sum_{j \in [d]} |x_i - x_j| = \left(x_{(2)} - x_{(1)}\right) + 2\left(x_{(3)} - x_{(2)}\right) + 3\left(x_{(4)} - x_{(3)}\right) + \cdots + (i^* - 1)\left(x_{(i^*)} - x_{(i^*-1)}\right) \\ + \left(x_{(d)} - x_{(d-1)}\right) + 2\left(x_{(d-1)} - x_{(d-2)}\right) + \cdots + (d - i^*)\left(x_{(i^*+1)} - x_{(i^*)}\right), \end{aligned}$$

which is minimized when $i^* = \lfloor \frac{d+1}{2} \rfloor$. Making this evaluation in the displayed sum, we find that the sum telescopes: when d is even, we obtain

$$-x_{(1)} - x_{(2)} - \cdots - x_{(\lfloor \frac{d+1}{2} \rfloor)} + x_{(\lfloor \frac{d+1}{2} \rfloor + 1)} + \cdots + x_{(d)},$$

and when d is odd, we obtain

$$-x_{(1)} - x_{(2)} - \cdots - x_{(\lfloor \frac{d+1}{2} \rfloor - 1)} + x_{(\lfloor \frac{d+1}{2} \rfloor + 1)} + \cdots + x_{(d)}.$$

In either case, this simplifies as

$$C(\mathbf{x}) = \sum_{i=1}^{\lfloor d/2 \rfloor} x_{(d-i+1)} - x_{(i)}. \qquad \square$$

Now that we have defined a cost array C, we can pose the d-fold analogue of the Monge problem from Section 2.2. Let $\lambda^1, \ldots, \lambda^d \in \mathcal{H}(m, n)$. We want to find a d-dimensional, $n \times \cdots \times n$ flow array $F = F(\mathbf{x})$ that solves the following:

$$\text{Minimize} \quad \sum_{\mathbf{x} \in [n]^d} C(\mathbf{x}) F(\mathbf{x}), \tag{6.2}$$

$$\text{subject to} \quad F(\mathbf{x}) \geqslant 0 \qquad \text{for all } \mathbf{x} \in [n]^d, \tag{6.3}$$

$$\text{and} \quad \sum_{\mathbf{x}:\, x_i = j} F(\mathbf{x}) = \lambda^i_j \qquad \text{for all } i \in [d] \text{ and } j \in [n]. \tag{6.4}$$

As before, we denote by $\mathcal{F}_\lambda$ the set of all flow arrays F satisfying the constraints (6.2) – (6.4). Geometrically, the constraint (6.4) dictates that in F, the sums inside the hyperplanes perpendicular to the ith standard basis vector must match the entries in λ^i. See Figure 8 to visualize this in the case $d = 3$ and $n = 5$.

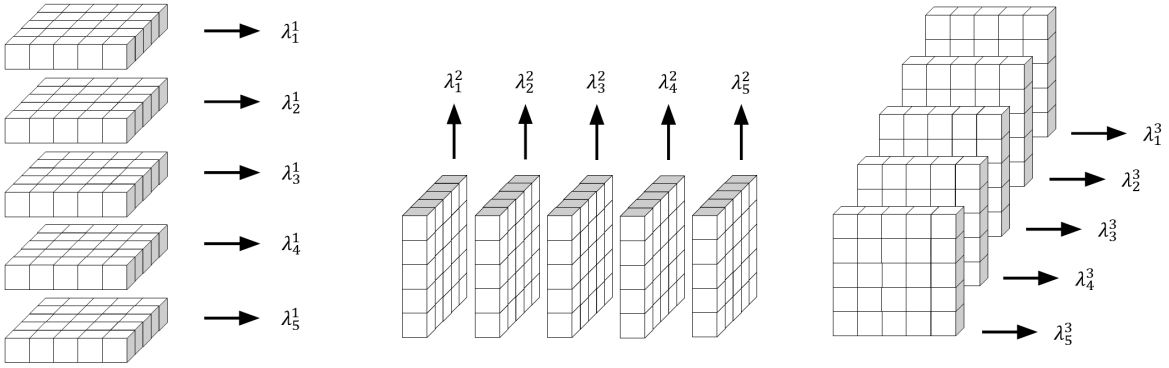


Figure 8: An illustration of the constraint (6.4) on F in the transport problem defining the EMC. (Here $d = 3$ and $n = 5$.) Each arrow represents the sum of the entries in the designated plane.

Once we have found an optimal solution $F \in \mathcal{F}_\lambda$ to the d-fold Monge problem above, the **earth mover's coefficient (EMC)** of $\boldsymbol{\lambda} = (\lambda^1, \ldots, \lambda^d)$ is defined to be the objective quantity in (6.2):

$$\mathbf{EMC}(\boldsymbol{\lambda}) := \inf_{F \in \mathcal{F}_\lambda} \left\{ \sum_{\mathbf{x} \in [n]^d} C(\mathbf{x}) F(\mathbf{x}) \right\}. \tag{6.5}$$

Remark. It is worth justifying why the taxicab distance from the main diagonal is the natural candidate to capture the cost of moving data points when comparing d histograms. The defining feature of the main diagonal is that it is the set of positions whose coordinates are all equal; hence for given $\mathbf{x} \in [n]^d$, its taxicab distance from the main diagonal is the fewest number of ± 1's we need to add in order to equalize all the coordinates x_i. For example, the most efficient way to equalize the coordinates of the position $\mathbf{x} = (5, 4, 5, 5, 5, 7, 5)$ is to add 1 to x_2, and then to subtract 2 from x_6, for a total cost of 3. But this is precisely the taxicab distance to the main diagonal, specifically to the position $(5, 5, 5, 5, 5, 5, 5)$. This distance-finding exercise corresponds explicitly to moving data points:

- When we added 1 to x_2 to make the change $4 \to 5$, we moved a data point in λ^2 from bin 4 to bin 5. (The cost is one unit of work.)
- When we subtracted 2 from x_6 to make the change $7 \to 5$, we moved a data point in λ^6 from bin 7 to bin 5. (The cost is two units of work.)

A recent paper [Kli19] defines a different cost array than ours for the earth mover's problem, namely $C'(\mathbf{x}) := x_{(d)} - x_{(1)}$. We can see from Proposition 6.1 that C' agrees with our C for $d = 2$ and $d = 3$, but not for $d > 3$. For example, letting $\mathbf{x} = (1, 1, 2, 2)$, we have $C(\mathbf{x}) = 2$ but $C'(\mathbf{x}) = 1$. In this thesis, we have chosen to define our C so that it counts *every* data-point move in the context of the EMC. For example, consider the four histograms $\lambda^1 = \lambda^2 = (1, 0)$ and $\lambda^3 = \lambda^4 = (0, 1)$. It is easy to see that an optimal flow array F is the $2 \times 2 \times 2 \times 2$ array whose only nonzero entry is a 1 at position $\mathbf{x} = (1, 1, 2, 2)$. Intuitively, we want $\mathbf{EMC}(\boldsymbol{\lambda})$ to be 2, not 1, since we must first move a data point by one bin, and then move *another* by one bin.

6.2 The Monge property, NW corner rule, and RSK in d dimensions

As we recall from Section 2.3, Hoffman [Hof63] showed that the two-histogram Monge problem can be solved by the greedy algorithm known as the northwest corner rule, if and only if the cost matrix C has the Monge property (Definition 2.3). This result was generalized by [AP89] and [BBPP95] to an arbitrary number of histograms: namely, the d-histogram Monge problem (6.2) – (6.4) can be solved by a d-dimensional analogue of the northwest corner rule if and only if the cost array C has the d-dimensional Monge property:

Definition 6.2. A d-dimensional array A has the **Monge property** if, for all $\mathbf{x} = (x_1, \ldots, x_d)$ and $\mathbf{y} = (y_1, \ldots, y_d)$, we have

$$A(\min\{x_1, y_1\}, \ldots, \min\{x_d, y_d\}) + A(\max\{x_1, y_1\}, \ldots, \max\{x_d, y_d\}) \leqslant A(\mathbf{x}) + A(\mathbf{y}).$$

Remark. If we regard an array as a function on a lattice, then the Monge property is equivalent to *submodularity*. Characterized by their "diminishing returns" property, submodular functions have found a vast expanse of modern applications, especially in machine learning (see [AMJJ18]).

Note that Definition 6.2 reduces to Definition 2.3 in the $d = 2$ case. Crucially, our particular cost array C defined in (6.1) retains the Monge property for all d; for the proof of this fact, see [Eri21, Section 8]. It is clear from the algorithm described in [BBPP95] that, just as in the two-dimensional case, the d-dimensional northwest corner rule outputs an optimal flow matrix with support on a chain, leading to the analogue of Proposition 2.5:

Proposition 6.3. *There exists an optimal $F \in \mathcal{F}_\lambda$ with entries in $\mathbb{N}$, whose support is a chain in $[n]^d$.*

Next, we generalize the RSK correspondence to allow for an input of d SSYT's; for experts, details on the existence of the multivariate generalization of RSK can be found in

[Cas09], but for our purposes we need only consider the one-row case:

$$\left\{\begin{array}{c} d\text{-tuples of SSYT's} \\ \text{with shape } (m) \\ \text{in the alphabet } [n] \end{array}\right\} \longleftrightarrow \left\{\begin{array}{c} d\text{-dimensional} \\ n \times \cdots \times n \text{ arrays} \\ \text{with entries in } \mathbb{N} \\ \text{summing to } m \\ \text{and support in a chain} \end{array}\right\}. \tag{6.6}$$

Just as in the classical construction, the chain follows from the fact that SSYT entries are weakly increasing. Given $\lambda^1, \ldots, \lambda^d \in \mathcal{H}(m,n)$, the d-tuple $(\mathrm{w}(\lambda^1), \ldots, \mathrm{w}(\lambda^d))$ is an element of the left-hand side of (6.6); from this we form the $d \times m$ matrix

$$W_{\boldsymbol{\lambda}} := \begin{bmatrix} \mathrm{w}(\lambda^1) \\ \vdots \\ \mathrm{w}(\lambda^d) \end{bmatrix}, \tag{6.7}$$

from which we construct the d-dimensional array $F_{\boldsymbol{\lambda}}$ so that $F_{\boldsymbol{\lambda}}(\mathbf{x})$ equals the number of occurrences of the column $\mathbf{x} = [x_1 \ \cdots \ x_d]^T$ in $W_{\boldsymbol{\lambda}}$. This $F_{\boldsymbol{\lambda}}$ is the RSK image of $(\mathrm{w}(\lambda^1), \ldots, \mathrm{w}(\lambda^d))$. As the reader can easily check, the coordinate hyperplane sums in $F_{\boldsymbol{\lambda}}$ are given by the λ^i, and so $F_{\boldsymbol{\lambda}}$ satisfies the constraint (6.4), meaning that $F_{\boldsymbol{\lambda}} \in \mathcal{F}_{\boldsymbol{\lambda}}$. Hence we conclude that this $F_{\boldsymbol{\lambda}}$ obtained via RSK is the unique optimal matrix described in Proposition 6.3; we now have the analogue of Lemma 3.3:

Lemma 6.4. *Let $\boldsymbol{\lambda} \in \mathcal{H}(m,n)^d$, with $W_{\boldsymbol{\lambda}} = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_m]$ defined as above, where $\mathbf{x}_j \in [n]^d$ is the jth column of $W_{\boldsymbol{\lambda}}$. Then we have*

$$\mathbf{EMC}(\boldsymbol{\lambda}) = \sum_{j=1}^{m} C(\mathbf{x}_j).$$

Example 6.5. Set $d = 4$, $m = 6$, and $n = 5$. We aim to compute the EMC of the histograms $\lambda^1, \ldots, \lambda^4 \in \mathcal{H}(6,5)$:

$$\lambda^1 = (4,1,1,0,0) \rightsquigarrow \mathrm{w}(\lambda^1) = \begin{array}{|c|c|c|c|c|c|}\hline 1&1&1&1&2&3\\\hline\end{array},$$

$$\lambda^2 = (3,0,0,0,3) \rightsquigarrow \mathrm{w}(\lambda^2) = \begin{array}{|c|c|c|c|c|c|}\hline 1&1&1&5&5&5\\\hline\end{array},$$

$$\lambda^3 = (0,4,2,0,0) \rightsquigarrow \mathrm{w}(\lambda^3) = \begin{array}{|c|c|c|c|c|c|}\hline 2&2&2&2&3&3\\\hline\end{array},$$

$$\lambda^4 = (1,1,2,1,1) \rightsquigarrow \mathrm{w}(\lambda^4) = \begin{array}{|c|c|c|c|c|c|}\hline 1&2&3&3&4&5\\\hline\end{array}.$$

The words above become the rows of the matrix W_λ:

$$W_\lambda = \begin{bmatrix} 1 & 1 & 1 & 1 & 2 & 3 \\ 1 & 1 & 1 & 5 & 5 & 5 \\ 2 & 2 & 2 & 2 & 3 & 3 \\ 1 & 2 & 3 & 3 & 4 & 5 \end{bmatrix}.$$

Following Lemma 6.4 and Proposition 6.1, we compute the cost $C(\mathbf{x}_j)$ for each of the six columns $\mathbf{x}_j$ of M:

$$C(1,1,2,1) = (2-1) + (1-1) = \mathbf{1},$$

$$C(1,1,2,2) = (2-1) + (2-1) = \mathbf{2},$$

$$C(1,1,2,3) = (3-1) + (2-1) = \mathbf{3},$$

$$C(1,5,2,3) = (5-1) + (3-2) = \mathbf{5},$$

$$C(2,5,3,4) = (5-2) + (4-3) = \mathbf{4},$$

$$C(3,5,3,5) = (5-3) + (5-3) = \mathbf{4}.$$

Finally, summing these costs, we conclude that

$$\mathbf{EMC}(\boldsymbol{\lambda}) = 1 + 2 + 3 + 5 + 4 + 4 = \mathbf{19}.$$

We conclude this section by observing that the trivariate EMC is just half the sum of the pairwise EMC's (equivalently, pairwise EMD's):

Proposition 6.6. *Let $\lambda, \mu, \nu \in \mathcal{H}(m,n)$. Then*

$$\mathbf{EMC}(\lambda,\mu,\nu) = \frac{1}{2}\Big(\mathbf{EMC}(\lambda,\mu) + \mathbf{EMC}(\lambda,\nu) + \mathbf{EMC}(\mu,\nu)\Big).$$

Proof. Fix $j \in [m]$. In the jth column of the matrix $W_{\lambda,\mu,\nu}$, call the three entries x_j, y_j, z_j, labeled so that $x_j \leqslant y_j \leqslant z_j$. Each of the three pairs (x_j, y_j), (x_j, z_j), and (y_j, z_j) corresponds naturally to one of the pairs (λ,μ), (λ,ν), and (μ,ν), if we let rows $1,2,3$ correspond to

λ, μ, ν respectively. Therefore by Lemmas 3.3 and 6.4, we have

$$\begin{aligned}
\mathbf{EMC}(\lambda,\mu) + \mathbf{EMC}(\lambda,\nu) + \mathbf{EMC}(\mu,\nu) &= \sum_{j=1}^{m} C(x_j,y_j) + C(x_j,z_j) + C(y_j,z_j) \\
&= \sum_j (y_j - x_j) + (z_j - x_j) + (z_j - y_j) \\
&= \sum_j 2z_j - 2x_j \\
&= 2\sum_j z_j - x_j \\
&= 2\sum_j C(x_j,y_j,z_j) \\
&= 2 \cdot \mathbf{EMC}(\lambda,\mu,\nu). \qquad \square
\end{aligned}$$

This relationship does not generalize to $d > 3$, because in general, the telescoping summand in the proof does not reduce in terms of a higher-dimensional cost function. For example, when $d = 4$, the analogue of the third line above is $\sum_j 3w_j + z_j - y_j - 3x_j$, or $\sum_j C(x_j,y_j,z_j,w_j) + 2(w_j - x_j)$.

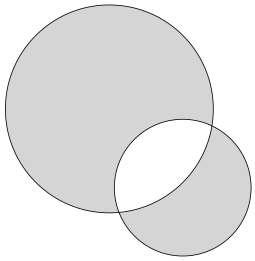
6.3 A generalization of the symmetric difference

Typically in the literature, the definition of symmetric difference $\triangle$ is extended associatively to any finite number of sets; in this way, the symmetric difference contains those elements that are contained in an odd number of the original sets. In order to capture the behavior of the EMC, however, we introduce a different generalization:

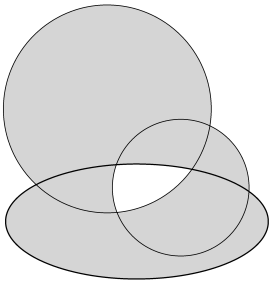
Definition 6.7. We say the **unimodal symmetric difference** of sets $X_1, \ldots, X_d$ is

$$\blacktriangle(X_1,\ldots,X_d) \coloneqq \sum_{k=0}^{d} \min\{k,\, d-k\} \cdot \#\{\text{elements contained in exactly } k \text{ of the } X_i\}.$$

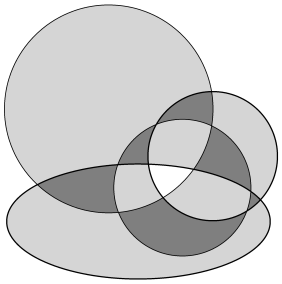
Note that $\blacktriangle$ returns a natural number rather than a set. Intuitively, the unimodal symmetric difference counts elements by assigning the weight 1 to those elements contained



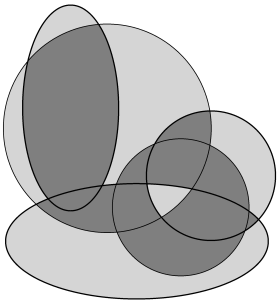
For $d = 2$, weights are $0, 1, 0$.



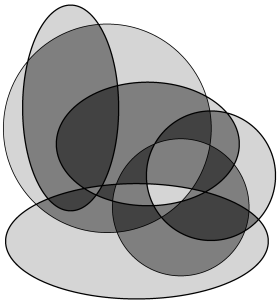
For $d = 3$, weights are $0, 1, 1, 0$.



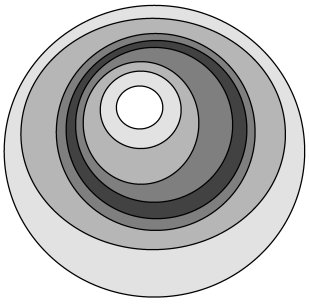
For $d = 4$, weights are $0, 1, 2, 1, 0$.



For $d = 5$, weights are $0, 1, 2, 2, 1, 0$. Note that in this example, the intersection of all five sets is empty.



For $d = 6$, weights are $0, 1, 2, 3, 2, 1, 0$.



For $d = 8$, weights are $0, 1, 2, 3, 4, 3, 2, 1, 0$. This image depicts a chain of eight subsets.

Figure 9: A visualization of the unimodal symmetric difference $\blacktriangle(X_1, \ldots, X_d)$.

in exactly one, or all but one, set; then the weight 2 to those elements in exactly two, or all but two, sets; and so forth, finally assigning the greatest weight $\lfloor d/2 \rfloor$ to those elements contained in exactly "half" of the sets. In each subfigure of Figure 9, we provide a visualization of this, where each ellipse is one of the sets X_i, and the weights $0, 1, \ldots, \lfloor d/2 \rfloor$ are depicted by successively darker shades of grey. Elements contained in 0 or d sets are not counted (white regions); elements contained in exactly 1 or $d - 1$ sets are counted once (lightest grey); elements in exactly 2 or $d - 2$ sets are counted twice (slightly darker grey); and so forth, until elements in the darkest regions are counted $\lfloor d/2 \rfloor$ times. Note that if

the X_i are all subsets of a set U, then $\blacktriangle$ is invariant under complements:

$$\begin{aligned}\blacktriangle(X_1', \ldots, X_d') &= \sum_{k=0}^{d} \min\{k, d-k\} \cdot \#\{\text{elements contained in exactly } k \text{ of the } X_i'\} \\ &= \sum_{k=0}^{d} \min\{k, d-k\} \cdot \#\{\text{elements contained in exactly } d-k \text{ of the } X_i\} \\ &= \blacktriangle(X_1, \ldots, X_d).\end{aligned}$$

We also observe that $\blacktriangle(X, Y) = |X \triangle Y|$, and hence the proposition below reduces to Proposition 3.4 in the $d = 2$ case.

Proposition 6.8. *For $\boldsymbol{\lambda} \in \mathcal{H}(m, n)^d$, we have* $\mathbf{EMC}(\boldsymbol{\lambda}) = \blacktriangle\big(\mathrm{Y}(\lambda^1), \ldots, \mathrm{Y}(\lambda^d)\big)$.

Proof. By Lemma 6.4, we have $\mathbf{EMC}(\boldsymbol{\lambda}) = \sum_{j=1}^{m} C(\mathbf{x}_j)$, where $\mathbf{x}_j$ is the jth column of the matrix $W_{\boldsymbol{\lambda}}$ defined in (6.7). Fix $j \in [m]$, and let $\mathbf{x}_j = (x_1, \ldots, x_d)$. Then by Proposition 6.1, we have

$$\begin{aligned}C(\mathbf{x}_j) &= \sum_{i=1}^{\lfloor d/2 \rfloor} x_{(d+1-i)} - x_{(i)} \\ &= \sum_{i=1}^{\lfloor d/2 \rfloor} \sum_{k=i}^{d-i} x_{(k+1)} - x_{(k)}.\end{aligned}$$

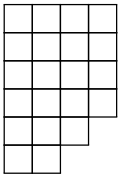
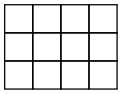
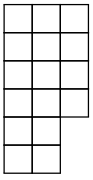
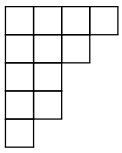
Now we switch the order of summation, using the fact that $i \leqslant k \leqslant d-i$ implies $i \leqslant \min\{k, d-k\}$:

$$\begin{aligned}C(\mathbf{x}_j) &= \sum_{k=1}^{d-1} \left(\sum_{i=1}^{\min\{k, d-k\}} x_{(k+1)} - x_{(k)} \right) \\ &= \sum_{k=1}^{d-1} \min\{k,\ d-k\} \cdot \big(x_{(k+1)} - x_{(k)}\big) \\ &= \blacktriangle\Big([x_1], \ldots, [x_d]\Big),\end{aligned}$$

where as usual $[x_i]$ denotes the set $\{1, \ldots, x_i\}$. Now recalling the construction of the matrix $W_{\boldsymbol{\lambda}}$, we know that x_i is the jth element from the word $\mathrm{w}(\lambda^i)$, and therefore $x_i' \coloneqq n - x_i$ is the length of the jth row of $\mathrm{Y}(\lambda^i)$. Hence $\blacktriangle([x_1], \ldots, [x_d]) = \blacktriangle([x_1'], \ldots, [x_d'])$ is the result

of applying ▲ to the collection containing the jth row from each of the Young diagrams $\mathrm{Y}(\lambda^1), \ldots, \mathrm{Y}(\lambda^d)$. Letting j range over all m rows and taking the sum, we conclude that $\mathbf{EMC}(\boldsymbol{\lambda}) = \blacktriangle\Big(\mathrm{Y}(\lambda^1), \ldots, \mathrm{Y}(\lambda^d)\Big)$. □

Example 6.9. We revisit Example 6.5, but this time we will calculate the EMC by using the symmetric difference from Proposition 6.8. We translate the complementary word $\mathrm{w}'(\lambda^i)$ of each histogram into its corresponding Young diagram:

$\mathrm{Y}(\lambda^1) =$  $\mathrm{Y}(\lambda^2) =$  $\mathrm{Y}(\lambda^3) =$  $\mathrm{Y}(\lambda^4) =$ 

Now we superimpose these four diagrams so that they share a common upper-left cell; see the two diagrams below in (6.8). On the left-hand side of (6.8), we label each cell with the number of diagrams $\mathrm{Y}(\lambda^i)$ that contain it (obtaining a plane partition); then on the right-hand side, to each cell labeled k we assign the weight $\min\{k,\, 4-k\}$, since $d = 4$, and shade as in Figure 9:

(6.8)

The shading is superfluous, of course, since we only need to sum the entries of the right-hand diagram to compute that $\mathbf{EMC}(\boldsymbol{\lambda}) = \blacktriangle\Big(\mathrm{Y}(\lambda^1), \ldots, \mathrm{Y}(\lambda^4)\Big) = \mathbf{19}$, agreeing with the calculation in Example 6.5.

6.4 Expected value of the EMC

This section contains the main result (Theorem 6.12) of the author's paper [Eri21], extending to arbitrary values of d the methods used in [BW20] for $d = 2$. First we will define a generating function, in which the exponents encode the values of **EMC**, and which we will differentiate in order to sum up all of these values. This will allow us to compute expected value for the **EMC** by reading off the coefficients of this derivative.

6.4.1 A generating function for the discrete setting

Just as in Section 3.4, because we are about to make a recursive definition, we must momentarily consider d-tuples $\boldsymbol{\lambda}$ consisting of histograms with different numbers of bins — i.e., different values n_i such that each $\lambda^i \in \mathcal{H}(m, n_i)$. Therefore, $\mathbf{n} = (n_1, \ldots, n_d)$ will denote this vector of bin numbers. Note that the EMC is still defined on $\mathcal{H}(m, n_1) \times \cdots \times \mathcal{H}(m, n_d)$, since we can append zeros to the end of each n_i-tuple as needed; in this way, we regard the histograms as sharing a common bin number $\max\{n_1, \ldots, n_d\}$. We do not, however, consider any statistical meaning for the EMC on these padded histograms; they are a formal means to an end, and our main result (Theorem 6.12) ultimately assumes equal bin numbers.

In order to encode an inclusion–exclusion argument, we define an indicator vector $\mathbf{e}_A$ for a subset $A \subseteq [d]$. Namely,

$$\mathbf{e}_A := \sum_{i \in A} \mathbf{e}_i$$

is the vector whose ith component is 1 if $i \in A$ and 0 otherwise. For example, if $d = 5$, and $A = \{2, 4, 5\}$, then $\mathbf{e}_A = (0, 1, 0, 1, 1)$.

For fixed $\mathbf{n} = (n_1, \ldots, n_d)$, we define a generating function in two indeterminates q and z:

$$H_{\mathbf{n}}(q, z) := \sum_{m=0}^{\infty} \left(\sum_{\boldsymbol{\lambda} \in \mathcal{H}(m,n_1) \times \cdots \times \mathcal{H}(m,n_d)} q^{\mathbf{EMC}(\boldsymbol{\lambda})} \right) z^m. \tag{6.9}$$

The coefficient of $q^r z^m$ is the number of elements $\boldsymbol{\lambda} \in \mathcal{H}(m, n_1) \times \cdots \times \mathcal{H}(m, n_d)$ such that $\mathbf{EMC}(\boldsymbol{\lambda}) = r$.

Proposition 6.10. *Fix $\mathbf{n} = (n_1, \ldots, n_d)$. The generating function $H_{\mathbf{n}} := H_{\mathbf{n}}(q, z)$ has the following recursive formula, where the sum is over all nonempty subsets $A \subseteq [d]$:*

$$H_{\mathbf{n}} = \frac{\sum_A (-1)^{|A|-1} \cdot H_{\mathbf{n}-\mathbf{e}_A}}{1 - q^{C(\mathbf{n})} z},$$

where $H_{(1,\ldots,1)} = \frac{1}{1-z}$, and $H_{\mathbf{n}-\mathbf{e}_A} = 0$ if $\mathbf{n} - \mathbf{e}_A$ contains a 0.

Proof. Each $\boldsymbol{\lambda} \in \mathcal{H}(m, n_1) \times \cdots \times \mathcal{H}(m, n_d)$ corresponds to a unique monomial

$$\boldsymbol{\lambda} \longleftrightarrow \prod_{\mathbf{x}} w_{\mathbf{x}}^{F_\lambda(\mathbf{x})}, \tag{6.10}$$

where $\boldsymbol{\lambda} \leftrightarrow F_\lambda$ is the generalized RSK correspondence. The variables $w_{\mathbf{x}}$ are indexed by multi-indices $\mathbf{x} \in [n_1] \times \cdots \times [n_d]$. Note that the degree of the monomial (6.10) equals m (the sum of the entries of F_λ). Now making the substitution

$$w_{\mathbf{x}} \longmapsto q^{C(\mathbf{x})} z, \tag{6.11}$$

the correspondence (6.10) gives us the maps

$$\boldsymbol{\lambda} \longleftrightarrow \prod_{\mathbf{x}} w_{\mathbf{x}}^{F_\lambda(\mathbf{x})} \longmapsto q^{\mathbf{EMC}(\lambda)} z^m.$$

Therefore the generating function $H_{\mathbf{n}}$ is just the image of the formal sum $H_{\mathbf{n}}^*$ of all monomials of the form (6.10), under the substitution (6.11); as m ranges over $\mathbb{N}$, there is one monomial in $H_{\mathbf{n}}^*$ for each possible $\boldsymbol{\lambda} \in \mathcal{H}(m, n_1) \times \cdots \times \mathcal{H}(m, n_d)$.

Since $\mathbf{x} \preceq \mathbf{n}$ for all $\mathbf{x}$, every RSK array F_λ is allowed to contain $\mathbf{n}$ in its support without violating the chain condition. This means that every monomial in $H_{\mathbf{n}}^*$ is allowed to contain the variable $w_{\mathbf{n}}$, and so we may factor out the sum of all possible powers of $w_{\mathbf{n}}$, rewriting as

$$H_{\mathbf{n}}^* = \sum_{\lambda} \left(\prod_{\mathbf{x}} w_{\mathbf{x}}^{F_\lambda(\mathbf{x})} \right) = \sum_{r=0}^{\infty} w_{\mathbf{n}}^r \cdot h(\mathbf{w}_{\mathbf{x} \neq \mathbf{n}}) = \frac{h(\mathbf{w}_{\mathbf{x} \neq \mathbf{n}})}{1 - w_{\mathbf{n}}},$$

where h is an infinite formal sum of monomials in the variables $w_{\mathbf{x}}$ where $\mathbf{x} \neq \mathbf{n}$. Now we focus on rewriting this numerator h. Suppose we subtract 1 from exactly one of the coordinates of $\mathbf{n}$; the possible results are $\mathbf{n} - \mathbf{e}_i$ for $i = 1, \ldots, d$. Now, on one hand, any monomial in h containing the variable $w_{\mathbf{n}-\mathbf{e}_i}$ appears in $H_{\mathbf{n}-\mathbf{e}_i}^*$. But on the other hand, note that all of these $\mathbf{n} - \mathbf{e}_i$ are mutually incomparable under the product order, and so at most *one* of them can be in the support of some F_λ, because of the chain condition. Therefore any monomial in h contains at most *one* of the variables $w_{\mathbf{n}-\mathbf{e}_i}$. But the sum $\sum_{i=1}^{d} H_{\mathbf{n}-\mathbf{e}_i}^*$ still overcounts the monomials appearing in h, since the same monomial may appear in several distinct summands.

In other words, we want h to be the formal sum of the *union* (without multiplicity) of the monomials which appear in the summands $H^*_{\mathbf{n}-\mathbf{e}_i}$. We can achieve this by using the inclusion–exclusion principle: subtract those monomials which appear in at least 2 of the summands, then add back the monomials which appear in at least 3 of the summands, then subtract those appearing in at least 4 summands, and so on, until we arrive at those monomials appearing in all d of the summands. We can write this inclusion–exclusion as an alternating sum over nonempty subsets $A \subseteq [d]$, adding when $|A|$ is odd and subtracting when $|A|$ is even:

$$h = \sum_A (-1)^{|A|-1} \cdot H^*_{\mathbf{n}-\mathbf{e}_A}.$$

Finally, applying the substitution (6.11), we obtain

$$\begin{aligned} H_{\mathbf{n}} = H^*_{\mathbf{n}}\Big|_{w_{\mathbf{x}}=q^{C(\mathbf{m})}z} &= \left.\frac{\sum_A (-1)^{|A|-1} \cdot H^*_{\mathbf{n}-\mathbf{e}_A}}{1-w_{\mathbf{n}}}\right|_{w_{\mathbf{x}}=q^{C(\mathbf{x})}z} \\ &= \frac{\sum_A (-1)^{|A|-1} \cdot H_{\mathbf{n}-\mathbf{e}_A}}{1-q^{C(\mathbf{n})}z}, \end{aligned}$$

proving the recursion.

As for the base case $H_{(1,\ldots,1)} = \frac{1}{1-z}$, there is only one element in $\mathcal{H}(m,1)$, and so since every $n_i = 1$, the inside sum in (6.9) has only one term; moreover, this unique $\boldsymbol{\lambda}$ is just d copies of the same trivial histogram with mass m and 1 bin, meaning that $\mathbf{EMC}(\boldsymbol{\lambda}) = 0$. Hence $H_{(1,\ldots,1)}(q,z) = \sum_m q^0 z^m = \sum_m z^m = \frac{1}{1-z}$. Likewise, since $\mathcal{H}(m,0)$ is empty, we must have $H_{\mathbf{n}} = 0$ whenever any of the n_i are 0. □

Example 6.11. We write out this recursive definition in a concrete case, where $d = 3$ and $\mathbf{n} = (5,2,2)$. It is easiest to arrange the terms of the numerator according to the size of the subset A. First, for $|A| = 1$, we add together all possible $H_{\mathbf{n}'}$ such that $\mathbf{n}'$ equals $\mathbf{n}$ with exactly 1 coordinate decreased; then for $|A| = 2$, we *subtract* all possible $H_{\mathbf{n}''}$ such that $\mathbf{n}''$ equals $\mathbf{n}$ with exactly 2 coordinates decreased; finally, for $|A| = 3$, we *add* the one possible $H_{\mathbf{n}'''}$ where $\mathbf{n}'''$ equals $\mathbf{n}$ with all 3 coordinates decreased. As for the denominator, c is the same cost function we defined in (6.1), meaning that $C(5,2,2) = 5-2 = 3$. Then the

recursion for $H_{\mathbf{n}}$ looks like this:

$$H_{(5,2,2)} = \frac{H_{(4,2,2)} + H_{(5,1,2)} + H_{(5,2,1)} - H_{(4,1,2)} - H_{(4,2,1)} - H_{(5,1,1)} + H_{(4,1,1)}}{1 - q^3 z.}$$

Having seen an example, we now study the important (and very well-studied) specialization that results from setting $q = 1$. In this case, the coefficient of z^m in $H_{\mathbf{n}}(1, z)$ is the total number of d-tuples $\boldsymbol{\lambda}$, which is $\prod_{i=1}^{d} \#\mathcal{H}(m, n_i) = \prod_{i=1}^{d} \binom{m+n_i-1}{n_i-1}$:

$$H_{\mathbf{n}}(1, z) = \sum_{m=0}^{\infty} \prod_{i=1}^{d} \binom{m + n_i - 1}{n_i - 1} z^m. \tag{6.12}$$

It is shown in [DR69] that the closed form of (6.12), after adjusting the index to match our setup, and writing $|\mathbf{n}| := n_1 + \cdots + n_d$, is

$$H_{\mathbf{n}}(1, z) = \frac{Q_{\mathbf{n}}(z)}{(1 - z)^{|\mathbf{n}|-d+1}}, \tag{6.13}$$

where the numerator $Q_{\mathbf{n}}(z)$ is a polynomial whose coefficients are the "Simon Newcomb" numbers. (For more on this natural generalization of Eulerian numbers to multisets, see [Abr75], [DR69], and [Mor13].) Specifically, denoting the coefficient of z^i in $Q_{\mathbf{n}}$ by the symbol $[z^i]Q_{\mathbf{n}}$, we have

$$[z^i]Q_{\mathbf{n}} = \text{ \# permutations of the multiset } \{1^{n_1-1}, \ldots, d^{n_d-1}\} \text{ containing } i \text{ descents.}$$

It follows that the evaluation $Q_{\mathbf{n}}(1)$ equals the total number of permutations of the multiset $\{1^{n_1-1}, \ldots, d^{n_d-1}\}$, a fact we will need later:

$$Q_{\mathbf{n}}(1) = \frac{\left(\sum_{i=1}^{d}(n_i - 1)\right)!}{\prod_{i=1}^{d}(n_i - 1)!} = \frac{(|\mathbf{n}| - d)!}{\prod(n_i - 1)!} \tag{6.14}$$

6.4.2 A partial derivative

Next, in order to transfer the EMC values from the exponents of q into coefficients, we take the partial derivative of $H_{\mathbf{n}}$ with respect to q. Applying the quotient rule to our

definition of $H_{\mathbf{n}}$ in Proposition 6.10, we obtain the following, where the sum still ranges over nonempty subsets $A \subseteq [d]$:

$$\frac{\partial H_{\mathbf{n}}}{\partial q} = \frac{\left(1 - q^{C(\mathbf{n})}z\right)\left(\sum_A (-1)^{|A|-1} \cdot \frac{\partial H_{\mathbf{n}-\mathbf{e}_A}}{\partial q}\right) + C(\mathbf{n}) \cdot q^{C(\mathbf{n})-1} \cdot z \cdot \left(\sum_A (-1)^{|A|-1} \cdot H_{\mathbf{n}-\mathbf{e}_A}\right)}{\left(1 - q^{C(\mathbf{n})}z\right)^2}$$

Now that the exponents have been changed into coefficients of z, we set $q = 1$:

$$\begin{aligned} H'_{\mathbf{n}} := \left.\frac{\partial H_{\mathbf{n}}}{\partial q}\right|_{q=1} &= \sum_{m=0}^{\infty} \left(\sum_{\boldsymbol{\lambda} \in \mathcal{H}(m,n_1) \times \cdots \times \mathcal{H}(m,n_d)} \mathbf{EMC}(\boldsymbol{\lambda}) \right) z^m \\ &= \frac{(1-z)\left(\sum_A (-1)^{|A|-1} \cdot H'_{\mathbf{n}-\mathbf{e}_A}\right) + z \cdot C(\mathbf{n}) \left(\sum_A (-1)^{|A|-1} \cdot H_{\mathbf{n}-\mathbf{e}_A}\right)}{(1-z)^2} \end{aligned} \tag{6.15}$$

At this point, q has played out its role, and so from now on we will write $H_{\mathbf{n}}$ in place of $H_{\mathbf{n}}(1, z)$.

Our goal is now in sight: to find the expected value of EMC for fixed $\mathbf{n}$, we need to divide the sum of all possible values of $\mathbf{EMC}(\boldsymbol{\lambda})$ (i.e., the coefficient of z^m in $H'_{\mathbf{n}}$) by the total number of possible inputs $\boldsymbol{\lambda}$ (i.e., the coefficient of z^m in $H_{\mathbf{n}}$). Therefore, once we find a way to simplify (6.15), we will be able to compute the result

$$\mathbb{E}\left(\mathbf{EMC}(\boldsymbol{\lambda}) \,\middle|\, \boldsymbol{\lambda} \in \mathcal{H}(m,n_1) \times \cdots \times \mathcal{H}(m,n_d)\right) = \frac{[z^m]H'_{\mathbf{n}}}{[z^m]H_{\mathbf{n}}} = \frac{[z^m]H'_{\mathbf{n}}}{\prod_{i=1}^{d} \binom{m+n_i-1}{n_i-1}}. \tag{6.16}$$

In order to make the expression (6.15) for $H'_{\mathbf{n}}$ more tractable to program, we will now focus only on the numerators of $H_{\mathbf{n}}$ and $H'_{\mathbf{n}}$. We have already defined the numerator of $H_{\mathbf{n}}$ as $Q_{\mathbf{n}}(t)$ in the previous section. We will let $N_{\mathbf{n}}(t)$ denote the numerator of $H'_{\mathbf{n}}$. By using Mathematica to observe patterns for small $\mathbf{n}$, we anticipate that the denominator of $H'_{\mathbf{n}}$ has exponent $|\mathbf{n}| - d + 2$, and so we set both

$$Q_{\mathbf{n}}(z) := (1-z)^{|\mathbf{n}|-d+1} H_{\mathbf{n}} \qquad \text{and} \qquad N_{\mathbf{n}}(z) := (1-z)^{|\mathbf{n}|-d+2} H'_{\mathbf{n}}. \tag{6.17}$$

Therefore, we can clear denominators in (6.15) by multiplying both sides by $(1-z)^{|\mathbf{n}|-d+2}$.

Proceeding carefully and clearing the remaining denominators using (6.17), we obtain

$$\begin{aligned} N_{\mathbf{n}} &= \sum_A (-1)^{|A|-1}(1-z)^{|A|-1} N_{\mathbf{n}-\mathbf{e}_A} + z \cdot C(\mathbf{n}) \cdot (1-z)^{|\mathbf{n}|-d} \cdot (1-z) \cdot H_{\mathbf{n}} \\ &= \sum_A (z-1)^{|A|-1} N_{\mathbf{n}-\mathbf{e}_A} + z \cdot C(\mathbf{n}) \cdot Q_{\mathbf{n}}. \end{aligned} \tag{6.18}$$

This provides us with a quick recursive code to obtain $N_{\mathbf{n}}(z)$, after which we need only divide by $(1-z)^{|\mathbf{n}|-d+2}$ to recover $H'_{\mathbf{n}}$. The rest is just a matter of extracting coefficients in order to apply the result in (6.16).

6.4.3 Expected value of EMC for probability distributions on $[n]$

We now shift our perspective from histograms to probability distributions. Let $\mathcal{P}(n)$ denote the set of probability distributions on $[n]$. We assume the uniform probability measure on the d-fold product $\mathcal{P}(n)^d$, defined by its embedding into $\mathbb{R}^{nd}$. The Monge problem (6.2) – (6.4) still makes sense when $\lambda_1, \ldots, \lambda_d \in \mathcal{P}(n)$, and therefore the definition of the EMC is still meaningful. Our goal is to write down a formula for the expected value of the EMC restricted to the domain $\mathcal{P}(n)^d$.

Starting with the expected value of EMC from (6.16), we scale by $1/m$ to normalize, and then take the limit as $m \to \infty$:

$$\mathcal{E}_{\mathbf{n}} := \mathbb{E}\left(\mathbf{EMC}(\boldsymbol{\lambda}) \,\middle|\, \boldsymbol{\lambda} \in \mathcal{P}(n_1) \times \cdots \times \mathcal{P}(n_d)\right) = \lim_{m\to\infty} \frac{1}{m} \cdot \frac{[z^m]H'_{\mathbf{n}}}{\prod_{i=1}^d \binom{m+n_i-1}{n_i-1}}.$$

First we focus on the $[z^m]H'_{\mathbf{n}}$ part, namely the coefficient of z^m in $H'_{\mathbf{n}} = \frac{N_{\mathbf{n}}(z)}{(1-z)^{|\mathbf{n}|-d+2}}$. It is a standard fact that the coefficient of z^m in the series $\frac{1}{(1-z)^{|\mathbf{n}|-d+2}}$ is just

$$\binom{m+|\mathbf{n}|-d+1}{|\mathbf{n}|-d+1} = \frac{m^{|\mathbf{n}|-d+1}}{(|\mathbf{n}|-d+1)!} + \text{ lower-order terms in } m.$$

Meanwhile, $N_{\mathbf{n}}(z)$ is a polynomial, with some finite degree b. Now, as $m \to \infty$, we also have $m - b \to \infty$, and so the coefficient of z^m in $H'_{\mathbf{n}}$ is asymptotic to $\frac{m^{|\mathbf{n}|-d+1}}{(|\mathbf{n}|-d+1)!}$ multiplied by the sum of the coefficients of $N_{\mathbf{n}}(z)$. But this sum is just $N_{\mathbf{n}}(1)$, and so we have

$$[z^m]H'_{\mathbf{n}} \sim N_{\mathbf{n}}(1) \cdot \frac{m^{|\mathbf{n}|-d+1}}{(|\mathbf{n}|-d+1)!}.$$

Accounting for the $1/m$, we currently have the following:

$$\mathcal{E}_{\mathbf{n}} = \lim_{m\to\infty} N_{\mathbf{n}}(1) \cdot \frac{m^{|\mathbf{n}|-d}}{(|\mathbf{n}|-d+1)! \prod_{i=1}^{d} \binom{m+n_i-1}{n_i-1}}.$$

Since $\prod_i \binom{m+n_i-1}{n_i-1} \sim \prod_i \frac{m^{n_i-1}}{(n_i-1)!} = \frac{m^{|\mathbf{n}|-d}}{\prod_i (n_i-1)!}$, this becomes

$$\mathcal{E}_{\mathbf{n}} = N_{\mathbf{n}}(1) \cdot \frac{\prod_{i=1}^{d}(n_i-1)!}{(|\mathbf{n}|-d+1)!}. \tag{6.19}$$

But when we evaluate $N_{\mathbf{n}}(1)$ from equation (6.18), the terms with $(z-1)$ all disappear; hence we need only consider singletons $A \subseteq [d]$, meaning we are now summing from 1 to d:

$$N_{\mathbf{n}}(1) = \sum_{i=1}^{d} N_{\mathbf{n}-\mathbf{e}_i}(1) + C(\mathbf{n}) Q_{\mathbf{n}}(1).$$

Substituting for $Q_{\mathbf{n}}(1)$ using (6.14), we have

$$N_{\mathbf{n}}(1) = \sum_{i=1}^{d} N_{\mathbf{n}-\mathbf{e}_i}(1) + \frac{C(\mathbf{n}) \cdot (|\mathbf{n}|-d)!}{\prod_{i=1}^{d}(n_i-1)!}$$

Finally, returning to (6.19) and substituting the above for $N_{\mathbf{n}}(1)$, we obtain the recursive definition

$$\begin{aligned}
\mathcal{E}_{\mathbf{n}} &= \left[\sum_{i=1}^{d} N_{\mathbf{n}-\mathbf{e}_i}(1) + \frac{C(\mathbf{n}) \cdot (|\mathbf{n}|-d)!}{\prod_{i=1}^{d}(n_i-1)!} \right] \cdot \frac{\prod_{i=1}^{d}(n_i-1)!}{(|\mathbf{n}|-d+1)!} \\
&= \frac{\sum_{i=1}^{d} N_{\mathbf{n}-\mathbf{e}_i}(1) \cdot \frac{\prod_{i=1}^{d}(n_i-1)!}{(|\mathbf{n}|-d)!}}{|\mathbf{n}|-d+1} + \frac{C(\mathbf{n}) \cdot (|\mathbf{n}|-d)!}{(|\mathbf{n}|-d+1)!} \\
&= \frac{\sum_{i=1}^{d}(n_i-1)\mathcal{E}_{\mathbf{n}-\mathbf{e}_i} + C(\mathbf{n})}{|\mathbf{n}|-d+1},
\end{aligned} \tag{6.20}$$

where $\mathcal{E}_{\mathbf{n}-\mathbf{e}_i} = 0$ if $\mathbf{n} - \mathbf{e}_i$ contains a 0.

We record this as the main theorem of Section 6.4. We state the result only in the case of equal bin numbers, writing $(n^d) = (n, \ldots, n)$, so that the EMC retains its statistical meaning.

Theorem 6.12. *The expected value of* **EMC** *on* $\mathcal{P}(n)^d$ *is given by* $\mathcal{E}_{(n^d)}$ *as defined in* (6.20).

Remark. Recall from Proposition 6.6 the special relationship between the $d = 2$ and $d = 3$ cases for the EMC of histograms (namely, the EMC of three histograms equals half the sum of their pairwise EMC's). This leads us to anticipate that $\mathcal{E}_{(n^3)} = \frac{3}{2} \cdot \mathcal{E}_{(n^2)}$. We confirm this in Mathematica:

n	$\mathcal{E}_{(n^2)}$	$\mathcal{E}_{(n^3)}$	$\mathcal{E}_{(n^3)}/\mathcal{E}_{(n^2)}$
2	0.3333	0.5000	1.5
3	0.5333	0.8000	1.5
4	0.6857	1.0286	1.5
5	0.8127	1.2191	1.5
6	0.9235	1.3853	1.5
7	1.0230	1.5345	1.5
8	1.1139	1.6709	1.5
9	1.1982	1.7972	1.5
10	1.2770	1.9155	1.5

6.4.4 The effect of the parity of d

It is convenient to unit normalize the value of the EMC so that its value falls between 0 and 1. Observe that the maximum value of the discrete EMC occurs when, in every column of the matrix W_λ, half the entries are 1's and the other half are n's; if d is odd, then "half" means $\lfloor d/2 \rfloor$, with the leftover entry being irrelevant by Proposition 6.1. For such a λ, then, $\mathbf{EMC}(\lambda)$ equals the cost $\lfloor d/2 \rfloor (n-1)$ multiplied by m (the number of columns). After dividing by m to pass to probability distributions, we see that the maximum value of **EMC** on $\mathcal{P}(n)^d$ is $\lfloor d/2 \rfloor (n-1)$. With this in mind, we define the **unit normalized EMC** and its expected value:

$$\widetilde{\mathbf{EMC}}(\lambda) := \frac{\mathbf{EMC}(\lambda)}{\lfloor d/2 \rfloor (n-1)} \qquad \text{and} \qquad \widetilde{\mathcal{E}}_{(n^d)} := \frac{\mathcal{E}_{(n^d)}}{\lfloor d/2 \rfloor (n-1).} \tag{6.21}$$

Using Theorem 6.12 and the definitions in (6.21), we record the unit normalized expected value $\widetilde{\mathcal{E}}_{(n^d)}$ in Table 4, for the first few values of n and d.

We observe a curious phenomenon in Table 4 when we fix n and let d increase: the unit normalized expected value alternately increases (d changing from even to odd) and

n \ d	2	3	4	5	6	7	8	9	10
2	0.3333	0.5000	0.4000	0.5000	0.4286	0.5000	0.4444	0.5000	0.4545
3	0.2667	0.4000	0.3175	0.3968	0.3388	0.3952	0.3505	0.3943	0.3579
4	0.2286	0.3429	0.2711	0.3389	0.2888	0.3370	0.2985	0.3358	0.3046
5	0.2032	0.3048	0.2405	0.3006	0.2560	0.2986	0.2644	0.2974	0.2697
6	0.1847	0.2771	0.2184	0.2730	0.2322	0.2710	0.2398	0.2698	0.2445
7	0.1705	0.2557	0.2014	0.2517	0.2141	0.2498	0.2210	0.2486	0.2253
8	0.1591	0.2387	0.1878	0.2348	0.1996	0.2329	0.2060	0.2318	0.2101
9	0.1498	0.2247	0.1767	0.2209	0.1877	0.2190	0.1937	0.2180	0.1975
10	0.1419	0.2128	0.1673	0.2092	0.1778	0.2074	0.1834	0.2064	0.1870

Table 4: Selected values of the unit normalized expected value $\widetilde{\mathcal{E}}_{(n)^d}$.

decreases (d changing from odd to even). This suggests that an even number of probability distributions are more likely to be "closer" together (in the sense of the EMC) than an odd number of distributions. We can explain this dependence on parity from both a geometric and statistical perspective.

Geometrically, in an odd number of dimensions, the taxicab distance from a point to the main diagonal remains unchanged as we vary the median of the point's coordinates, and the variability of the median increases as the distance increases (i.e., as the other coordinates become more spread out). For example, in a 3-dimensional array with side lengths n, the greatest possible distance from the main diagonal is $\lfloor 3/2 \rfloor (n-1) = n-1$; this is attained by all those positions whose coordinates (up to reordering) are $(1, \_\_, n)$, where the blank can take any value in $[n]$. In a 4-dimensional array, however, there is a much smaller proportion of positions at maximal distance $\lfloor 4/2 \rfloor (n-1) = 2(n-1)$ from the main diagonal: namely, those positions whose coordinates (up to reordering) are $(1, 1, n, n)$. In effect, the free "odd-man-out" coordinate does not exist in an even number of dimensions, so there is a smaller proportion of array positions farther away from the main diagonal compared to an odd number of dimensions. This remains true when we restrict to arrays with support in a chain, and so we expect to see the even–odd disparity reflected in the EMC data.

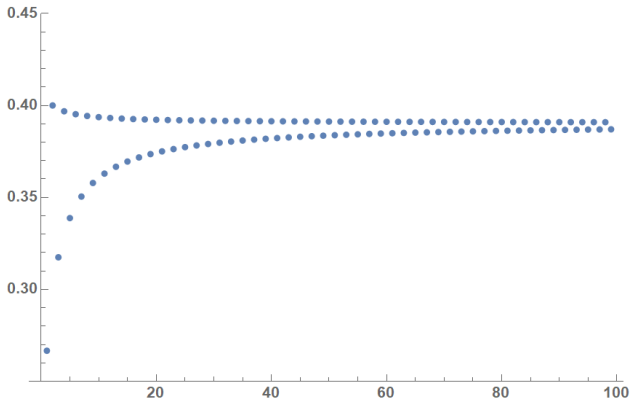


Figure 10: A plot of $\widetilde{\mathcal{E}}_{(3^d)}$ (vertical axis), for values of d up to 100 (horizontal axis).

Likewise, statistically, three probability distributions achieve maximum EMC when one of them is $(1, 0, \ldots, 0)$, another is $(0, \ldots, 0, 1)$, and the third is *anything* at all. But the maximum EMC for *four* probability distributions is attained only when two distributions are $(1, 0, \ldots, 0)$ and the other two are $(0, \ldots, 0, 1)$. For d odd, the variability in the free "odd-man-out" distribution is greater for higher EMC values, and so it is no surprise that there is a greater proportion of higher EMC values when d is odd.

Although this even–odd disparity is significant when d is relatively small (as in Table 4), the values of $\widetilde{\mathcal{E}}_{(n^d)}$ do stabilize fairly quickly. We plot these values in Figure 10, fixing $n = 3$ and letting d grow from 2 to 100 on the horizontal axis. On the far left side of the plot, the disparity between even and odd values of d is indeed striking — and the variability is much greater when d is even than when d is odd — but we see that this difference quickly becomes negligible, with the sequence of expected values converging to approximately 0.39.

6.4.5 Real-world data

As a basic example, we apply our generalized discrete EMC to four mathematics courses at the University of Wisconsin–Milwaukee, during the Fall 2019, Spring 2020, and Fall 2020 semesters. (Needless to say, these three semesters are of added interest because of the drastic changes brought on by the Covid-19 pandemic in Spring 2020.) These

Course	Fall 2019		Spring 2020		Fall 2020	
	Sections	$\widetilde{\text{EMC}}$	Sections	$\widetilde{\text{EMC}}$	Sections	$\widetilde{\text{EMC}}$
MATH 105	29	0.2059	13	0.1829	27	0.1356
MATH 231	10	0.1095	6	0.1150	7	0.1883
MATH 232	7	0.1308	7	0.1950	6	0.2617
MATH 233	7	0.1650	6	0.1867	6	0.1900

Table 5: Grade data from three semesters of mathematics courses at UW–Milwaukee.

data are contained in the Section Attrition and Grade Report, published by the Office of Assessment and Institutional Research at UWM; final letter grades are A, B, C, D, and F, so $n = 5$. The four courses are MATH 105 (Introduction to College Algebra) along with 231, 232, and 233 (Calculus I–III), and we analyze the grade distributions from the individual sections of each course. In Table 5, we record the number d of course sections in each semester, along with the unit normalized EMC among those sections. Before computing the EMC values using Lemma 6.4, we needed to account for the unequal "masses" (enrollments) among course sections. Although a generalization of our crystal EMD's from Chapter 5 would cetianly be a viable approach, here we decided to treat the sections as probability distributions, and hence we simply scaled each grade distribution so that $m = 100$, rounding as necessary to maintain whole numbers.

The results in Table 5 are also plotted in Figure 11, from which we can make one immediate observation: the EMC of each of the calculus courses *increased* from one semester to the next, whereas the EMC of the College Algebra course *decreased*. In fact, in the semester before the pandemic, College Algebra had a *higher* EMC than any of the calculus courses, but its EMC was *lower* than all of the calculus courses by the second semester of the pandemic. Of course, the EMC is only one statistic among many, and we would need to examine other measures (particularly the grade-point averages) and more previous semesters before drawing any conclusions. But we can at least say that the grade distributions of the various calculus sections (within each course) had grown farther apart since the shift to online learning in Spring 2020, while the distributions of the various sections of Col-

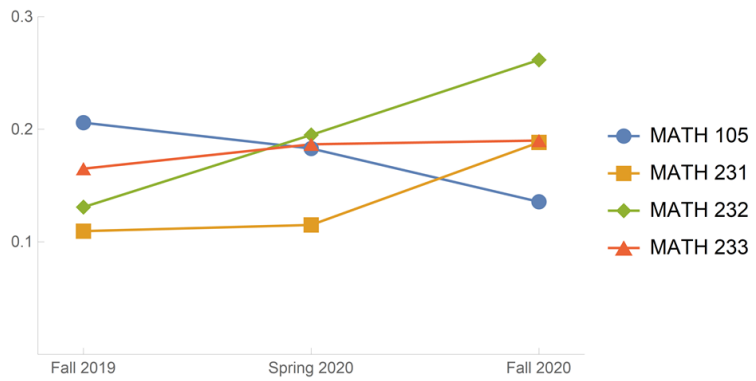


Figure 11: The unit normalized EMC values for the courses in Table 5.

lege Algebra had become closer together. In future applications, it might be interesting to observe courses' EMC values as their course coordinators vary, or to compare course EMC values between departments.

Finally, we compare these real-world EMC values to the expected value defined in Theorem 6.12. From Table 4, for $n = 5$, we observe that the approximate range of $\widetilde{\mathcal{E}}_{(5^d)}$ is between 0.25 and 0.30 for the section numbers d in our data set. The actual EMC values lie significantly below this range (with the sole exception of MATH 232 in Fall 2020). We should not be too surprised by this, of course, since college grades are (hopefully) not assigned at random, and therefore not all grade distributions are equally likely.

Remark. In the $d = 2$ case, the authors of [BW20] used the EMD to perform cluster analysis on selected grade distributions. Each pair of courses was connected by an edge if the EMD fell below a fixed threshold, and then this threshold was varied until the number of connected components became meaningful. (In a striking example with two connected components, it turned out that all of the English courses formed one cluster, while all of the mathematics and psychology courses formed the other cluster.) In the setting of this chapter, now that d is allowed to be arbitrary, we can now extend this cluster analysis so that multiple courses (or sections) form a higher-dimensional simplicial complex if their EMD is below a given threshold; this is the purview of *persistent homology* in data science. We believe that this is an interesting direction for further research and potential

applications.

6.5 A generalization of D

We leave probability distributions behind, and return to our accustomed setting of histograms. When comparing two histograms, the weighted difference **D** was a natural statistic to study, both in its own right and in relation to **EMD**. In this section we adapt the notion of weighted difference to an arbitrary number of histograms, and we study its combinatorics through representation theory.

6.5.1 **D** as a weight for $\mathfrak{sl}_d$

First we must define a weighted difference **D** that makes sense for d histograms rather than only two. We want **D** to capture just the right amount of information — namely, each of the pairwise differences between weighted totals, but nothing more — and so the most natural setting will be the root lattice Λ_R for the Lie algebra $\mathfrak{sl}_d$, that is, for Type A_{d-1}. Rather than the standard ("epsilon") coordinates, however, we will write elements of the root lattice in terms of the simple roots $\alpha_i = \varepsilon_i - \varepsilon_{i+1}$ for $1 \leqslant i < d$, along with the **affine root** $\alpha_0 := \varepsilon_d - \varepsilon_1$. (Recall this root from the final remark in Section 5.2.) Specifically, we write $(t_1, \ldots, t_d) = t_1\alpha_1 + \cdots + t_{d-1}\alpha_{d-1} + t_d\alpha_0$, with each $t_i \in \mathbb{Z}$. Note that the simple roots already form a $\mathbb{Z}$-basis for Λ_R, and the addition of the affine root introduces the relation $\alpha_1 + \cdots + \alpha_{d-1} + \alpha_0 = 0$; hence any multiple of $(1, \ldots, 1)$ must equal 0. (As we will see, this is our motivation for including the affine root: from the perspective of statistics, adding the same amount to the weighted total of every histogram has no effect on the pairwise weighted differences.) In this way, we regard elements of Λ_R as equivalence classes of integer d-tuples modulo $(1, \ldots, 1)$; hence as a $\mathbb{Z}$-module we have $\Lambda_R \cong \mathbb{Z}^d/\mathbb{Z}(1, \ldots, 1)$. We write an element of Λ_R using square brackets around any of its representatives; for example, $[1, 5, 2] = [4, 8, 5] = [-1, 3, 0]$. Note that each element

of Λ_R has a unique representative whose last component is 0.

Definition 6.13. Let $d \geqslant 2$. We (re)define the **weighted difference** to be the function $\mathbf{D} : \mathcal{H}(m,n)^d \longrightarrow \Lambda_R$ given by

$$\mathbf{D}(\lambda^1, \ldots, \lambda^d) := \big[t(\lambda^1), \ldots, t(\lambda^d)\big].$$

Remark. As mentioned above, each element of Λ_R can be represented uniquely by a d-tuple whose final coordinate is 0, which can then be truncated into a $(d-1)$-tuple. With this convention in mind, our new definition of $\mathbf{D}$ agrees with the original definition in the $d = 2$ case, since $\mathbf{D}(\lambda, \mu) = \big[t(\lambda), t(\mu)\big] = \big[t(\lambda) - t(\mu),\, 0\big] \rightsquigarrow t(\lambda) - t(\mu)$, just as before.

We pose the natural combinatorial question: how many elements in $\mathcal{H}(m,n)^d$ have a given $\mathbf{D}$-value? It turns out that the answer can be interpreted as a weight multiplicity in the $\mathfrak{sl}_d$-representation described in the following theorem. By a **virtual representation** we mean a formal $\mathbb{Z}$-combination of representations; in this context it makes sense not only to add but also to subtract characters.

Theorem 6.14. *Let $m, n \geqslant 1$ and $d \geqslant 2$. For $1 \leqslant i < d$, put $q_i = e^{\alpha_i}$, and put $q_d = e^{\alpha_0}$. Then the vector space*

$$U = \bigotimes^d S^m(\mathbb{C}^n)$$

admits a virtual representation of $\mathfrak{sl}_d$ whose character

$$\mathrm{ch}_{\mathfrak{sl}_d}(U) = \prod_{i=1}^{d} \begin{bmatrix} m+n-1 \\ m \end{bmatrix}_{q_i} \Bigg|_{q_d = (q_1 \cdots q_{d-1})^{-1}} \tag{6.22}$$

is the distribution of $\mathbf{D}$-values on $\mathcal{H}(m,n)^d$. Specifically, the coefficient of $q_1^{t_1} \cdots q_{d-1}^{t_{d-1}}$ equals the number of histogram d-tuples whose $\mathbf{D}$-value is $[t_1, \ldots, t_{d-1}, 0]$.

Proof. The easy part to prove is that the expression (6.22) is the generating function for the distribution of $\mathbf{D}$-values on $\mathcal{H}(m,n)^d$: we simply recall from (3.9) that $\begin{bmatrix} m+n-1 \\ m \end{bmatrix}_{q_i} = \sum_{\lambda \in \mathcal{H}(m,n)} q_i^{t(\lambda)}$. Hence the coefficient of $q_1^{t_1} \cdots q_d^{t_d}$ in the product in (6.22) equals the number of elements $(\lambda^1, \ldots, \lambda^d) \in \mathcal{H}(m,n)^d$ such that $t(\lambda^i) = t_i$ for each $i = 1, \ldots, d$. Carrying

out the substitution $q_d = (q_1 \cdots q_{d-1})^{-1}$ then identifies terms whose ratio is some power of $q_1 \cdots q_d$, and leaves the expansion expressed in terms of $q_1, \ldots, q_{d-1}$ only. Hence the coefficient of $q_1^{t_1} \cdots q_{d-1}^{t_{d-1}}$ now equals the number of d-tuples $\boldsymbol{\lambda}$ such that $[t(\lambda^1), \ldots, t(\lambda^d)] = [t_1, \ldots, t_{d-1}, 0]$. (Compare this to the identical calculation for the $d = 2$ case in the proof of Theorem 3.12.)

The harder part to prove is that (6.22) really is the virtual character of U under an action of $\mathfrak{sl}_d$. Consider the irreducible action on U by $(\mathfrak{gl}_n)^d = \mathfrak{gl}_n \oplus \cdots \oplus \mathfrak{gl}_n$, with each copy of $\mathfrak{gl}_n$ acting on the corresponding tensor factor. But $\mathfrak{gl}_n = \mathrm{M}_n \cong \mathrm{End}(F_2^{n-1})$, where $F_2^{n-1} \cong S^{n-1}(\mathbb{C}^2)$ is the n-dimensional irreducible representation of $\mathfrak{sl}_2$; since $\mathfrak{sl}_2$ is simple, the homomorphism defined by F_2^{n-1} has trivial kernel and thus defines an embedding $\mathfrak{sl}_2 \hookrightarrow \mathfrak{gl}_n$. The embedding is (up to conjugation) unique, and its image is called the **principal $\mathfrak{sl}_2$-subalgebra** of $\mathfrak{gl}_n$. Restricting to the principal $\mathfrak{sl}_2$-subalgebra in each copy of $\mathfrak{gl}_n$, we see that U is also a representation of $(\mathfrak{sl}_2)^d$. Crucially, these d copies of $\mathfrak{sl}_2$ embed simultaneously into $\mathfrak{sl}_d$ so as to induce a virtual $\mathfrak{sl}_d$-action, as follows. For $1 \leqslant i < d$, let $\mathfrak{s}_i \subset \mathfrak{sl}_d$ denote the image of the $\mathfrak{sl}_2$-embedding given by the Lie algebra homomorphism

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \longmapsto \quad \begin{array}{ccccc} & & a \cdot E_{i,i} & + & b \cdot E_{i,i+1} \\ + & & c \cdot E_{i+1,i} & + & d \cdot E_{i+1,i+1}, \end{array}$$

where $E_{i,j}$ is the elementary matrix with 1 in position (i, j) and 0's elsewhere. Similarly, define $\mathfrak{s}_d \subset \mathfrak{sl}_d$ as the image of the embedding

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \longmapsto \quad \begin{array}{ccccc} - & & a \cdot E_{1,1} & - & c \cdot E_{1,d} \\ - & & b \cdot E_{d,1} & - & d \cdot E_{d,d} \end{array}$$

via the negative transpose in the "four corners."

Not only does each $\mathfrak{s}_i$ inherit (from $\mathfrak{sl}_2$) a natural action on the ith tensor factor of U, but these actions extend to a (virtual) $\mathfrak{sl}_d$-action, as follows.[1] We have

$$\begin{aligned} \mathbb{C}^d &\cong \mathrm{span}\{e_i,\, e_{i+1}\} \oplus \mathrm{span}\{e_1, \ldots, e_{i-1},\, e_{i+2}, \ldots, e_d\} \\ &\cong \mathbb{C}^2 \oplus \mathbb{C}^{d-2}, \end{aligned}$$

[1] In the following calculations, we must treat the special case $i = d$ differently: namely, replace all instances of $i + 1$ by 1, to respect the embedding $\mathfrak{s}_d$, and change $\mathbb{C}^2$ to its dual $(\mathbb{C}^2)^*$.

where $\mathbb{C}^2$ is the defining representation of $\mathfrak{s}_i \cong \mathfrak{sl}_2$, and $\mathbb{C}^{d-2}$ is the direct sum of $d-2$ copies of the trivial representation of $\mathfrak{sl}_d$. In other words, we let $\mathfrak{h} \subset \mathfrak{sl}_d$ act on $\mathbb{C}^d$ not as in the defining representation, but via

$$\operatorname{diag}(h_1, \ldots, h_d) \cdot (v_1, \ldots, v_d)^\top = (0, \ldots, 0, h_i v_i, h_{i+1} v_{i+1}, 0, \ldots, 0)^\top.$$

Hence we are setting $\varepsilon_j = 0$ for all j except i and $i+1$, so that the character of this action is $\operatorname{ch}_{\mathfrak{sl}_d}(\mathbb{C}^d) = e^{\varepsilon_i} + e^{\varepsilon_{i+1}} + (d-2)$. We then regard $\mathbb{C}^2$ as the formal difference $\mathbb{C}^d - \mathbb{C}^{d-2}$ of $\mathfrak{sl}_d$-representations. Recalling that $\varepsilon_i = -\varepsilon_{i+1}$ as $\mathfrak{s}_i$-weights, we have the virtual character

$$\begin{aligned}
\operatorname{ch}_{\mathfrak{sl}_d}(\mathbb{C}^2) &= \operatorname{ch}_{\mathfrak{sl}_d}(\mathbb{C}^d) - (d-2)(\operatorname{ch}_{\mathfrak{sl}_d} \mathbb{C}) \\
&= e^{\varepsilon_i} + e^{\varepsilon_{i+1}} + (d-2) - (d-2)(1) \\
&= e^{\varepsilon_i} + e^{\varepsilon_{i+1}} \\
&= e^{\varepsilon_i} + e^{-\varepsilon_i} \\
&= x_i + x_i^{-1} = \operatorname{ch}_{\mathfrak{s}_i}(\mathbb{C}^2),
\end{aligned}$$

where we write $x_i = e^{\varepsilon_i}$. Hence, the ith tensor factor $S^m(\mathbb{C}^n) \cong S^m(S^{n-1}(\mathbb{C}^2))$ of U also admits a virtual $\mathfrak{sl}_d$-representation, whose character, by (4.5), is

$$\operatorname{ch}_{\mathfrak{sl}_d} S^m(\mathbb{C}^n) = \binom{m+n-1}{m}_{x_i.}$$

But since $\varepsilon_i = -\varepsilon_{i+1}$, we have the substitution $q_i = e^{\alpha_i} = e^{\varepsilon_i - \varepsilon_{i+1}} = e^{2\varepsilon_i} = x_i^2$. Thus by (4.6), we have

$$\operatorname{ch}_{\mathfrak{sl}_d} S^m(\mathbb{C}^n) = \binom{m+n-1}{m}_{x_i} = x_i^{-m(n-1)} \begin{bmatrix} m+n-1 \\ m \end{bmatrix}_{q_i.}$$

Multiplying all these $\mathfrak{sl}_d$-characters for $i = 1, \ldots, d$ in order to obtain the character of the d-fold tensor product U, we have

$$\operatorname{ch}_{\mathfrak{sl}_d}(U) = \prod_{i=1}^{d} x_i^{-m(n-1)} \begin{bmatrix} m+n-1 \\ m \end{bmatrix}_{q_i.}$$

But since $\varepsilon_1 + \cdots + \varepsilon_d = 0$, we have $x_1 \cdots x_d = 1$, and so $\prod_{i=1}^{d} x_i^{-m(n-1)} = 1$, leaving us with

$$\operatorname{ch}_{\mathfrak{sl}_d}(U) = \prod_{i=1}^{d} \begin{bmatrix} m+n-1 \\ m \end{bmatrix}_{q_i.} \tag{6.23}$$

Finally, since $\alpha_0 = -(\alpha_1 + \cdots + \alpha_{d-1})$, we have $q_0 = (q_1 \cdots q_{d-1})^{-1}$, and so we can perform this substitution to express the character in terms of $q_1, \ldots, q_{d-1}$, just as in (6.22). This completes the proof. □

Example 6.15. Let $d = 3$, with $m = 1$ and $n = 2$. Then by the character formula in Theorem 6.14, we have

$$\begin{aligned}\mathrm{ch}_{\mathfrak{sl}_3}(U) &= \begin{bmatrix}2\\1\end{bmatrix}_{q_1}\begin{bmatrix}2\\1\end{bmatrix}_{q_2}\begin{bmatrix}2\\1\end{bmatrix}_{q_3} \\ &= \begin{bmatrix}2\\1\end{bmatrix}_{q_1}\begin{bmatrix}2\\1\end{bmatrix}_{q_2}\begin{bmatrix}2\\1\end{bmatrix}_{(q_1q_2)^{-1}} \\ &= q_1 + q_2 + q_1q_2 + q_1^{-1} + q_2^{-2} + q_1^{-1}q_2^{-1} + 2.\end{aligned}$$

Lining up each term's exponent vector (t_1, t_2) with its corresponding **D**-value $[t_1, t_2, 0]$, we conclude that in $\mathcal{H}(1,2) \times \mathcal{H}(1,2) \times \mathcal{H}(1,2)$, there is one triple having each of the **D**-values

$$[1,0,0], \quad [0,1,0], \quad [1,1,0], \quad [-1,0,0], \quad [0,-1,0], \quad \text{and} \quad [-1,-1,0],$$

and two triples for which $\mathbf{D} = [0,0,0]$. The reader can easily verify this distribution by hand.

6.5.2 Weight diagrams

We would like to visualize the distribution of **D**-values for a given d, m, and n. Theorem 6.14 tells us that such a visualization is nothing other than the weight diagram of U under the $\mathfrak{sl}_d$-action described in the proof, and the theorem even tells us the data of the diagram explicitly via the character. Since the root lattice Λ_{R} for $\mathfrak{sl}_d$ can be depicted in $(d-1)$ dimensions, we can easily plot the weight diagrams when $d \leqslant 3$, and even the three-dimensional lattice for $d = 4$ is quite possible to visualize.

For $d = 2$, of course, we can obtain the desired diagram simply by taking our planar plots from Section 3.5 and "projecting" (via summation) each vertical strip down onto the horizontal axis; in this way, each integer on the number line is labeled with the total

number of histogram pairs having that **D**-value. (Essentially, this is just throwing away the EMC information from the plots; we have already done this in the proof of Theorem 3.12.) In this case Λ_R is generated by the sole simple root α_1, pointing in the positive **D**-direction. On one hand, we know that these "column sums" from our old plots count the number of Young diagram pairs which differ in size by a fixed number of cells; on the other hand, Theorem 6.14 says that these sums are just the weights of $S^m(\mathbb{C}^n) \otimes S^m(\mathbb{C}^n)$ as an $\mathfrak{sl}_2$-representation.[2]

For $d = 3$, the root lattice Λ_R is a two-dimensional triangular lattice generated by α_1 (pointing due east, at 0 degrees) and α_2 (pointing northwest, at 120 degrees). Now we can plot each exponent vector in $\mathrm{ch}_{\mathfrak{sl}_3}(U)$, which we label with the corresponding coefficient. From Example 6.15, we obtain a regular hexagon of dots labeled "1," along with a dot at the origin labeled "2." (See the first diagram in Figure 12, which the reader may recognize as the weight diagram of the adjoint representation of $\mathfrak{sl}_3$.) In Figure 12, for $d = 3$, we display the weight diagrams (i.e., **D**-distributions) for several other values of m and n. Incidentally, Theorem 6.14 gives a much more efficient way to create these graphics than does generalizing the generating function from Section 3.4, especially for large values of n.

Remark. We consider another, perhaps more natural way to think about the plots of **D**-distributions; we remain with the $d = 3$ case, although this approach holds in any dimension. For any histogram triple (λ, μ, ν), draw three axes radiating from the origin (as we do in Figure 12): the λ-axis in the direction of α_1, the μ-axis in the direction of α_2, and the ν-axis in the direction of α_0, i.e., southwest at 240 degrees. (These are the roots corresponding to q_1, q_2, and q_3 in the character formula.) Now we can plot a **D**-value $[a, b, c]$, whether or not its final coordinate is 0, by starting at the origin and moving a units parallel to the λ-axis, b units parallel to the μ-axis, and c units parallel to the ν-axis. Notice that any multiple of $[1, 1, 1]$ lands on the origin, just as we would expect.

[2]When $d = 2$, we can drop the term "virtual" since $U \cong S^m(S^{n-1}(\mathbb{C}^2)) \otimes S^m(S^{n-1}(\mathbb{C}^2)^*)$ under the standard action of $\mathfrak{sl}_2$.

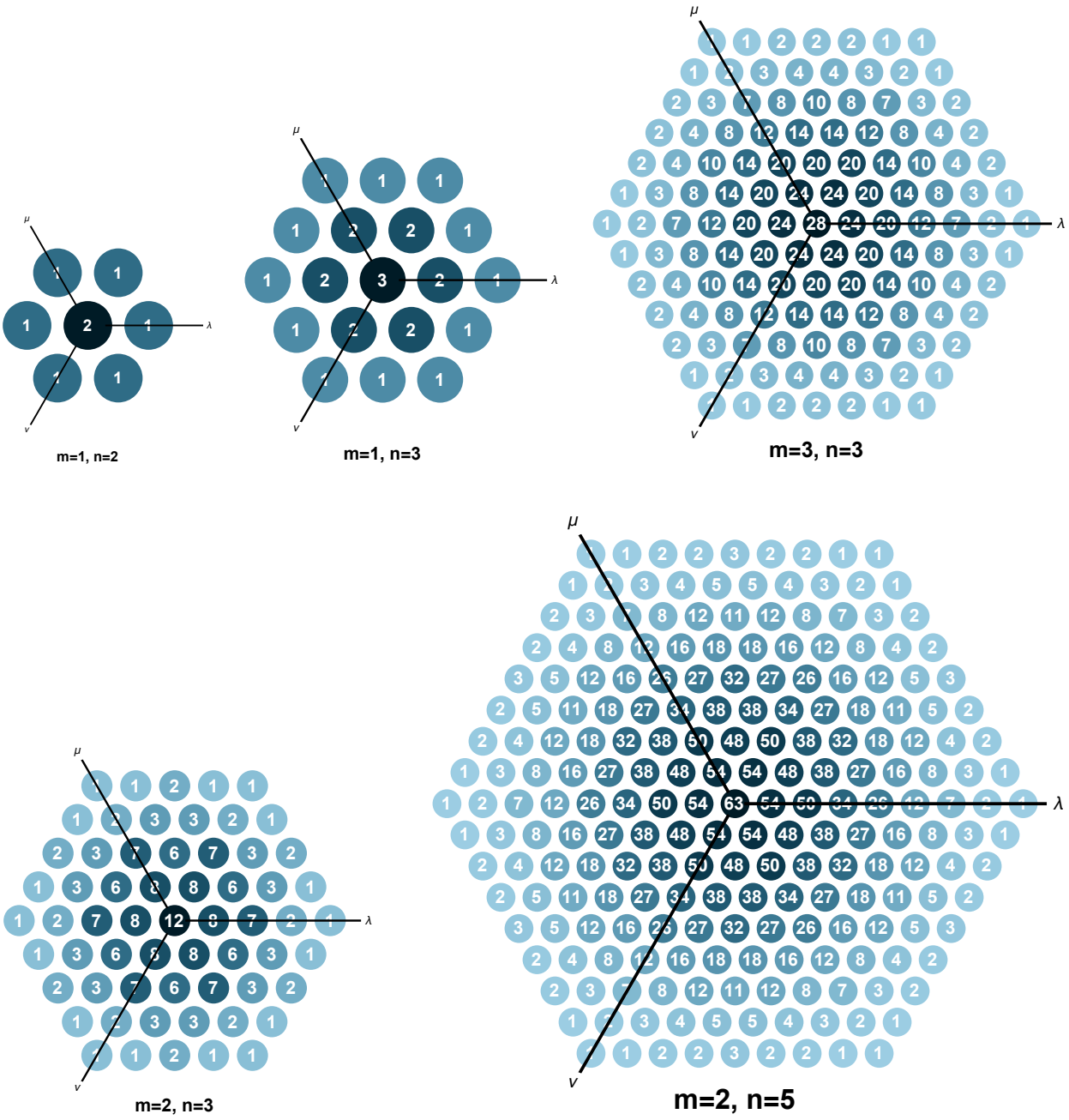


Figure 12: Plots of the distribution of **D**-values for $d = 3$. By Theorem 6.14, these are also the weight diagrams of the space $U = \bigotimes^3 S^m(\mathbb{C}^n)$ as a virtual $\mathfrak{sl}_3$-representation.

Now the angular position of a **D**-value tells us immediately the ordering of the three weighted totals: for example, at all points strictly between 120 and 180 degrees, we must have $t(\mu) > t(\nu) > t(\lambda)$. The radial distance is a measure of the "lopsidedness" among the three weighted totals. By the symmetry inherent in our setup, we can get a full picture of the diagram by restricting our attention to just one sector, such as that between 0 and 60 degrees where $t(\lambda) \geqslant t(\mu) \geqslant t(\nu)$. Although we have focused on **D** at the expense of the EMC in this section, we could nevertheless include the EMC frequencies in the $d = 3$ diagrams by plotting them in the third dimension, along strips directly above each **D**-value, just as in our $d = 2$ plots.

From the perspective of representation theory, rather than work with the entire weight diagram of U, it is natural to seek the decomposition of U into irreducible subrepresentations of $\mathfrak{sl}_3$. Since finite-dimensional irreducible representations are indexed by their highest weight, we can actually convey the same information from the weight diagram with a much sparser diagram, on which we plot only those highest weights corresponding to the irreducible decomposition of U. (See Figure 13, where we show only the region of the $\mathfrak{sl}_3$ root lattice inside the dominant chamber.) As experts will have anticipated, we programmed these decompositions by multiplying $\mathrm{ch}_{\mathfrak{sl}_3}(U)$ by the Weyl denominator $(1 - q_1^{-1})(1 - q_2^{-1})(1 - (q_1 q_2)^{-1})$, and then plotting the dominant weights in the resulting expansion. In these plots, the number inside each dot indicates the multiplicity of that irreducible representation inside U. Note that some multiplicities are negative (depicted by red dots), reflecting the fact that U is actually a virtual representation of $\mathfrak{sl}_3$. When $m = 1$, we see that U is isomorphic to the irreducible representation of $\mathfrak{sl}_3$ with highest weight $(n-1)(\alpha_1 + \alpha_2)$. As m or n increases, certain patterns are emerging from the diagrams, and this decomposition will be an interesting subject of further research. The three-dimensional diagrams in the $d = 4$ case will likely be of interest as well.

Remark. From this perspective of representation theory, the $(m, n) \longleftrightarrow (n-1,\ m+1)$

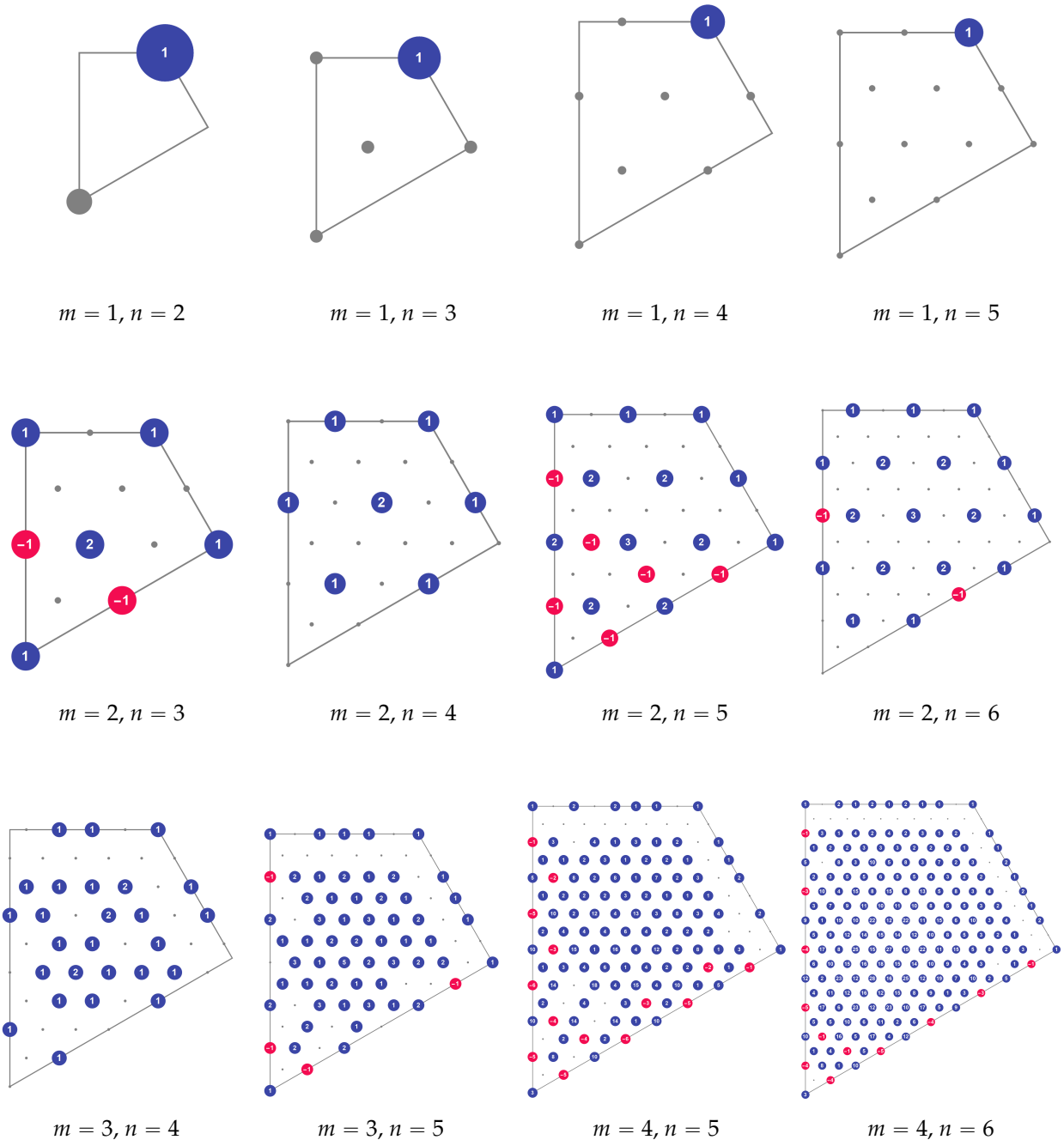


Figure 13: The decomposition of $U = \bigotimes^3 S^m(\mathbb{C}^n)$ as a virtual $\mathfrak{sl}_3$-representation. Each irreducible representation of $\mathfrak{sl}_3$ is plotted (with multiplicity) at its highest weight. Negative multiplicities are colored red.

duality in (3.3) is a manifestation of **Hermite reciprocity**: as $\mathfrak{sl}_2$-representations,

$$\begin{aligned} S^m(\mathbb{C}^n) &\cong S^m\left(S^{n-1}(\mathbb{C}^2)\right) \\ &\cong S^{n-1}\left(S^m(\mathbb{C}^2)\right) \\ &\cong S^{n-1}(\mathbb{C}^{m+1}). \end{aligned}$$

This can be seen directly from the q-binomial coefficient that expresses the $\mathfrak{sl}_2$-character in (4.5). Since the distribution of **D**-values is completely determined by the $\mathfrak{sl}_d$-action on U, which in turn is completely determined by the actions of copies of $\mathfrak{sl}_2$, it is clear why $\mathcal{H}(m,n)^d$ must have the same **D**-distribution as $\mathcal{H}(n-1,\ m+1)^d$.

7 ENRIGHT RESOLUTIONS AND BLATTNER'S FORMULA IN TYPE A

As a technical coda to the thesis, this final chapter describes a striking connection that the author observed between Howe duality in Type A — in particular, the complicated structure of the infinite-dimensional $\mathfrak{su}(p,q)$-modules $\widetilde{F}^{\lambda}_{p,q}$ from Section 4.4, which include the first Wallach representation — and certain non-holomorphic discrete series representations of $\mathfrak{su}(k,p+q)$. This chapter leaves the EMD far behind, and some familiarity with the representation theory of real groups is required. The results in this chapter are for the most part contained in the author's preprint [Erib].

The first of the two settings is that of Section 4.4, in particular the Howe duality decomposition (4.15), and the $\mathfrak{su}(p,q)$-modules therein which we denoted by $\widetilde{F}^{\lambda}_{p,q}$. (Now that there is no danger of confusing Greek letters for histograms in this chapter, we will let λ denote an integral weight of $\mathfrak{gl}_k$. In particular, λ will from now on play the role of ξ from the Howe duality decomposition (4.15).) Although the modules $\widetilde{F}^{\lambda}_{p,q}$ have a complicated structure (outside the stable range $k \geqslant p+q$), they nonetheless have finite resolutions, constructed by Enright and Willenbring in [EW04], in terms of generalized Verma modules (4.9). The Enright resolution of $\widetilde{F}^{\lambda}_{p,q}$ is determined by the (non)occurrence of each possible $M_{\mu,\nu}$, which we will encode via the integer $\varepsilon^{\lambda}_{\mu,\nu}$ (see Definition 7.1).

The second setting is, it seems, far removed from the first: namely, the discrete series of the special indefinite unitary group $\mathrm{SU}(k,p+q)$. Harish–Chandra first conjectured (1954), and subsequently proved in [HC66], that a linear connected semisimple real Lie group G_0 has discrete series representations if and only if its rank is equal to that of a maximal compact subgroup K_0. In this case, it is natural to restrict a discrete series representation of G_0 to the action of the complexified Lie algebra $\mathfrak{k}$, and to study the multiplicities of irreducible $\mathfrak{k}$-modules ("$\mathfrak{k}$-types") in the resulting decomposition. A formula for these $\mathfrak{k}$-type multiplicities — an alternating sum over the Weyl group of $\mathfrak{k}$ — was (according

to Harish–Chandra) conjectured by Robert J. Blattner, and this "Blattner's formula" was eventually proved by Hecht and Schmid in [HS75]. We will write $B(\delta, \eta)$ to denote the value of Blattner's formula for the $\mathfrak{k}$-type with highest weight δ inside the discrete series representation of G_0 with Blattner parameter η. (Outside certain conditions, the outputs of Blattner's formula do not correspond to actual discrete series representations, and can assume negative values.)

When we put $G_0 = \mathrm{SU}(k, p+q)$ with a certain embedding of K_0 (see Figure 15), we obtain our main result uniting the two settings above, namely the equation

$$\varepsilon^{\lambda}_{\mu,\nu} = B\big(0, [\![\lambda, \mu, \nu]\!]\big),$$

with the $[\![-,-,-]\!]$ notation to be defined below in (7.16). Theoretically, the heart of the paper lies in Lemmas 7.8 and 7.9, which spell out the explicit connection between the character theory of $\mathbb{C}[V]$ and the generating function $\mathbf{b}(0) := \sum_{\eta} B(0,\eta)e^{\eta}$ introduced by Willenbring and Zuckerman in [WZ08]. Our result allows us (see Example 7.7) to write down the Enright resolution of $\widetilde{F}^{\lambda}_{p,q}$ — without any knowledge of the machinery used in [EW04] — simply by reading off the coefficient of e^{λ} in $\mathbf{b}(0)$.

7.1 Preliminaries

7.1.1 Howe duality in Type A, continued

The reader is invited to review Section 4.4, from which we carry forward all definitions and notation. In addition, we supplement Section 4.4 with the following preliminary observations.

Recall the maximal compact subalgebra $\mathfrak{k}_0 = \mathfrak{s}(\mathfrak{u}_p \oplus \mathfrak{u}_q) \subset \mathfrak{g}_0$, embedded block-diagonally. In the Howe duality setting, the complexification $\mathfrak{k} = \mathfrak{s}(\mathfrak{gl}_p \oplus \mathfrak{gl}_q) \subset \mathfrak{g}$ acts on V by the derived action (up to a central character) of the complexified group $K = \mathrm{S}(\mathrm{GL}_p \times \mathrm{GL}_q)$:

$$(g,h)\cdot(X,Y) = (Xg^{-1}, Yh^{-1}), \tag{7.1}$$

where $g \in \mathrm{GL}_p$ and $h \in \mathrm{GL}_q$. Since the action in (7.1) commutes with the GL_k-action in (4.12), we have an action of the product $M' = \mathrm{S}(\mathrm{GL}_k \times \mathrm{GL}_p \times \mathrm{GL}_q)$ upon V:

$$(g_k, g_p, g_q) \cdot (X, Y) = (g_k^{-T} X g_p^{-1}, g_k Y g_q^{-1}). \tag{7.2}$$

(We choose the notation M' because this group will soon play the role of a Levi factor in Section 7.4, in which we will decorate all groups and Lie algebras with prime symbols in order to distinguish them from the present context.)

Recall from (4.15) the Howe duality decomposition

$$\mathbb{C}[V] \cong \bigoplus_{\lambda} F_k^{\lambda} \otimes \widetilde{F}_{p,q}^{\lambda}. \tag{7.3}$$

We now clarify which λ's appear in this decomposition: the sum ranges over all weakly decreasing k-tuples λ such that $\ell(\lambda^+) \leqslant p$ and $\ell(\lambda^-) \leqslant q$. These length conditions motivate a notion of "stability": we say that the parameters k, p, q lie in the **stable range** when

$$k \geqslant p + q.$$

Inside the stable range, any pair of partitions (λ^+, λ^-) such that $\ell(\lambda^+) \leqslant p$ and $\ell(\lambda^-) \leqslant q$ determines a weakly decreasing k-tuple $\lambda = [\lambda^+, \lambda^-]$, i.e., the highest weight of the rational GL_k-representation F_k^{λ}. Furthermore, inside the stable range, upon restriction to $K = \mathrm{S}(\mathrm{GL}_p \times \mathrm{GL}_q)$, the module $\widetilde{F}_{p,q}^{\lambda}$ has the especially nice structure

$$\widetilde{F}_{p,q}^{\lambda} \cong \mathbb{C}[\mathrm{M}_{p,q}] \otimes \left(F_p^{\lambda^+} \otimes F_q^{\lambda^-} \right), \tag{7.4}$$

where K acts on the second tensor factor in the obvious way, and on the first tensor factor via

$$(g, h) \cdot f(X) = f(g^T X h). \tag{7.5}$$

The reader may recognize (7.4) as the generalized Verma module for $\mathfrak{g}$ induced from the irreducible $\mathfrak{k}$-module $F_p^{\lambda^+} \otimes F_q^{\lambda^-}$, as in (4.9). In this context, $\mathbb{C}[\mathrm{M}_{p,q}] \cong S(\mathfrak{p}^-)$ as $\mathfrak{k}$-modules, given the Cartan decomposition in (4.7).

Outside the stable range, the structure of $\widetilde{F}^{\lambda}_{p,q}$ is more complicated than in (7.4). Regardless of stability, however, $\widetilde{F}^{\lambda}_{p,q}$ admits a certain finite resolution, as proved by Enright and Willenbring in [EW04]. This idea is an extension of the BGG resolution of a finite-dimensional module in terms of Verma modules, and of the Lepowsky resolution of a finite-dimensional module in terms of generalized Verma modules. We describe these resolutions in the next section.

7.1.2 Enright resolutions

There exists an **Enright resolution** (also called a "generalized BGG resolution" in [EH04]) for any unitarizable highest-weight representation of a simple connected real Lie group ($G_0 = \mathrm{SU}(p,q)$ in our case), given a maximal compact subgroup ($K_0 = \mathrm{S}(\mathrm{U}_p \times \mathrm{U}_q)$ for us) such that (G_0, K_0) is a Hermitian symmetric pair. (As we will see later, the block-diagonal embedding $K_0 \subset G_0$ implies the Hermitian symmetric property.) Each of the finitely many terms in the Enright resolution is the direct sum of generalized Verma modules

$$U(\mathfrak{g}) \otimes_{U(\mathfrak{q})} L_{\mathfrak{k}}(\xi) \tag{7.6}$$

associated with the standard maximal parabolic subalgebra $\mathfrak{q} = \mathfrak{k} \oplus \mathfrak{p}^+$, where $L_{\mathfrak{k}}(\xi)$ denotes the irreducible $\mathfrak{k}$-module with highest weight ξ, regarded as a $\mathfrak{q}$-module by letting $\mathfrak{p}^+$ act trivially. As a $\mathfrak{k}$-module, (7.6) is isomorphic to $S(\mathfrak{p}^-) \otimes L_{\mathfrak{k}}(\xi)$. Therefore in our case, namely the Enright resolution of $\widetilde{F}^{\lambda}_{p,q}$, the generalized Verma modules restrict as $\mathfrak{k}$-modules to the form

$$M_{\mu,\nu} := \mathbb{C}[\mathrm{M}_{p,q}] \otimes (F^{\mu}_{p} \otimes F^{\nu}_{q}), \tag{7.7}$$

where μ and ν are partitions of lengths at most p and q respectively. See Theorem 2 and the preceding discussion in [EW04] for the explicit construction of the individual terms in the resolution; for our purposes, we will appeal only to its existence, because our main result will actually show us how to find all the data from the Enright resolution hidden within a certain generating function. (See Example 7.7.)

We will now assign to each generalized Verma module $M_{\mu,\nu}$ an integer $\varepsilon^{\lambda}_{\mu,\nu}$ which describes its occurrence in the Enright resolution of $\widetilde{F}^{\lambda}_{p,q}$. Let $\mathrm{ch}_{\mathfrak{k}}(-)$ denote the character of a $\mathfrak{k}$-module. Let

$$\mathcal{X} := \left\{ \mathrm{ch}_{\mathfrak{k}}\big(M_{\mu,\nu}\big) \;\middle|\; \ell(\mu) \leqslant p \text{ and } \ell(\nu) \leqslant q \right\},$$

which is a basis for its integer span $\mathbb{Z}\mathcal{X}$. Hence $\mathrm{ch}_{\mathfrak{k}}(\widetilde{F}^{\lambda}_{p,q})$ has a unique expansion as an element of $\mathbb{Z}\mathcal{X}$, which we can regard as the Euler characteristic of the Enright resolution of $\widetilde{F}^{\lambda}_{p,q}$.

Definition 7.1. For fixed k, p, q, suppose λ, μ, ν satisfy the following conditions:

$$\lambda \text{ a weakly decreasing } k\text{-tuple}; \quad \ell(\lambda^+),\, \ell(\mu) \leqslant p; \quad \ell(\lambda^-),\, \ell(\nu) \leqslant q. \tag{7.8}$$

Then we define

$$\varepsilon^{\lambda}_{\mu,\nu} := \text{the coefficient of } \mathrm{ch}_{\mathfrak{k}}\big(M_{\mu,\nu}\big) \text{ in } \mathrm{ch}_{\mathfrak{k}}(\widetilde{F}^{\lambda}_{p,q}),$$

where $\mathrm{ch}_{\mathfrak{k}}(\widetilde{F}^{\lambda}_{p,q})$ is expanded as an element of $\mathbb{Z}\mathcal{X}$.

Whether or not we are in the stable range, we thus have a $\mathfrak{k}$-character of the form

$$\mathrm{ch}_{\mathfrak{k}}(\widetilde{F}^{\lambda}_{p,q}) = \mathrm{ch}_{\mathfrak{k}}\big(\mathbb{C}[\mathrm{M}_{p,q}]\big) \cdot \sum_{\mu,\nu} \varepsilon^{\lambda}_{\mu,\nu} \cdot \mathrm{ch}_{\mathfrak{k}}(F^{\mu}_{p} \otimes F^{\nu}_{q}), \tag{7.9}$$

where $\varepsilon^{\lambda}_{\mu,\nu}$ is nonzero for only finitely many μ, ν. Note that inside the stable range, (7.4) implies that

$$\varepsilon^{\lambda}_{\mu,\nu} = \delta_{(\lambda^+,\lambda^-),\,(\mu,\nu)} \tag{7.10}$$

where δ is the Kronecker delta.

We can now write out the character of $\mathbb{C}[V]$ as a representation of M', under the action in (7.2). Combining (4.15) and (7.9), we have the factorization

$$\mathrm{ch}\,\mathbb{C}[V] = \mathrm{ch}\,\mathbb{C}[\mathrm{M}_{p,q}] \cdot \sum_{\lambda,\mu,\nu} \varepsilon^{\lambda}_{\mu,\nu} \cdot \mathrm{ch}(F^{\lambda}_{k} \otimes F^{\mu}_{p} \otimes F^{\nu}_{q}) \tag{7.11}$$

as a character of M' and therefore of $\mathfrak{m}'$. This time the sum is infinite, since λ ranges over an infinite set.

7.1.3 Discrete series and Blattner's formula

In this section, we temporarily forget the concrete setting of the previous section, and so we regard G_0, K_0, $\mathfrak{g}$, $\mathfrak{k}$, etc. in the abstract. Nonetheless, the reader may profitably keep in mind the specific notation of Sections 4.4 and 7.1, which is one example of the general setting below.

A **discrete series representation** of a Lie group G_0 is a topologically closed subspace of $L^2(G_0)$ that is invariant and irreducible under the usual G_0-action on functions. (See the end of our introduction for the analytical background of the discrete series.) As mentioned above, Harish–Chandra proved in [HC66] that a connected semisimple group G_0 has discrete series representations if and only if $\operatorname{rank} G_0 = \operatorname{rank} K_0$, where $K_0 \subset G_0$ is a maximal compact subgroup. (Hence G_0 must have finite center.)

It is natural to restrict a discrete series representation from G_0 to K_0, thus to the complexified Lie algebra $\mathfrak{k}$, then decompose into a direct sum of irreducible finite-dimensional $\mathfrak{k}$-representations ("$\mathfrak{k}$-types"), and then seek the multiplicity of a given $\mathfrak{k}$-type in this decomposition. Before recording this multiplicity formula (named after Robert J. Blattner), we explain the preliminary notation, following that used by Willenbring and Zuckerman in [WZ08].

Let G_0 be a connected, semisimple Lie group with finite center, with complexified Lie algebra $\mathfrak{g}$. Let $K_0 \subset G_0$ be a maximal compact subgroup, with complexified Lie algebra $\mathfrak{k}$. Fix a Cartan subalgebra $\mathfrak{h} \subset \mathfrak{g}$. By Harish–Chandra's equal-rank criterion, we may assume that $\mathfrak{h} \subset \mathfrak{k}$. Let $\Phi = \Phi(\mathfrak{g}, \mathfrak{h})$ denote the corresponding root system.

Let Φ^+ denote a choice of positive roots and let let $\Phi^- = -\Phi^+$ denote the set of negative roots. We call a root $\alpha \in \Phi$ a **compact root** if $\mathfrak{g}_\alpha \subset \mathfrak{k}$; otherwise, we call the root **noncompact**. We denote the sets of compact and noncompact roots by Φ_c and Φ_{nc} respectively. Thus we have $\mathfrak{k} = \mathfrak{h} \oplus \bigoplus_{\alpha \in \Phi_c} \mathfrak{g}_\alpha$. We write $\mathfrak{p} = \bigoplus_{\alpha \in \Phi_{nc}} \mathfrak{g}_\alpha$ for the sum of the noncompact root spaces, and so we have $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$. Finally, since we will need to

distinguish the positive and negative roots, we let

$$\Phi_c^+ = \Phi^+ \cap \Phi_c, \quad \Phi_{nc}^+ = \Phi^+ \cap \Phi_{nc}, \quad \Phi_c^- = \Phi^- \cap \Phi_c, \quad \Phi_{nc}^- = \Phi^- \cap \Phi_{nc}$$

in the obvious way. Then we have $\mathfrak{p} = \mathfrak{p}^+ \oplus \mathfrak{p}^-$, where

$$\mathfrak{p}^+ = \bigoplus_{\alpha \in \Phi_{nc}^+} \mathfrak{g}_\alpha \qquad \text{and} \qquad \mathfrak{p}^- = \bigoplus_{\alpha \in \Phi_{nc}^-} \mathfrak{g}_\alpha.$$

Let $W_\mathfrak{g}$ and $W_\mathfrak{k}$ be the Weyl groups for $\mathfrak{g}$ and $\mathfrak{k}$ respectively, and let $\ell(w)$ denote the length of a Weyl group element $w \in W_\mathfrak{k}$, so that $\ell(w) = |w(\Phi_c^+) \cap \Phi_c^-|$. Let $\Pi = \{\alpha_1, \ldots, \alpha_r\} \subset \Phi^+$ be the set of simple roots, and set $\Pi_c = \Pi \cap \Phi_c$ and $\Pi_{nc} = \Pi \cap \Phi_{nc}$. Then either of the sets Π_c or Π_{nc} determines the decomposition $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$. The case where $|\Pi_{nc}| = 1$ is especially important in the theory of maximal parabolic subalgebras, and in this case we say that (G_0, K_0) is a **Hermitian symmetric pair**. (This is the algebraic definition, as opposed to the analytical one we gave in Section 4.3.1. See [EHP14] for a detailed exposition of this theory.)

Let $(-,-)$ denote the Killing form on $\mathfrak{g}$, which restricts to a nondegenerate form on $\mathfrak{h}$ and thus allows us to identify $\mathfrak{h}$ with $\mathfrak{h}^*$. Under this identification, $\alpha^\vee := \frac{2\alpha}{(\alpha,\alpha)}$ is identified with the simple coroot in $\mathfrak{h}$ corresponding to α, for each $\alpha \in \Pi$. A weight $\xi \in \mathfrak{h}^*$ is called an integral weight for $\mathfrak{g}$ if $(\xi, \alpha^\vee) \in \mathbb{Z}$ for all $\alpha \in \Pi$; the set of $\mathfrak{g}$-integral weights is denoted by $P(\mathfrak{g})$. The same condition defines the $\mathfrak{k}$-integral weights $P(\mathfrak{k})$ if we replace Π by Π_c. Moreover, ξ is said to be $\mathfrak{g}$-dominant if $(\xi, \alpha) \geqslant 0$ for all $\alpha \in \Pi$, and likewise $\mathfrak{k}$-dominant if $(\xi, \alpha) \geqslant 0$ for all $\alpha \in \Pi_c$; the sets of dominant integral weights are denoted by $P_+(\mathfrak{g})$ and $P_+(\mathfrak{k})$. We say that ξ is $\mathfrak{g}$-regular if $(\xi, \alpha) \neq 0$ for all $\alpha \in \Phi$. For $\delta \in P_+(\mathfrak{k})$ we let $L_\mathfrak{k}(\delta)$ denote the finite-dimensional representation of $\mathfrak{k}$ with highest weight δ. We distinguish the elements

$$\rho_c = \frac{1}{2} \sum_{\alpha \in \Phi_c^+} \alpha \qquad \text{and} \qquad \rho_{nc} = \frac{1}{2} \sum_{\alpha \in \Phi_{nc}^+} \alpha$$

to be half the sum of the positive compact (respectively, noncompact) roots. Finally, define $Q(\xi)$ to be the number of ways of writing ξ as a sum of *non*compact positive roots; in other words, Q is the Φ_{nc}^+-partition function.

For $\delta, \eta \in P(\mathfrak{k})$, **Blattner's formula** is given by

$$B(\delta, \eta) = \sum_{w \in W_{\mathfrak{k}}} (-1)^{\ell(w)} Q\big(w(\delta + \rho_c) - \rho_c - \eta\big). \tag{7.12}$$

Under certain assumptions on δ and η, the output of Blattner's formula gives a $\mathfrak{k}$-type multiplicity in a discrete series representation of G_0. In order to index these representations, we appeal to Vogan's theory (see [Vog79]) of the **lowest $\mathfrak{k}$-type**, i.e., the unique $\mathfrak{k}$-type which appears with multiplicity 1 in the $\mathfrak{k}$-decomposition of a discrete series representation. For $\eta \in P_+(\mathfrak{k})$, we write $\mathcal{D}^\eta$ for the discrete series representation of G_0 whose lowest $\mathfrak{k}$-type is $L_{\mathfrak{k}}(\eta)$. This η is sometimes called the "Blattner parameter" of the representation $\mathcal{D}^\eta$. (See [Kna86], page 310.) We now state the precise interpretation of Blattner's formula in terms of $\mathfrak{k}$-type multiplicities, as proved in [HS75].

Theorem 7.2 (Hecht and Schmid, 1975)**.** *Assume $\delta, \eta \in P_+(\mathfrak{k})$ such that $\eta + \rho_c - \rho_{nc}$ is $\mathfrak{g}$-dominant regular. Then $B(\delta, \eta)$ equals the multiplicity of $L_{\mathfrak{k}}(\delta)$ in $\mathcal{D}^\eta$.*

In [WZ08], Willenbring and Zuckerman introduce the formal power series

$$\mathbf{b}(\delta) := \sum_{\eta \in \mathfrak{h}^*} B(\delta, \eta) e^\eta$$

which encodes the values of Blattner's formula when the first argument $\delta \in P_+(\mathfrak{k})$ is held fixed. Their main result is the closed form

$$\mathbf{b}(\delta) = \operatorname{ch} L_{\mathfrak{k}}(\delta) \cdot \frac{\prod_{\alpha \in \Phi_c^+} 1 - e^{-\alpha}}{\prod_{\alpha \in \Phi_{nc}^+} 1 - e^{-\alpha}}$$

which specializes to

$$\mathbf{b}(0) = \frac{\prod_{\alpha \in \Phi_c^+} 1 - e^{-\alpha}}{\prod_{\alpha \in \Phi_{nc}^+} 1 - e^{-\alpha}}. \tag{7.13}$$

In proving our main result in Section 7.4, we will reinterpret $\mathbf{b}(0)$ in terms of the character of $\mathbb{C}[V]$ from the Howe duality setting.

7.2 Discrete series and Howe duality: one noncompact simple root

We return to the concrete setting in which G_0, K_0, $\mathfrak{g}$, $\mathfrak{k}$, etc. denote the specific objects in Section 4.3.1. Note that

$$\operatorname{rank} G_0 = \operatorname{rank} \mathrm{SU}(p,q) = p + q - 1 = \operatorname{rank} \mathrm{S}(\mathrm{U}_p \times \mathrm{U}_q) = \operatorname{rank} K_0,$$

so that $\mathrm{SU}(p,q)$ satisfies Harish–Chandra's criterion and thus has discrete series representations. As it turns out, we have already encountered one example of discrete series representations in the Howe decomposition (4.15). Recall that inside the stable range $k \geqslant p+q$, the infinite-dimensional $\widetilde{F}^{\lambda}_{p,q}$ decomposes as a $\mathfrak{k}$-module as in (7.4). In this case, $\widetilde{F}^{\lambda}_{p,q}$ is a discrete series representation (or limit thereof) of a covering group of $\mathrm{SU}(p,q)$ — hence of $\mathfrak{su}(p,q)$ — and by (7.4) its lowest $\mathfrak{k}$-type is $F_p^{\lambda^+} \otimes F_q^{\lambda^-}$. (See [HTW05, Theorem 3.2(b)].) Therefore, inside the stable range, by regarding λ as the pair (λ^+, λ^-) of highest weights, we can write

$$\widetilde{F}^{\lambda}_{p,q} \cong \mathcal{D}^{\lambda}$$

as a discrete series representation of $\mathfrak{su}(p,q)$. Since in the setting of Section 4.3.1 the embedding $\mathfrak{k} \subset \mathfrak{g}$ is block diagonal in nature (see Figure 14), there is exactly one noncompact simple root, namely α_p, and so (G_0, K_0) is a Hermitian symmetric pair.

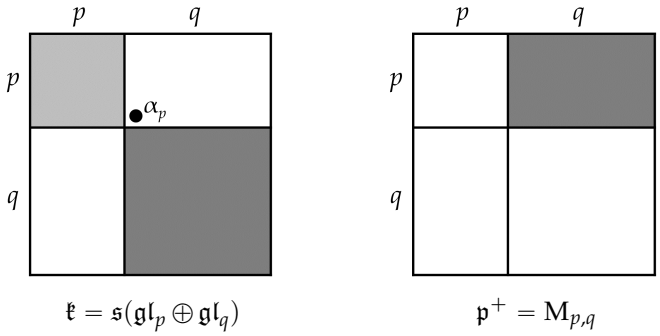


Figure 14: The Hermitian symmetric case (i.e., one noncompact simple root) in Type A, where $\mathfrak{g} = \mathfrak{sl}_{p+q}$.

In this Hermitian symmetric case for $G_0 = \mathrm{SU}(p,q)$, we can derive a combinatorial expression, *without* signs, for Blattner's formula. Since $\mathfrak{k} = \mathfrak{s}(\mathfrak{gl}_p \oplus \mathfrak{gl}_q)$, the weights δ

and η in Blattner's formula are actually pairs $\delta = (\delta^p, \delta^q)$ and $\eta = (\eta^p, \eta^q)$. Here δ^p and η^p are partitions of length at most p (thus highest weights of polynomial representations of GL_p), while δ^q and η^q are partitions of length at most q (and thus highest weights of polynomial representations of GL_q). Recall that for partitions α, β, γ, with lengths at most k, the **Littlewood–Richardson coefficient** $c^{\gamma}_{\alpha,\beta}$ is the multiplicity of F^{γ}_k in the tensor product $F^{\alpha}_k \otimes F^{\beta}_k$, under the restriction of $\mathrm{GL}_k \times \mathrm{GL}_k$ to its diagonal subgroup $\Delta(\mathrm{GL}_k) \cong \mathrm{GL}_k$.

Proposition 7.3. *Let $\mathfrak{g} = \mathfrak{sl}_{p+q}$, with $\Pi_{\mathrm{nc}} = \{\alpha_p\}$; hence $\mathfrak{k} = \mathfrak{s}(\mathfrak{gl}_p \oplus \mathfrak{gl}_q)$ embedded block diagonally in $\mathfrak{g}$. Let $\delta = (\delta^p, \delta^q)$, $\eta = (\eta^p, \eta^q) \in P_+(\mathfrak{k})$. Then*

$$B(\delta, \eta) = \sum_{\xi} c^{\delta^p}_{\xi,\eta^p} c^{\delta^q}_{\xi,\eta^q} \tag{7.14}$$

where the sum is over all partitions ξ such that $\ell(\xi) \leqslant \min(p, q)$.

Proof. By Cauchy's identity, we have the following decomposition as a K-module:

$$\mathbb{C}[\mathrm{M}_{p,q}] \cong \bigoplus_{\xi} F^{\xi}_p \otimes F^{\xi}_q, \tag{7.15}$$

where ξ ranges over all partitions of length at most $\min(p, q)$. Starting from (7.4), and putting $\eta^p = \eta^+$ and $\eta^q = \eta^-$, we have

$$\begin{aligned}
\mathcal{D}^{\eta} \cong \widetilde{F}^{\eta}_{p,q} &\cong \mathbb{C}[\mathrm{M}_{p,q}] \otimes \left(F^{\eta^p}_p \otimes F^{\eta^q}_q \right) \\
&\cong \left(\bigoplus_{\xi} F^{\xi}_p \otimes F^{\xi}_q \right) \otimes \left(F^{\eta^p}_p \otimes F^{\eta^q}_q \right) \\
&\cong \bigoplus_{\xi} \left(F^{\xi}_p \otimes F^{\eta^p}_p \right) \otimes \left(F^{\xi}_q \otimes F^{\eta^q}_q \right) \\
&\cong \bigoplus_{\xi} \left(\bigoplus_{\delta^p} c^{\delta^p}_{\xi,\eta^p} F^{\delta^p}_p \right) \otimes \left(\bigoplus_{\delta^q} c^{\delta^q}_{\xi,\eta^q} F^{\delta^q}_q \right) \\
&\cong \bigoplus_{\delta^p, \delta^q} \left(\sum_{\xi} c^{\delta^p}_{\xi,\eta^p} c^{\delta^q}_{\xi,\eta^q} \right) F^{\delta^p}_p \otimes F^{\delta^q}_q \\
&= \bigoplus_{\delta^p, \delta^q} B(\delta, \eta)\, F^{\delta^p}_p \otimes F^{\delta^q}_{q,}
\end{aligned}$$

where δ^p and δ^q range over all partitions of length at most p and q respectively. Comparing the coefficients in the last two lines, we see that the proposition holds. □

One benefit of Proposition 7.3 is that there are no negative signs in the sum on the right-hand side of (7.14), whereas Blattner's formula itself is an alternating sum. One application of our main result will be a similar combinatorial expression, without signs, for discrete series representations *outside* the Hermitian symmetric case (in particular, with *two* noncompact simple roots). For this, we will need a generalization of the classical Littlewood–Richardson coefficients, which we present in the following section.

7.3 Generalized Littlewood–Richardson coefficients

In the Howe decomposition (4.15), the infinite-dimensional modules $\widetilde{F}^{\lambda}_{p,q}$ actually provide combinatorial information about the representation theory of the finite-dimensional modules F^{λ}_k. Using the method of seesaw pairs, the authors of [CEW22] describe a generalization of the Littlewood–Richardson coefficients $c^{\gamma}_{\alpha,\beta}$ which applies to *multi*-tensor product multiplicities for *rational* (not only polynomial) representations of GL_k. Below we outline the generalization in the case of only two tensor factors, since this will suffice for our purposes here.

Let α, β, γ be k-tuples of weakly decreasing integers (i.e., highest weights indexing rational GL_k-representations). We wish to obtain the multiplicity of F^{γ}_k inside $F^{\alpha}_k \otimes F^{\beta}_k$, which we will denote by $\mathrm{LRC}^{\gamma}_{\alpha,\beta}$. To this end, we consider all possible **hollow contingency tables** of the form

$$\begin{array}{c} \\ \alpha^+ \\ \beta^+ \\ \gamma^- \end{array} \begin{array}{c} \begin{array}{ccc} \alpha^- & \beta^- & \gamma^+ \end{array} \\ \begin{bmatrix} \mathbf{0} & * & * \\ * & \mathbf{0} & * \\ * & * & \mathbf{0} \end{bmatrix} \end{array}$$

where each star denotes an arbitrary partition (**0** being the empty partition, corresponding to the trivial GL_k-representation). Note that each row and column naturally corresponds to a (classical) Littlewood–Richardson coefficient; for example, the first row corresponds

to $c_{*,*}^{\alpha^+}$. To obtain the multiplicity $\mathrm{LRC}_{\alpha,\beta}^{\gamma}$, multiply these six Littlewood–Richardson coefficients together, and then sum over all such tables. We share an easy example below.

Example 7.4. Let $\alpha = (1, 0, \ldots, 0)$ and $\beta = (0, \ldots, 0, -1)$. Hence $F_k^\alpha \cong \mathbb{C}^k$ is the defining representation of GL_k, while F_k^β is its dual $(\mathbb{C}^k)^*$. Then $F_k^\alpha \otimes F_k^\beta \cong \mathrm{M}_k \cong \mathfrak{gl}_k$, the adjoint representation of GL_k. We have $\alpha^+ = (1)$, $\alpha^- = \beta^+ = \mathbf{0}$, and $\beta^- = (1)$. Treating γ as the unknown, there are only two ways of filling the contingency table so that the product of Littlewood–Richardson coefficients is nonzero, namely

$$\begin{array}{c} \\ (1) \\ \mathbf{0} \\ \gamma^- \end{array}\!\!\begin{array}{c} \begin{array}{ccc} \mathbf{0} & (1) & \gamma^+ \end{array} \\ \begin{bmatrix} \mathbf{0} & \mathbf{0} & (1) \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (1) & \mathbf{0} \end{bmatrix} \end{array} \quad \text{and} \quad \begin{array}{c} \\ (1) \\ \mathbf{0} \\ \gamma^- \end{array}\!\!\begin{array}{c} \begin{array}{ccc} \mathbf{0} & (1) & \gamma^+ \end{array} \\ \begin{bmatrix} \mathbf{0} & (1) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \end{array}.$$

The first table forces $\gamma = [(1), (1)] = (1, 0, \ldots, 0, -1)$, and the second table forces $\gamma = \mathbf{0}$; in each case, by a simple application of the Pieri rule, all six Littlewood–Richardson coefficients are 1. Hence $F_k^\alpha \otimes F_k^\beta = F_k^{(1,0,\ldots,0,-1)} \oplus F_k^{\mathbf{0}}$, or in other words, the adjoint representation $\mathfrak{gl}_k$ is the direct sum $\mathfrak{sl}_k \oplus \mathbb{C}I_k$.

As an immediate advantage, which the reader can check, we can rewrite (7.14) as a single generalized Littlewood–Richardson coefficient:

$$B(\delta, \eta) = \mathrm{LRC}_{\delta^p,\,[0,\delta^q]}^{[\eta^p,\eta^q]}$$

with the $[\ ,\ ]$ notation as in (4.3). We will use these generalized coefficients again in Theorem 7.10 to write down a combinatorial expression for Blattner's formula with two noncompact simple roots.

7.4 Main result

It turns out that the classical decomposition in (4.15) is related, in a very different way, to certain discrete series representations *outside* the Hermitian symmetric case outlined in Section 7.2. In fact, the thrust of our main result is that we can read off the character

theory of $\mathbb{C}[V]$ — in particular the integers $\varepsilon^{\lambda}_{\mu,\nu}$ and hence the Enright resolutions of the modules $\widetilde{F}^{\lambda}_{p,q}$ — from the generating function $\mathbf{b}(0)$ for Blattner's formula for the group $\mathrm{SU}(k, p+q)$, with *two* noncompact simple roots. To distinguish this new setting from the Hermitian symmetric setting of Section 7.2, we will now decorate all of our notation with prime symbols; a complete summary is found in Table 6, accompanied by illustrations in Figure 15.

	Sections 4.4 and 7.2		Sections 7.4 and 7.5	
Classical representation	V	$\mathrm{M}_{k,p} \oplus \mathrm{M}_{k,q}$	–	–
Group	G_0	$\mathrm{SU}(p,q)$	G_0'	$\mathrm{SU}(k, p+q)$
Compl. Lie algebra	$\mathfrak{g}$	$\mathfrak{sl}_{p+q}$	$\mathfrak{g}'$	$\mathfrak{sl}_{p+k+q}$
Noncomp. simple roots	Π_{nc}	$\{\alpha_p\}$	Π'_{nc}	$\{\alpha_p, \alpha_{p+k}\}$
Compl. max. compact	$\mathfrak{k}$	$\mathfrak{s}(\mathfrak{gl}_p \oplus \mathfrak{gl}_q)$	$\mathfrak{k}'$	$\mathfrak{s}(\mathfrak{gl}_k \oplus \mathfrak{gl}_{p+q})$
Levi subalgebra	–	–	$\mathfrak{m}'$	$\mathfrak{s}(\mathfrak{gl}_k \oplus \mathfrak{gl}_p \oplus \mathfrak{gl}_q)$
$\bigoplus$ pos nc root spaces	$\mathfrak{u}^+$	$\mathrm{M}_{p,q}$	$\mathfrak{u}'^+$	$\mathrm{M}_{p,k} \oplus \mathrm{M}_{k,q}$
$\bigoplus$ neg nc root spaces	$\mathfrak{u}^-$	$\mathrm{M}_{q,p}$	$\mathfrak{u}'^-$	$\mathrm{M}_{k,p} \oplus \mathrm{M}_{q,k}$

Table 6: Compendium of notation for Chapter 7.

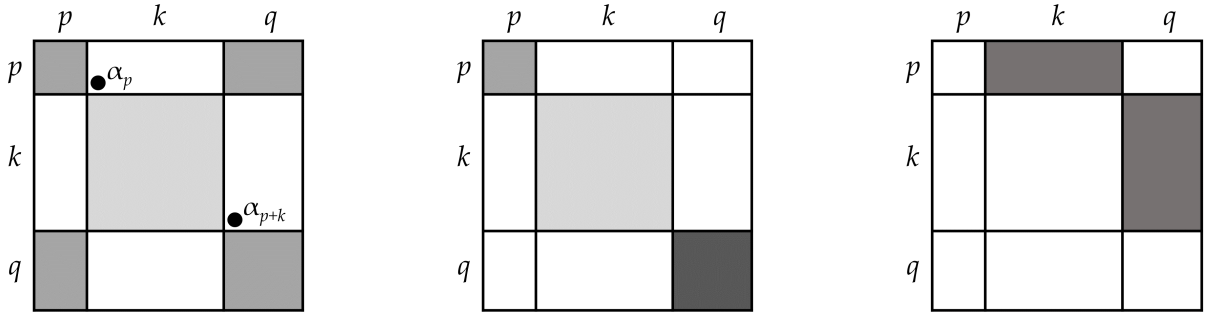


Figure 15: A visual companion to the notation in Table 6; regard as subspaces of $\mathfrak{g}' = \mathfrak{sl}_{p+k+q}$.

Let $G_0' = \mathrm{SU}(k, p+q)$, whose complexified Lie algebra is $\mathfrak{g}' = \mathfrak{sl}_{p+k+q}$. Suppose that $\mathfrak{g}'$ has two noncompact simple roots, namely $\Pi'_{\mathrm{nc}} = \{\alpha_p, \alpha_{p+k}\}$. This implies that $\mathfrak{g}'$ has maximal compact subalgebra $\mathfrak{k}' = \mathfrak{s}(\mathfrak{gl}_k \oplus \mathfrak{gl}_{p+q})$, where the $\mathfrak{gl}_k$ is embedded in the "middle" and $\mathfrak{gl}_{p+q}$ is embedded in the "four corners." Hence the direct sum $\mathfrak{u}'^+$ of the positive noncompact root spaces is embedded in $\mathfrak{g}'$ as the two blocks in the upper-right, so

that $\mathfrak{u}'^{+} \cong \mathrm{M}_{p,k} \oplus \mathrm{M}_{k,q}$ as a vector space. Likewise, the sum $\mathfrak{u}'^{-}$ of the negative noncompact root spaces is embedded in the two corresponding blocks in the lower-left, so that $\mathfrak{u}'^{-} \cong \mathrm{M}_{k,p} \oplus \mathrm{M}_{q,k}$ as a vector space. Special importance will be played by the Levi subalgebra $\mathfrak{m}' = \mathfrak{s}(\mathfrak{gl}_k \oplus \mathfrak{gl}_p \oplus \mathfrak{gl}_q) \subset \mathfrak{k}'$, which is the Lie algebra of $M' = \mathrm{S}(\mathrm{GL}_k \times \mathrm{GL}_p \times \mathrm{GL}_q)$.

We will write a weight of $\mathfrak{g}'$ as a triple

$$[\![\lambda, \mu, \nu]\!] := (\underbrace{-\mu_p, \dots, -\mu_1}_{\mu^*}, \underbrace{\lambda_1, \dots, \lambda_k}_{\lambda}, \underbrace{\nu_1, \dots, \nu_q}_{\nu}) \tag{7.16}$$

where $\lambda \in P(\mathfrak{gl}_k)$, $\mu \in P(\mathfrak{gl}_p)$, and $\nu \in P(\mathfrak{gl}_q)$. The resulting $(p+k+q)$-tuple on the right-hand side of (7.16) is written in terms of the standard coordinates $\varepsilon_i : \mathrm{diag}[h_1, \dots, h_{p+k+q}] \longmapsto h_i$. Notice that the order of the three weights in $[\![\lambda, \mu, \nu]\!]$ follows the alphabetical order k, p, q, whereas the explicit tuple transposes λ and μ in order to respect the embedding of $\mathfrak{m}'$ in $\mathfrak{g}'$. The reason for the dual on μ will soon become apparent, in the proof of Lemma 7.8: in order to line up the actions of M' in the Howe duality setting and in the Blattner setting, we will need to regard $\mathfrak{gl}_p$ as being embedded in $\mathfrak{g}'$ with a twist (i.e., negative transpose).

A weight $[\![\lambda, \mu, \nu]\!] \in P(\mathfrak{g}')$ is $\mathfrak{k}'$-dominant if and only if

$$\lambda_1 \geqslant \cdots \geqslant \lambda_k \qquad \text{and} \qquad -\mu_p \geqslant \cdots \geqslant -\mu_1 \geqslant \nu_1 \geqslant \cdots \geqslant \nu_q.$$

We will need this fact only in the proof of Theorem 7.10. For most of this paper, we will actually be concerned with $\mathfrak{m}'$-dominant weights, in which we drop the condition $-\mu_1 \geqslant \nu_1$; in other words, $\lambda \in P_+(\mathfrak{gl}_k)$ and $\mu \in P_+(\mathfrak{gl}_p)$ and $\nu \in P_+(\mathfrak{gl}_q)$. Note that for any λ, μ, ν satisfying the conditions (7.8), the weight $[\![\lambda, \mu, \nu]\!]$ is automatically $\mathfrak{m}'$-dominant.

We now state the main result of this chapter, uniting the Enright resolutions of Section 4.4 with Blattner's formula for $G_0' = \mathrm{SU}(n, p+q)$ in the present section.

Theorem 7.5. *Let λ, μ, ν satisfy the conditions* (7.8)*, with $\varepsilon^{\lambda}_{\mu,\nu}$ as in Definition 7.1. Let $B(\ ,\ )$ denote Blattner's formula as defined in* (7.12)*, for the special case described in the rightmost column*

of Table 6, and with $[\![-,-,-]\!]$ *as defined in* (7.16)*. Then*

$$\varepsilon^{\lambda}_{\mu,\nu} = B\big(0,\, [\![\lambda,\mu,\nu]\!]\big).$$

We reserve the proof for Section 7.5, opting first to provide some concrete examples. By (7.13), Theorem 7.5 implies that $\varepsilon^{\lambda}_{\mu,\nu}$ equals the coefficient of $e^{[\![\lambda,\mu,\nu]\!]}$ in $\mathbf{b}(0)$.

Example 7.6. We work out the case when $k = p = q = 1$ in full detail. Note that this is outside the stable range. In this case, $\mathbb{C}[V] = \mathbb{C}[x,y]$, and the action of $M' = \mathrm{S}(\mathrm{GL}_1 \times \mathrm{GL}_1 \times \mathrm{GL}_1)$ upon a typical monomial follows from (7.2):

$$(g_k, g_p, g_q) \cdot x^a y^b = \big(g_k g_p x\big)^a \left(\frac{g_q}{g_k} y\right)^b = \left(g_k^{a-b} g_p^a g_q^b\right) x^a y^b,$$

where $g_k, g_p, g_q \in \mathbb{C}^\times$. Hence each monomial spans a one-dimensional subrepresentation of M'. Note that by setting $\lambda = a - b$ and $c = \min(a,b)$, we can rewrite the formal sum of these monomials as

$$\begin{aligned}\sum_{a,b\in\mathbb{N}} x^a y^b &= \sum_{c=0}^{\infty} (xy)^c \left(\sum_{\lambda\geqslant 0} x^\lambda + \sum_{\lambda<0} y^{-\lambda}\right) \\ &= \sum_{\lambda\geqslant 0} x^\lambda \sum_{c=0}^{\infty} (xy)^c + \sum_{\lambda<0} y^{-\lambda} \sum_{c=0}^{\infty} (xy)^c.\end{aligned}$$

When $k = p = q = 1$, each rational representation of GL_k and GL_p and GL_q is one-dimensional, indexed by a single integer λ (nonnegative if the representation is polynomial), where the group action is $g \cdot z = g^\lambda z$. Hence we will write $\mathbb{C}_\lambda$ for the representation F_1^λ of $\mathrm{GL}_1 \cong \mathbb{C}^\times$. Comparing the two calculations above, we see that the Howe decomposition (4.15) in this case is

$$\mathbb{C}[x,y] = \bigoplus_{\lambda\in\mathbb{Z}} \mathbb{C}_\lambda \otimes \widetilde{F}^\lambda_{1,1}$$

where

$$\widetilde{F}^\lambda_{1,1} = \begin{cases} \mathbb{C}[xy] \otimes (\mathbb{C}_\lambda \otimes \mathbb{C}_0) = M_{\lambda,0}, & \lambda \geqslant 0, \\ \mathbb{C}[xy] \otimes (\mathbb{C}_0 \otimes \mathbb{C}_{-\lambda}) = M_{0,-\lambda}, & \lambda < 0. \end{cases}$$

Therefore the Enright resolution of $\widetilde{F}^{\lambda}_{1,1}$ contains only one term, and so for $\lambda, \mu, \nu \in \mathbb{Z}$ we have

$$\varepsilon^{\lambda}_{\mu,\nu} = \begin{cases} 1, & \lambda = \mu \geqslant 0 \text{ and } \nu = 0, \\ 1, & \lambda = -\nu < 0 \text{ and } \mu = 0, \\ 0 & \text{otherwise,} \end{cases}$$

so that the triples (λ, μ, ν) satisfying the first two cases are of the form

$$\mathbb{N}(1,1,0) \text{ and } \mathbb{N}(-1,0,1). \tag{7.17}$$

Next we compute $\mathbf{b}(0)$ in order to check it against (7.17). We have $\mathfrak{g}' = \mathfrak{sl}_3$, with $\Pi'_{\mathrm{nc}} = \{\alpha_1, \alpha_2\} = \Pi'$. Then $\Phi'^{+}_{\mathrm{c}} = \{\varepsilon_1 - \varepsilon_3\}$ and $\Phi'^{+}_{\mathrm{nc}} = \{\varepsilon_1 - \varepsilon_2, \varepsilon_2 - \varepsilon_3\}$. Recall that the triple $[\![\ell, m, n]\!]$ corresponds to the $\mathfrak{g}'$-weight $(-m, \ell, n)$ in standard ε-coordinates. Setting $t_i = e^{\varepsilon_i}$ for $i = 1, 2, 3$, we have

$$\begin{aligned} \mathbf{b}(0) = \frac{1 - \frac{t_3}{t_1}}{\left(1 - \frac{t_2}{t_1}\right)\left(1 - \frac{t_3}{t_2}\right)} &= \frac{1}{1 - \frac{t_2}{t_1}} + \frac{\frac{t_3}{t_2}}{1 - \frac{t_3}{t_2}} \\ &= \sum_{k=0}^{\infty} e^{k(\varepsilon_2 - \varepsilon_1)} + \sum_{k=1}^{\infty} e^{k(\varepsilon_3 - \varepsilon_2)} \\ &= \sum_{k=0}^{\infty} e^{k(-1,1,0)} + \sum_{k=1}^{\infty} e^{k(0,-1,1)} \\ &= \sum_{k=0}^{\infty} e^{k[\![1,1,0]\!]} + \sum_{k=1}^{\infty} e^{k[\![-1,0,1]\!]}, \end{aligned}$$

coinciding exactly with (7.17) and Theorem 7.5. (See Figure 16 for a visualization of $\mathbf{b}(0)$, in which we plot the support of $B(0, -)$ on the weight lattice of $\mathfrak{sl}_3$.)

Before presenting a more interesting example, we outline how Theorem 7.5 allows us to write down explicitly the Enright resolution for $\widetilde{F}^{\lambda}_{p,q}$. Note that the theorem could just

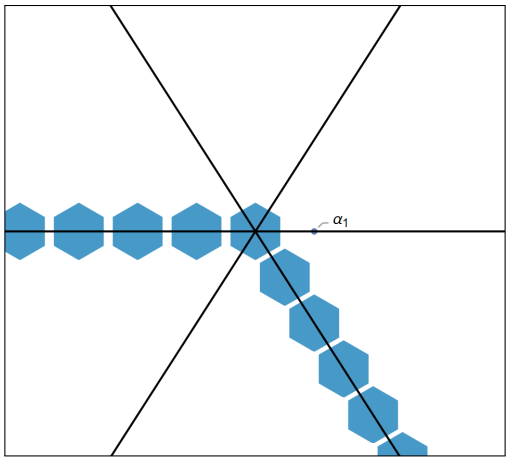


Figure 16: A visualization of $\mathbf{b}(0)$ from Example 7.6. By programming Blattner's formula (7.12) directly in Mathematica, we plot a hexagon at each weight $\eta \in P(\mathfrak{sl}_3)$ such that $B(0,\eta) = 1$. Note that these weights are the nonnegative multiples of $-\alpha_1 = (-1,1,0)$ and of $-\alpha_2 = (0,-1,1)$. (All other weights return 0.)

as well be rewritten as the following (in fact, this will be the last step (7.22) of the proof):

$$\begin{aligned}\mathbf{b}(0) &= \sum_{\substack{\lambda,\mu,\nu \\ \text{as in (7.8)}}} \varepsilon^{\lambda}_{\mu,\nu}\, e^{[\![\lambda,\mu,\nu]\!]} + \sum_{\xi \notin P_+(\mathfrak{m}')} c_\xi e^\xi, && c_\xi \in \mathbb{Z} \\ &= \sum_{\lambda} e^{[\![\lambda,0,0]\!]} \underbrace{\left(\sum_{\mu,\nu} \varepsilon^{\lambda}_{\mu,\nu} \cdot e^{[\![0,\mu,\nu]\!]} \right)}_{:= \text{ ``coefficient'' of } e^\lambda} + \sum_{\xi \notin P_+(\mathfrak{m}')} c_\xi e^\xi. && (7.18)\end{aligned}$$

Therefore we begin by computing $\mathbf{b}(0)$, expanding to a sufficiently high order, and then ignoring all terms corresponding to nondominant $\mathfrak{m}'$-weights. Then given λ, we should collect all the remaining terms in $\mathbf{b}(0)$ of the form $e^{[\![\lambda,*,*]\!]}$ (where the stars are arbitrary), and then factor out $e^{[\![\lambda,0,0]\!]}$ to obtain the multi-term "coefficient" of e^λ indicated in (7.18). The terms inside this coefficient tell us exactly which generalized Verma modules $M_{\mu,\nu}$ appear in the Enright resolution of $\widetilde{F}^{\lambda}_{p,q}$, along with appropriate signs depending on the parity of the term.

Remark. In order to recover the complete data of the Enright resolution, we clearly need to supplement the method outlined above so as to determine the exact term (rather than just the parity) in which each generalized Verma module occurs. Although not *a priori* obvious, it can be seen from the construction in [EW04] that as we move from right to left

in the resolution, the partitions μ and ν strictly increase in size. This fact will allow us to easily recover the ordering of the terms once we have found the coefficient of e^λ.

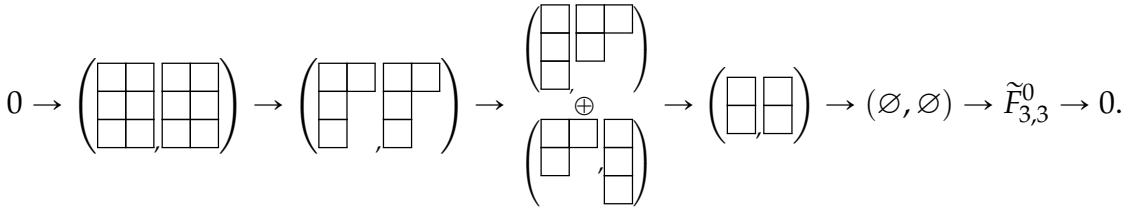
Example 7.7. We use software to present an example of the method outlined above. On one hand, we will compute the terms of the Enright resolution of $\widetilde{F}^\lambda_{p,q}$ directly, using Maple code written by Jeb Willenbring. On the other hand, we will expand $\mathbf{b}(0)$ and isolate the coefficient of e^λ, using Mathematica code written by the author of the thesis.

Let $k = 1$, with $p = q = 3$. Set $\lambda = 0$, the empty partition; then $\widetilde{F}^0_{3,3} = \mathbb{C}[V]^{\mathrm{GL}_1}$ is the invariant subalgebra of $\mathbb{C}[V]$ in the setting of Section 4.4. This example is of special interest since it is the first Wallach representation of $\mathfrak{su}(3,3)$, from Section 4.3. See Enright and Hunziker's paper [EH04] on minimal resolutions for the Wallach representations; in fact, our present example is exactly Example 7.11 in [EHP14] where $p = 3$.

In Maple, we compute the following resolution for $\widetilde{F}^0_{3,3}$, with $M_{\mu,\nu}$ as in (7.7):

$$0 \to M_{(2,2,2),(2,2,2)} \to M_{(2,1,1),(2,1,1)} \to M_{(1,1,1),(2,1,0)} \oplus M_{(2,1,0),(1,1,1)}$$
$$\to M_{(1,1,0),(1,1,0)} \to M_{(0,0,0),(0,0,0)} \to \widetilde{F}^0_{3,3} \to 0.$$

We can also write this more compactly with Young diagrams:



In Mathematica, we define the generating function $\mathbf{b}(0)$ directly from the definition (7.13), setting $x_i = e^{\varepsilon_i}$ for $i = 1,2,3$, and $w = e^{\varepsilon_4}$, and $y_i = e^{\varepsilon_{i+4}}$ for $i = 1,2,3$. Note that the alphabetical order w, x, y mirrors that of k, p, q as we visualize the embedding $\mathfrak{m}' \subset \mathfrak{g}'$; in this way, the exponent vector for the variables x_i encodes (the dual of) a weight of $\mathfrak{gl}_p$, the exponent vector for the y_i encodes a weight of $\mathfrak{gl}_q$, and the exponent of w encodes a

weight of $\mathfrak{gl}_k$. Using (7.13), we obtain

$$\mathbf{b}(0) = \frac{\displaystyle\prod_{1\leqslant i<j\leqslant 3}\left(1-\frac{x_j}{x_i}\right)\prod_{1\leqslant i<j\leqslant 3}\left(1-\frac{y_j}{y_i}\right)\prod_{1\leqslant i,j\leqslant 3}\left(1-\frac{y_j}{x_i}\right)}{\displaystyle\prod_{1\leqslant i\leqslant 3}\left(1-\frac{w}{x_i}\right)\prod_{1\leqslant i\leqslant 3}\left(1-\frac{y_i}{w}\right)}.$$

Upon expanding $\mathbf{b}(0)$ to a sufficiently high order, we program Mathematica to retain only those terms in which the exponent vectors for (x_1, x_2, x_3) and (y_1, y_2, y_3) are both weakly decreasing, corresponding to dominant weights for $\mathfrak{gl}_3$. (Since $k = 1$ in this example, there is no need to do the same for the lone variable w.) In the remaining sum, we then find the "coefficient" of e^λ, namely, of w^0 — that is, we collect all terms in which the power of w is 0. This "coefficient" is the following sum of terms in the variables x_i and y_i; in light of the remark before this example, we write the terms in descending order with respect to total degree in the y_i:

$$\frac{y_1^2y_2^2y_3^2}{x_1^2x_2^2x_3^2} - \frac{y_1^2y_2y_3}{x_1x_2x_3^2} + \frac{y_1^2y_2}{x_1x_2x_3} + \frac{y_1y_2y_3}{x_2x_3^2} - \frac{y_1y_2}{x_2x_3} + 1$$

$$\begin{aligned} = \quad & e^{(-2,-2,-2,0,2,2,2)} - e^{(-1,-1,-2,0,2,1,1)} + e^{(-1,-1,-1,0,2,1,0)} + e^{(0,-1,-2,0,1,1,1)} \\ & - e^{(0,-1,-1,0,1,1,0)} + e^{(0,0,0,0,0,0,0)} \end{aligned}$$

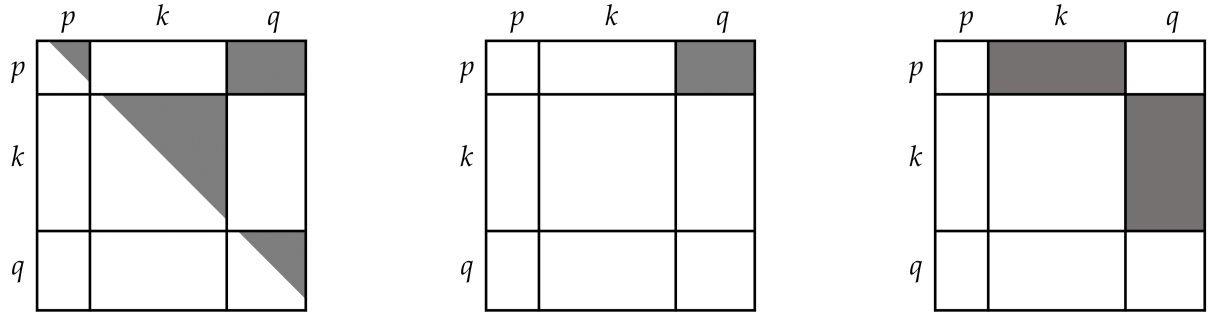
$$\begin{aligned} = \quad & e^{[\![0,(2,2,2),(2,2,2)]\!]} - e^{[\![0,(2,1,1),(2,1,1)]\!]} + e^{[\![0,(1,1,1),(2,1,0)]\!]} + e^{[\![0,(2,1,0),(1,1,1)]\!]} \\ & - e^{[\![0,(1,1,0),(1,1,0)]\!]} + e^{[\![0,(0,0,0),(0,0,0)]\!]}. \end{aligned}$$

We thus arrive at the Enright resolution produced by Maple.

7.5 Proof of Theorem 7.5

All notation remains as in Section 7.4. We will prove Theorem 7.5 by rewriting the closed form for $\mathbf{b}(0)$ in (7.13), in terms of the character theory of $\mathbb{C}[V]$ from Section 4.4. To this end, we partition the positive compact roots $\Phi_c'^+$ into two subsets, as in Figures

17a and 17b. Let $\Phi_{\mathfrak{m}'} = \{\alpha \in \Phi_c'^+ \mid \mathfrak{g}'_\alpha \subset \mathfrak{m}'\}$, which contains those positive compact roots whose root spaces span the three triangular regions in Figure 17a. Let $\overline{\Phi_{\mathfrak{m}'}}$ denote the complement of $\Phi_{\mathfrak{m}'}$ in $\Phi_c'^+$, which contains those positive compact roots whose root spaces span the upper-right $p \times q$ block in Figure 17b.



(a) $\Phi_{\mathfrak{m}'}$ (b) $\overline{\Phi_{\mathfrak{m}'}}$ (c) $\Phi_{\mathrm{nc}}'^+$

Figure 17: Root spaces corresponding to three subsets of Φ'^+.

Now define the products

$$\Delta_{\mathfrak{m}'} = \prod_{\alpha \in \Phi_{\mathfrak{m}'}} 1 - e^{-\alpha}, \qquad \overline{\Delta_{\mathfrak{m}'}} = \prod_{\alpha \in \overline{\Phi_{\mathfrak{m}'}}} 1 - e^{-\alpha}, \qquad \Delta_{\mathfrak{u}'^+} = \prod_{\alpha \in \Phi_{\mathrm{nc}}'^+} 1 - e^{-\alpha},$$

where the "Δ" notation is meant to evoke the Weyl denominator from the Weyl character formula. Now we can rewrite (7.13) as

$$\mathbf{b}(0) = \frac{\prod_{\alpha \in \Phi_c^+} 1 - e^{-\alpha}}{\prod_{\alpha \in \Phi_{\mathrm{nc}}^+} 1 - e^{-\alpha}} = \frac{\Delta_{\mathfrak{m}'} \cdot \overline{\Delta_{\mathfrak{m}'}}}{\Delta_{\mathfrak{u}'^+}}. \tag{7.19}$$

The following two lemmas capture the connection between the expression for $\mathbf{b}(0)$ in (7.19) and the Howe duality setting from Section 4.4.

Lemma 7.8. *Consider* $\mathbb{C}[\mathrm{M}_{p,q}]$ *from the Howe duality setting in Section 4.4. Then* $\overline{\Delta_{\mathfrak{m}'}} = \left(\operatorname{ch} \mathbb{C}[\mathrm{M}_{p,q}]\right)^{-1}$ *as a character of* $\mathfrak{m}'$.

Proof. Define the block "upper-right" embedding

$$\psi : \mathrm{M}_{p,q} \longrightarrow \bigoplus_{\alpha \in \overline{\Phi_{\mathfrak{m}'}}} \mathfrak{g}'_\alpha,$$

$$E_{i,j} \longmapsto E_{i,p+k+j}$$

for $i = 1, \ldots, p$ and $j = 1, \ldots, q$. (Here $E_{i,j}$ denotes the matrix with 1 in the (i, j) position and 0's elsewhere.) Clearly ψ is a vector space isomorphism; we claim that ψ is in fact an isomorphism of $\mathfrak{m}'$-modules.

To see this, recall the action of $K = \mathrm{S}(\mathrm{GL}_p \times \mathrm{GL}_q)$ on $\mathrm{M}_{p,q}$ given in (7.5) in the Howe duality setting, where $(g, h) \cdot X = g^{-T} X h^{-1}$. Here in the Blattner setting, we also have $K \subset M' \subset G' = \mathrm{SL}_{p+k+q}$ embedded block-diagonally with GL_p in the upper-left and GL_q in the lower-right. In this embedding, $\mathrm{im}(\psi) \subset \mathfrak{g}'$ is a K-module via the adjoint action, and the explicit action is $(g, h) \cdot X = gXh^{-1}$. Therefore, the K-actions in the Howe duality setting and in the Blattner setting are the same, up to a twist in the GL_p-action. (This is remedied by embedding GL_p in the upper-left via its inverse transpose, and at the Lie algebra level, by embedding $\mathfrak{gl}_p$ via its negative transpose.) Extending this K-action to M' by letting the factor GL_k act trivially, we conclude that $\mathrm{M}_{p,q} \cong \mathrm{im}(\psi)$ as modules for M', and thus for $\mathfrak{m}'$, which proves the claim.

Now, $\mathrm{im}(\psi)^T := \{X^T \mid X \in \mathrm{im}(\psi)\}$ is the embedding of $\mathrm{M}_{p,q}$ via the transpose into the lower-left block of $\mathfrak{g}'$. Furthermore, $\mathrm{im}(\psi)^T \cong \mathrm{im}(\psi)^*$ as an M'-module, which is clear from the adjoint action of $M' \subset G'$ on $\mathfrak{g}'$. Therefore

$$S\big(\mathrm{im}(\psi)^T\big) \cong \mathbb{C}[\mathrm{im}(\psi)] \cong \mathbb{C}[\mathrm{M}_{p,q}]$$

as $\mathfrak{m}'$-modules. Now we can conclude that

$$\begin{aligned} \overline{\Delta_{\mathfrak{m}'}} &= \prod_{\alpha \in \overline{\Phi_{\mathfrak{m}'}}} (1 - e^{-\alpha}) \\ &= \Big(\mathrm{ch}\, S\big(\mathrm{im}(\psi)^T\big) \Big)^{-1} \\ &= \big(\mathrm{ch}\, \mathbb{C}[\mathrm{M}_{p,q}] \big)^{-1} \end{aligned}$$

as a character of $\mathfrak{m}'$. □

Lemma 7.9. *We have* $(\Delta_{\mathfrak{u}'^+})^{-1} = \mathrm{ch}\, \mathbb{C}[V]$ *as a character of* $\mathfrak{m}'$.

Proof. Let $\varphi : V \longrightarrow \mathfrak{u}'^+$ be given by $\varphi(X, Y) = (X^T, Y)$. We claim that φ is an isomorphism of $\mathfrak{m}'$-modules.

To see this, note that an element of V is of the form (X, Y), while an element of $\mathfrak{u}'^+$ is of the form (X^T, Y), where $X \in \mathrm{M}_{k,p}$ and $Y \in \mathrm{M}_{k,q}$. Recall the action of M' on V given in (7.2), from the Howe duality setting. Here in the Blattner setting (as explained in the proof of Lemma 7.8), regard GL_p as being embedded in the upper-left block of G' via inverse transpose; then the adjoint action of $M' \subset G'$ on $\mathfrak{u}'^+ \subset \mathfrak{g}'$ is given by

$$(g_k, g_p, g_q) \cdot (X^T, Y) = (g_p^{-T} X^T g_k^{-1},\ g_k Y g_q^{-1}).$$

For $g = (g_k, g_p, g_q) \in M'$, we must show that $g \circ \varphi = \varphi \circ g$. But

$$\begin{aligned} g \circ \varphi(X, Y) &= g(X^T, Y) \\ &= (g_p^{-T} X^T g_k^{-1},\ g_k Y g_q^{-1}) \\ &= \varphi(g_k^{-T} X g_p^{-1},\ g_k Y g_q^{-1}) \\ &= \varphi \circ g(X, Y), \end{aligned}$$

by (7.2), which proves the claim.

Now, observing from the adjoint M'-action that $\mathfrak{u}'^+ \cong (\mathfrak{u}'^-)^*$ as M'-modules, we have

$$S(\mathfrak{u}'^-) \cong \mathbb{C}[V]$$

as modules for M', and therefore for $\mathfrak{m}'$. Therefore we have

$$\begin{aligned} (\Delta_{\mathfrak{u}'^+})^{-1} &= \prod_{\alpha \in \Phi^+_{\mathrm{nc}}} \frac{1}{1 - e^{-\alpha}} \\ &= \mathrm{ch}\, S(\mathfrak{u}'^-) \\ &= \mathrm{ch}\, \mathbb{C}[V] \end{aligned}$$

as a character of $\mathfrak{m}'$. □

At this point, we should observe that for $\xi \in P_+(\mathfrak{m}')$, the Weyl character formula can be written as

$$\mathrm{ch}\, L_{\mathfrak{m}'}(\xi) = \frac{\sum_{w \in W_{\mathfrak{m}'}} (-1)^{\ell(w)} e^{w(\xi+\rho)}}{e^{\rho} \prod_{\alpha \in \Phi_{\mathfrak{m}'}} 1 - e^{-\alpha}} = \frac{\sum_{w \in W_{\mathfrak{m}'}} (-1)^{\ell(w)} e^{w(\xi+\rho)-\rho}}{\Delta_{\mathfrak{m}'}},$$

where $W_{\mathfrak{m}'}$ is the Weyl group for $\mathfrak{m}'$ and $\rho = \frac{1}{2}\sum_{\alpha\in\Phi_{\mathfrak{m}'}}\alpha$. Upon rearranging, this says that the product $\Delta_{\mathfrak{m}'}\cdot \operatorname{ch} L_{\mathfrak{m}'}(\xi)$ is the alternating sum of terms of the form $e^{w(\xi+\rho)-\rho}$. But $e^{w(\xi+\rho)-\rho}\in P_+(\mathfrak{m}')$ if and only if $w=1$, which means that

$$\Delta_{\mathfrak{m}'}\cdot \operatorname{ch} L_{\mathfrak{m}'}(\xi) = e^{\xi} + \text{alternating sum of } e^{\text{nondominant } \mathfrak{m}'\text{-weight}}\text{'s}.$$

More generally, consider an arbitrary $\mathfrak{m}'$-module $L = \bigoplus_{\xi} m_\xi L_{\mathfrak{m}'}(\xi)$, ranging over $\xi \in P_+(\mathfrak{m}')$, with multiplicities $m_\xi \in \mathbb{N}$. Then we have

$$\Delta_{\mathfrak{m}'}\cdot \operatorname{ch} L = \sum_{\xi} m_\xi e^{\xi} + \text{alternating sum of } e^{\text{nondominant } \mathfrak{m}'\text{-weight}}\text{'s}. \tag{7.20}$$

The upshot is that multiplying the character of an $\mathfrak{m}'$-module L by $\Delta_{\mathfrak{m}'}$ produces a sum of formal weights, in which the $\mathfrak{m}'$-dominant weights are precisely the highest weights of the irreducible $\mathfrak{m}'$-modules in the decomposition of L, the coefficients of which are their multiplicities in L.

Proof of Theorem 7.5. From (7.19), we have

$$\mathbf{b}(0) = \frac{\Delta_{\mathfrak{m}'}\cdot\overline{\Delta_{\mathfrak{m}'}}}{\Delta_{\mathfrak{u}'+}}$$

which, by Lemmas 7.8 and 7.9, becomes

$$\mathbf{b}(0) = \Delta_{\mathfrak{m}'}\cdot(\operatorname{ch}\mathbb{C}[\mathrm{M}_{p,q}])^{-1}\cdot\operatorname{ch}\mathbb{C}[V].$$

Substituting for $\operatorname{ch}\mathbb{C}[V]$ from (7.11), we find that the two instances of $\operatorname{ch}\mathbb{C}[\mathrm{M}_{p,q}]$ cancel each other out:

$$\begin{aligned}\mathbf{b}(0) &= \Delta_{\mathfrak{m}'}\cdot(\operatorname{ch}\mathbb{C}[\mathrm{M}_{p,q}])^{-1}\cdot\operatorname{ch}\mathbb{C}[\mathrm{M}_{p,q}]\cdot\sum_{\lambda,\mu,\nu}\varepsilon^{\lambda}_{\mu,\nu}\cdot\operatorname{ch}(F^{\lambda}_{k}\otimes F^{\mu}_{p}\otimes F^{\nu}_{q})\\ &= \Delta_{\mathfrak{m}'}\cdot\sum_{\lambda,\mu,\nu}\varepsilon^{\lambda}_{\mu,\nu}\cdot\operatorname{ch}(F^{\lambda}_{k}\otimes F^{\mu}_{p}\otimes F^{\nu}_{q}).\end{aligned} \tag{7.21}$$

The sum in the last line is a virtual $\mathfrak{m}'$-character, and so by (7.20), we have

$$\mathbf{b}(0) = \sum_{\lambda,\mu,\nu}\varepsilon^{\lambda}_{\mu,\nu}\cdot e^{[\![\lambda,\mu,\nu]\!]} + \sum_{\xi\notin P_+(\mathfrak{m}')} c_\xi e^{\xi} \tag{7.22}$$

with $c_\xi\in\mathbb{Z}$. This completes the proof. □

The proof above suggests a combinatorial expression for Blattner's formula in Type A, similar to that in Proposition 7.3, but this time for *two* noncompact roots (under a stability condition). We record this as a final result of the thesis.

Theorem 7.10. *Suppose $k \geqslant p+q$. Let $\mathfrak{g}' = \mathfrak{sl}_{p+k+q}$ with $\Pi'_{\mathrm{nc}} = \{\alpha_p, \alpha_{p+k}\}$; hence $\mathfrak{k}' = \mathfrak{s}(\mathfrak{gl}_k \oplus \mathfrak{gl}_{p+q})$, just as in Section 7.4. Let $\delta, \eta \in P_+(\mathfrak{k}')$, where $\delta = [\![\delta^k, \delta^p, \delta^q]\!]$ and $\eta = [\![\eta^k, \eta^p, \eta^q]\!]$, and write $[\![\delta^p, \delta^q]\!] := (-\delta^p_p, \dots, -\delta^p_1, \delta^q_1, \dots, \delta^q_q) \in P_+(\mathfrak{gl}_{p+q})$. Then*

$$B(\delta, \eta) = \sum_{\lambda,\mu,\nu} \mathrm{LRC}^{[\![\delta^p,\delta^q]\!]}_{\mu,\nu} \mathrm{LRC}^{\eta^k}_{\delta^k,\lambda} \mathrm{LRC}^{\eta^p}_{\mu,\lambda^+} \mathrm{LRC}^{\eta^q}_{\nu,\lambda^-}$$

where the sum is over all λ, μ, ν satisfying (7.8)*.*

Proof. Using (7.21) to substitute for $\mathbf{b}(0)$, and then (7.10) to simplify since we are in the stable range, we calculate that

$$\begin{aligned}
\mathbf{b}(\delta) &= \operatorname{ch} L_{\mathfrak{k}'}(\delta) \cdot \mathbf{b}(0) \\
&= \operatorname{ch}\left(F^{\delta^k}_k \otimes F^{[\![\delta^p,\delta^q]\!]}_{p+q}\right) \cdot \left(\Delta_{\mathfrak{m}'} \cdot \sum_{\lambda,\alpha,\beta} \varepsilon^{\lambda}_{\alpha,\beta} \cdot \operatorname{ch} F^{\lambda}_k \otimes F^{\alpha}_p \otimes F^{\beta}_q\right) \\
&= \operatorname{ch}\left(F^{\delta^k}_k \otimes \bigoplus_{\mu,\nu} \mathrm{LRC}^{[\![\delta^p,\delta^q]\!]}_{\mu,\nu} F^{\mu}_p \otimes F^{\nu}_q\right) \cdot \left(\sum_{\lambda} \operatorname{ch} F^{\lambda}_k \otimes F^{\lambda^+}_p \otimes F^{\lambda^-}_q\right) \cdot \Delta_{\mathfrak{m}'} \\
&= \operatorname{ch}\left(\bigoplus_{\mu,\nu} \mathrm{LRC}^{[\![\delta^p,\delta^q]\!]}_{\mu,\nu} \left(F^{\delta^k}_k \otimes F^{\mu}_p \otimes F^{\nu}_q\right) \otimes \bigoplus_{\lambda} F^{\lambda}_k \otimes F^{\lambda^+}_p \otimes F^{\lambda^-}_q\right) \cdot \Delta_{\mathfrak{m}'} \\
&= \Delta_{\mathfrak{m}'} \cdot \operatorname{ch} \bigoplus_{\substack{\lambda,\ \mu,\ \nu,\\ \eta^k,\eta^p,\eta^q}} \mathrm{LRC}^{[\![\delta^p,\delta^q]\!]}_{\mu,\nu} \mathrm{LRC}^{\eta^k}_{\delta^k,\lambda} \mathrm{LRC}^{\eta^p}_{\mu,\lambda^+} \mathrm{LRC}^{\eta^q}_{\nu,\lambda^-} \left(F^{\eta^k}_k \otimes F^{\eta^p}_p \otimes F^{\eta^q}_q\right) \\
&= \sum_{\eta := [\![\eta^k,\eta^p,\eta^q]\!]} \left(\sum_{\lambda,\mu,\nu} \mathrm{LRC}^{[\![\delta^p,\delta^q]\!]}_{\mu,\nu} \mathrm{LRC}^{\eta^k}_{\delta^k,\lambda} \mathrm{LRC}^{\eta^p}_{\mu,\lambda^+} \mathrm{LRC}^{\eta^q}_{\nu,\lambda^-}\right) e^{\eta}
\end{aligned}$$

plus terms of the form $\_\_e^{\text{nondominant } \mathfrak{m}'\text{-weight}}$. Since $\mathfrak{m}' \subset \mathfrak{k}'$, nondominant $\mathfrak{m}'$-weights are necessarily nondominant $\mathfrak{k}'$-weights, and so we ignore all such nondominant terms as falling outside the hypothesis on η. Hence by the definition of $\mathbf{b}(\delta)$, we conclude that

$$B(\delta, \eta) = \sum_{\lambda,\mu,\nu} \mathrm{LRC}^{[\![\delta^p,\delta^q]\!]}_{\mu,\nu} \mathrm{LRC}^{\eta^k}_{\delta^k,\lambda} \mathrm{LRC}^{\eta^p}_{\mu,\lambda^+} \mathrm{LRC}^{\eta^q}_{\nu,\lambda^-}.$$

□

REFERENCES

[Abr75] M. Abramson. A simple solution of Simon Newcomb's problem. *J. Combin. Theory Ser. A*, 18:223–225, 1975.

[AMJJ18] D. Alvarez-Melis, T. Jaakkola, and S. Jegelka. Structured optimal transport. *Proceedings of Machine Learning Research*, 84:1771–1780, 2018.

[AP89] A. Aggarwal and J.K. Park. Sequential searching in multidimensional monotone arrays. In *Research Report RC 15128*. IBM T.J. Watson Research Center, Yorktown Heights, NY, 1989.

[BBPP95] W. Bein, P. Brucker, J. Park, and P. Pathak. A Monge property for the d-dimensional transport problem. *Discrete Appl. Math.*, 58(2):97–109, 1995.

[BL01] P. J. Bickel and E. Levina. The earth mover's distance is the Mallows distance: some insights from statistics. *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, 2:251–256, 2001.

[BS17] D. Bump and A. Schilling. *Crystal bases: representations and combinatorics*. World Scientific Publishing Co. Pte. Ltd., 2017.

[BW20] R. Bourn and J. Willenbring. Expected value of the one-dimensional earth mover's distance. *Algebr. Stat.*, 11(1):53–78, 2020.

[Cas09] F. Caselli. On the multivariate Robinson–Schensted correspondence. *Bollettino dell'Unione Matematica Italiana*, 9(1):591–602, 2009.

[CEW22] M. Colarusso, W. Erickson, and J. Willenbring. Contingency tables and the generalized Littlewood–Richardson coefficients. *Proc. Amer. Math. Soc.*, 150(1), 2022.

[Col19] H. Colakoğlu. On the distance formulae in the generalized taxicab geometry. *Turkish J. Math.*, 43(3):1578–1594, 2019.

[DES91] M. Davidson, T. Enright, and R. Stanke. Differential operators and highest weight representations. *Mem. Amer. Math. Soc.*, 94(455):1–102, 1991.

[DR69] J. Dillon and D. Roselle. Simon Newcomb's problem. *SIAM J. Appl. Math.*, 17(6):1086–1093, 1969.

[EH04] T. Enright and M. Hunziker. Resolutions and Hilbert series of determinantal varieties and unitary highest weight modules. *J. Algebra*, 273(2):608–639, 2004.

[EHP14] T. Enright, M. Hunziker, and W.A. Pruett. Diagrams of Hermitian type, highest weight modules, and syzygies of determinantal varieties. In *Symmetry: representation theory and its applications*, volume 257 of *Progr. Math.*, pages 121–184. Birkhäuser/Springer, New York, 2014.

[Eria] W. Erickson. The earth mover's distance as the symmetric difference of Young diagrams. arXiv:2104.07273. Under review.

[Erib] W. Erickson. Enright resolutions encoded by a generating function for Blattner's formula. arXiv:2108.08469.

[Eri21] W. Erickson. A generalization for the expected value of the earth mover's distance. *Algebr. Stat.*, 12(2):139–166, 2021. DOI 10.2140/astat.2021.12.139.

[EW04] T. Enright and J. Willenbring. Hilbert series, Howe duality and branching for classical groups. *Ann. of Math. (2)*, 159(1):337–375, 2004.

[FH91] W. Fulton and J. Harris. *Representation theory: a first course*, volume 129 of *Graduate Texts in Mathematics*. Springer–Verlag, 1991.

[FMM02] U. Frisch, S. Matarrese, and R. Mohayaee. A reconstruction of the initial conditions of the universe by optimal mass transportation. *Nature*, 417:260–262, 2002. https://doi.org/10.1038/417260a.

[Ful97] W. Fulton. *Young tableaux*. Cambridge University Press, 1997.

[GW09] R. Goodman and N. Wallach. *Symmetry, representations, and invariants*, volume 255 of *Graduate Texts in Mathematics*. Springer, 2009.

[HC66] Harish-Chandra. Discrete series for semisimple Lie groups II. Explicit determination of the characters. *Acta Math.*, 116:1–111, 1966.

[Her91] R. Herb. Harish–Chandra and his work. *Bull. Amer. Math. Soc. (N.S.)*, 25(1):1–17, 1991.

[HK02] J. Hong and S. Kang. *Introduction to quantum groups and crystal bases*, volume 42 of *Graduate Studies in Mathematics*. American Mathematical Society, 2002.

[Hof63] A. Hoffman. On simple linear programming problems. In V. Klee, editor, *Convexity: Proceedings of the Seventh Symposium in Pure Mathematics of the AMS*, pages 317–327. American Mathematical Society, Providence, RI, 1963.

[HS75] H. Hecht and W. Schmid. A proof of Blattner's conjecture. *Invent. Math.*, 31(2):129–154, 1975.

[HS80] L. Hansen and T. Sargent. Formulating and estimating dynamic linear rational expectations models. *J. Econom. Dynamics Control*, 2(1):7–46, 1980.

[HTW05] R. Howe, E.-C. Tan, and J. Willenbring. Stable branching rules for classical symmetric pairs. *Trans. Amer. Math. Soc.*, 357(4):1601–1626, 2005.

[Hum72] J. Humphreys. *Introduction to Lie algebras and representation theory*, volume 9 of *Graduate Texts in Mathematics*. Springer–Verlag, 1972.

[Jos92] A. Joseph. Annihilators and associated varieties of unitary highest weight modules. *Ann. Sci. École Norm. Sup. (4)*, 25(1):1–45, 1992.

[Kan42] L. Kantorovich. On mass transportation. (russian). *CR Doklady Acad. Sci. URSS NS*, 37(7-8), 1942.

[Kas91] M. Kashiwara. On crystal bases of the q-analogue of universal enveloping algebras. *Duke Math. J.*, 63(2):465–516, 1991.

[Kli19] J. Kline. Properties of the d-dimensional earth mover's problem. *Discrete Appl. Math.*, 265:128–141, 2019.

[KMT19] P. Komiske, E. Metodiev, and J. Thaler. Metric space of collider events. *Phys. Rev. Lett.*, 123:041801, 2019.

[Kna86] A. Knapp. *Representation theory of semisimple groups: an overview based on examples*. Princeton University Press, 1986.

[Kol41] A. N. Kolmogorov. Stationary sequences in hilbert's space. (russian). *Bolletin Moskovskogo Gosudarstvenogo Universiteta. Matematika*, 2(6):1–40, 1941.

[Kre20] J. Kretschmann. Earth mover's distance between grade distribution data with fixed mean. Master's thesis, University of Wisconsin–Milwaukee, 2020.

[KV78] M. Kashiwara and M. Vergne. On the Segal–Shale–Weil representations and harmonic polynomials. *Invent. Math.*, 44(1):1–47, 1978.

[Lit94] P. Littelmann. Littlewood–Richardson rule for symmetrizable Kac–Moody algebras. *Invent. Math.*, 116(1-3):329–346, 1994.

[LSW17] N. Lupu, L. Selios, and Z. Warner. A new measure of congruence: the earth mover's distance. *Political Analysis*, 25:95–113, 2017.

[Lus90] G. Lusztig. Canonical bases arising from quantized enveloping algebras. *J. Amer. Math. Soc.*, 3(2):447–498, 1990.

[Mac16] P. MacMahon. *Combinatory Analysis*, volume 2. Cambridge University Press, 1916.

[MKV+20] A. Melnyk, S. Knyazev, F. Vannberg, L. Bunimovich, P. Skums, and A. Zelikovsky. Using earth mover's distance for viral outbreak investigations. *BMC Genomics*, 5:582–590, 2020. https://doi.org/10.1186/s12864-020-06982-4.

[Mor13] M. Morales. Segre embeddings, Hilbert series, and Newcomb's problem. HAL ID: hal-00839652, 2013.

[PBRT99] J. Puzicha, J.M. Buhmann, Y. Rubner, and C. Tomasi. Empirical evaluation of dissimilarity measures for color and texture. In *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, volume 2, pages 1165–1172, 1999.

[RDG08] J. Rabin, J. Delon, , and Y. Gousseau. Circular earth mover's distance for the comparison of local features. In *Proceedings of the IEEE International Conference on Pattern Recognition (ICPR)*. IEEE Computer Society, 2008.

[RDG11] J. Rabin, J. Delon, and Y. Gousseau. Transportation distances on the circle. *J. Math. Imaging Vision*, 41(1-2):147–167, 2011.

[RTG98] Y. Rubner, C. Tomasi, and L.J. Guibas. A metric for distributions with applications to image databases. In *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, pages 59–66, 1998.

[Sta99] R. P. Stanley. *Enumerative combinatorics, volume 2*, volume 62 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 1999.

[Sta12] R. P. Stanley. *Enumerative combinatorics, volume 1*, volume 49 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 2 edition, 2012.

[Var89] V.S. Varadarajan. *An introduction to harmonic analysis on semisimple Lie groups*, volume 16 of *Cambridge studies in advanced mathematics*. Cambridge University Press, 1989.

[Vil09] C. Villani. *Optimal transport: old and new*, volume 338 of *Fundamental Principles of Mathematical Sciences*. Springer–Verlag, 2009.

[Vog79] D. Vogan. The algebraic structure of the representation of semisimple Lie groups I. *Ann. of Math. (2)*, 109(1):1–60, 1979.

[Wal79] N. Wallach. The analytic continuation of the discrete series, I and II. *Trans. Amer. Math. Soc.*, 251:1–37, 1979.

[Wey39] H. Weyl. *The Classical Groups. Their Invariants and Representations*. Princeton University Press, 1939.

[Wie47] H. Wiener. Structural determination of paraffin boiling points. *Journal of the American Chemical Society*, 69(1):17–20, 1947. doi:10.1021/ja01193a005.

[Wie49] N. Wiener. *Extrapolation, interpolation, and smoothing of stationary time series. With engineering applications*. The Technology Press of the Massachusetts Institute of Technology, 1949.

[WPR85] M. Werman, S. Peleg, and A. Rosenfeld. A distance metric for multidimensional histograms. *Computer Vision, Graphics, and Image Processing*, 32(3):328–336, 1985.

[WZ08] J. Willenbring and G. Zuckerman. A generating function for Blattner's formula. *Proc. Amer. Math. Soc.*, 136(6):2261–2270, 2008.

CURRICULUM VITAE

WILLIAM Q. ERICKSON

Place of birth: Waukesha, WI

EDUCATION

University of Wisconsin–Milwaukee
Ph.D., Mathematics, expected August 2022
Advisor: Jeb F. Willenbring
Dissertation: *The earth mover's distance through the lens of algebraic combinatorics*

University of Wisconsin–Milwaukee
M.S., Mathematics, May 2018

University of Notre Dame
B.A., Mathematics and Philosophy, May 2010
Advisor: Kenneth M. Sayre
Honors Thesis: *The nature of Gödel's mathematical Platonism*

RESEARCH INTERESTS

Representation theory of Lie groups, algebraic combinatorics, algebraic statistics, and classical invariant theory.

PUBLICATIONS

M. Colarusso, W. Erickson, and J. Willenbring. Contingency tables and the generalized Littlewood–Richardson coefficients. *Proc. Amer. Math. Soc.* 150(2022), 79–94. https://doi.org/10.1090/proc/15731.

W. Erickson. A generalization for the expected value of the earth mover's distance. *Algebr. Stat.* 12(2021), 139–166. DOI 10.2140/astat.2021.12.139.

W. Erickson. Haste makes waste: an optimization problem. *College Math. J.* 2022, 1–12. https://doi.org/10.1080/07468342.2021.2022955.

PREPRINTS

W. Erickson. The earth mover's distance as the symmetric difference of Young diagrams. arXiv:2104.07273. Under review.

W. Erickson. Enright resolutions encoded by a generating function for Blattner's formula. arXiv:2108.08469. Under review.

TEACHING EXPERIENCE

Teacher, Trinity Academy, Pewaukee, WI, 2010–2018

Math courses:

- Pre-algebra (Grade 7)
- Algebra I (Grade 8)
- Algebra II (Grade 10)
- Precalculus (Grade 11)
- AP Calculus AB/BC (Grades 11, 12)

Latin courses:

- Introduction to Latin I, II (Grades 6, 7)
- Latin I, II, III (High school)
- AP Latin

Graduate Teaching Assistant, University of Wisconsin–Milwaukee, Fall 2016–present

Instructor of record for 1 to 2 sections each semester:

- Math 98/108, Algebraic Literacy I and II, co-requisite (Fall 2018)
- Math 103, Contemporary Applications of Mathematics (Spring 2018, Spring 2021)
- Math 105, Introduction to College Algebra (Fall 2016)
- Math 108, Algebraic Literacy II (Spring 2017)
- Math 115, Precalculus (Fall 2017)
- Math 213, Calculus with Life Science Applications (Fall 2019, Spring 2020)
- Math 231, Calculus and Analytic Geometry I (Spring 2019)
- Math 233, Calculus and Analytic Geometry III (Summer 2020, Fall 2020, Fall 2021, Spring 2022)

Counselor, Ross Mathematics Program, Summer 2020

- Prepared and led a one-hour presentation for the campers in Family 20, every weekday for 6 weeks, covering material from the Ross number theory course.

- Graded the campers' problem sets every day and met with them individually to talk about their current set.

AWARDS

Ernst Schwandt Teaching Award, UW–Milwaukee, 2019
Morris and Miriam Marden Award in Mathematics, UW–Milwaukee, 2018, 2019, 2020

AFFILIATIONS

American Mathematical Society (member), 2016–present