Any-Dimensional Data Science: Learning, Optimization, and Sampling

Thesis by

Eitan Levin

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2026
Defended May 18, 2026

Eitan Levin
ORCID: 0009-0002-5789-4191

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor Venkat Chandrasekaran for his guidance and mentorship over the past six years. Throughout my time at Caltech, Venkat has always pushed me to ask big questions, to pursue correspondingly big answers to them, and to explain these answers to the broadest audience possible. I am also grateful to Venkat for creating the intellectual environment that made this work possible.

I must also thank my co-advisor Joel Tropp for his advice and support. Joel was always available when I needed his input and gave me feedback and advice at key junctures in my PhD, for which I am very grateful.

I am grateful to Babak Hassibi, Leonard Schulman, and Adam Wierman for serving on my candidacy and defense committees, for engaging with my work, and for their invaluable feedback and advice.

I would like to express my gratitude to my mentors, colleagues, and collaborators, in particular to Tamir Bendory, Nicolas Boumal, Mateo Díaz, Jordan Ellenberg, Joe Kileel, Victoria Kostina, Alex Meltzer, Amit Singer, and Soledad Villar. I have learned a lot from them and I am grateful for their advice and support over the years.

I am also thankful to my friends, with whom I have been fortunate to share my time at Caltech. I want to thank in particular Pau Batlle, Dima Burov, Nico Chistianson, Lauren Conger, Matthieu Darcy, Emily Fourney, Jorge Garza-Vargas, Sebastian Lamas, John Lathrop, Elvira Moreno, Kevin Shu, George Stepaniants, Margaret Trautner, Cameron Voloshin, and Chris Yeh. You have made these last several years fantastic.

My partner Siki Wang has been a constant source of love and support during the ups and downs of my PhD, and I have been very lucky to have met her at Caltech and share many wonderful adventures together.

I am eternally grateful to my family, and in particular my parents Gideon and Elena Levin, for their unwavering support, encouragement, and love. I would (literally) not be here without you!

ABSTRACT

An exciting direction in computational math driven by recent trends in AI involves automating the solution process of entire problem classes. For example, algorithms for sophisticated data analysis tasks are increasingly being learned from training data instead of being designed by hand. There is also increasing interest in computationally producing proofs of mathematical theorems. However, computationally-derived algorithms and proofs, in sharp contrast to manually-designed ones, rarely generalize to inputs of size not seen during training. How can we train a model on inputs of a few small sizes, and generalize it to inputs of any other size? How can we computationally search over proofs of theorems that hold for objects of all sizes? This thesis develops foundations for tackling such **any-dimensional problems** by using random sampling maps to compare and summarize objects of different sizes. Our framework allows us to systematically derive finitely-parametrized families of functions that take inputs of any size, to prove inequalities between such functions, and to learn them from data. Our methodology leverages new de Finetti-type theorems and the recently-identified phenomenon of representation stability. We illustrate the resulting framework for any-dimensional problems in several applications.

PUBLISHED CONTENT AND CONTRIBUTIONS

[1] Eitan Levin and Venkat Chandrasekaran. "Any Dimensional Learning by Sampling". In: *In Preparation* (2026).
E.L participated in the conception of the project, performed the analysis, and participated in the writing of the manuscript.

[2] Eitan Levin and Venkat Chandrasekaran. "Any-Dimensional Polynomial Optimization via de Finetti Theorems". In: *arXiv preprint arXiv:2507.15632* (2025).
E.L participated in the conception of the project, performed the analysis, undertook the numerical experiments, and participated in the writing of the manuscript.
DOI: 10.48550/arXiv.2507.15632.

[3] Eitan Levin and Venkat Chandrasekaran. "Limits of Weighted Graphs via Random Quotients". In: *arXiv preprint arXiv:2512.23149* (2025).
E.L participated in the conception of the project, performed the analysis, and participated in the writing of the manuscript.
DOI: 10.48550/arXiv.2512.23149.

[4] Eitan Levin et al. "On Transferring Transferability: Towards a Theory for Size Generalization". In: *Advances in Neural Information Processing Systems*. Ed. by D. Belgrave et al. Vol. 38.
E.L participated in the conception of the project, performed some of the analysis, and participated in the writing of the manuscript.
Curran Associates, Inc., 2025, pp. 67015–67083
URL: https://proceedings.neurips.cc/paper_files/paper/2025/file/60b393761eea35bf491687c46e4f1a5a-Paper-Conference.pdf.

[5] Eitan Levin and Mateo Díaz. "Any-dimensional equivariant neural networks". In: *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*. Ed. by Sanjoy Dasgupta, Stephan Mandt, and Yingzhen Li. Vol. 238. Proceedings of Machine Learning Research.
E.L participated in the conception of the project, performed the analysis, and participated in the writing of the manuscript.
PMLR, 2024, pp. 2773–2781
URL: https://proceedings.mlr.press/v238/levin24a.html.

[6] Eitan Levin and Venkat Chandrasekaran. "Dimension-free descriptions of convex sets". In: *arXiv preprint arXiv:2307.04230* (2023).
E.L participated in the conception of the project, performed the analysis, undertook the numerical experiments, and participated in the writing of the manuscript.
DOI: 10.48550/arXiv.2307.04230.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

INTRODUCTION

This thesis concerns data that comes in different sizes. We model such data as elements of sequences of vector spaces $(\mathbb{V}_n)$ of growing dimensionality, and ask several basic questions pertaining to such data.

(Description) How can we describe a function $f\colon \bigsqcup_n \mathbb{V}_n \to \mathbb{R}$ accepting inputs of arbitrarily-large size using finitely-many parameters?

(Learning) How do we fit an any-dimensional function to data? How do we ensure that the fitted function performs well on inputs of any possible size?

(Sketching and Property Testing) Suppose we wish to evaluate an any-dimensional function on a high-dimensional input, but doing so is expensive. Can we sketch the input to obtain a low-dimensional one with a similar function value? Doing so amounts to testing properties of the original large input quantified by the function of interest.

(Computational Inequality Proving) Mathematical theorems are often formulated as inequalities that hold for objects of all sizes. Examples include Cauchy–Schwarz, AM-GM, Jensen's inequality, or Mantel's theorem. How can we computationally prove such any-dimensional inequalities?

We tackle these questions by imposing relations between the different vector spaces $(\mathbb{V}_n)$ that model how problem instances or inputs of different sizes can be viewed as equivalent. We begin by studying the systematic description of any-dimensional functions by finitely-many parameters. In Chapter 2 we consider conic descriptions of convex sets that can be instantiated in any dimension, and construct finitely-parametrized families of convex sets and functions. Our approach proceeds by imposing symmetries and relations between dimensions, allowing us to exploit the recently-identified phenomenon of representation stability. In Chapter 3, we explain how the same ideas yield finitely-parametrized any-dimensional neural networks. Next, in Chapter 4 we turn to the problem of computationally proving any-dimensional inequalities. We focus on polynomials and prove any-dimensional inequalities between them by solving finite-dimensional optimization problems.

The construction of these problems uses a new representation for any-dimensional symmetric polynomials as moments of fixed-dimensional random samples of their inputs. The proof of this representation, and the analysis of the resulting inequalities, are based on new generalizations of de Finetti's theorem from probability. Turning to the question of sketching and property testing, in Chapter 5 we focus on testing properties pertaining to hubs in large graphs. We develop a new notion of limits of graphs of growing size based on our new de Finetti theorems, and use it obtain non-asymptotic rates for approximation of various graph parameters quantifying hub structure by sampling a constant number of edges from the graph. Finally, in Chapter 6 we investigate the ability of any-dimensional functions fitted to limited-size data to generalize to unseen input sizes. We show that the key to analyzing this generalization is an appropriate metric between inputs of different sizes. We obtain such a metric systematically by comparing distributions of samples of these inputs using an application-adapted sampling map, and use this perspective to obtain general non-asymptotic sketching and generalization rates using a unified proof technique.

We now describe the questions we tackle in each chapter in more detail.

1.1 Any-Dimensional Convex Sets (Chapter 2).

Convex sets play a central role in numerous areas of the mathematical sciences such as optimization, statistical inference, control, inverse problems, and information theory. The convex sets arising in all these domains are often well-defined in every relevant dimension. Indeed, unit balls of standard regularizers used in inverse problems (e.g., ℓ_p or Schatten norms) are defined for vectors and matrices of any size. Polytopes associated to graph problems and their relaxations (e.g., the cut polytope and the elliptope approximating it) are defined for graphs of any size. Information-theoretic quantities (e.g., relative entropy) as well as their quantum analogues, are defined for distributions and states on any number of (qu)bits. Consequently, all these convex sets and many others should be viewed not as single sets but as sequences indexed by dimension. Furthermore, there is often a single "dimension-free" description of all the sets in such a sequence, as the following example illustrates.

Example 1.1.1. *Here are a few examples of sequences of convex sets and their descriptions.*

(a) *The simplex in n dimensions is* $\Delta^{n-1} = \{x \in \mathbb{R}^n : x \geq 0,\ \mathbb{1}_n^\top x = 1\}$*, which* *is the set of probability distributions over n items. Here $x \geq 0$ denotes an entrywise nonnegative vector x, and $\mathbb{1}_n \in \mathbb{R}^n$ is the vector of all-1's.*

(b) *The spectraplex, or the set of density matrices, of size n is $\mathcal{D}^{n-1} = \{X \in \mathbb{S}^n : X \succeq 0,\ \mathrm{Tr}(X) = 1\}$. It is the set of density matrices describing mixed states in quantum mechanics [165, §2.4]. Here $X \succeq 0$ denotes a symmetric positive-semidefinite (PSD) matrix X.*

(c) *The ℓ_2 ball in $\mathbb{R}^n$ is given by $\mathcal{B}^n_{\ell_2} = \left\{x \in \mathbb{R}^n : \begin{bmatrix} 1 & x^\top \\ x & I_n \end{bmatrix} \succeq 0\right\}$ where I_n denotes the $n \times n$ identity.*

(d) *The elliptope of size n is $\{X \in \mathbb{S}^n : X \succeq 0, \mathrm{diag}(X) = \mathbb{1}_n\}$. It arises in a standard relaxation of the max-cut problem [89].*

Note that all of the above descriptions are clearly defined for any relevant dimension. Indeed, these descriptions are composed of elements such as $\mathbb{1}_n$ and I_n, linear maps such as $\mathrm{diag}(\cdot)$ and $\mathrm{Tr}(\cdot)$, and inequalities such as $\geq$ and $\succeq$, all of which are *"dimension-free"* in the sense that they are well-defined in every relevant dimension. In Chapter 2, we develop a systematic framework to study such *freely-described convex sets*.

One motivation for this effort is to facilitate structural understanding of convex sets that can be instantiated in any dimension and of sequences of optimization problems over such sets. Such sequences arise in several applications. For example, in extremal combinatorics [184, 185] it is of interest to certify inequalities between graph homomorphism densities involving graphs of every size. In quantum information and control theory [105, 125], many problems reduce to optimizing (traces of) polynomials in matrix variables of every size. Despite the ubiquity of sequences of convex sets in these problem domains, the existence and interplay between the sets in different dimensions is typically not explicitly discussed or exploited in the literature. In Chapter 2, we explicitly consider sequences of convex sets that are related in a concrete fashion across dimensions, and we present a framework to derive several structural results about such sets as well as optimization programs over them; as one notable illustration, we show that sequences of invariant conic programs (including some arising in the above applications) can be solved in time independent of dimension.

A second motivation for our effort stems from the growing interest in learning solution methods for various problem families from data. Here one is typically given input-output data, and the goal is to fit a mapping that approximately fits the data. This framework has been fruitfully applied to domains including integer programming, inverse problems, and numerical solvers for PDEs [11, 43, 131, 158, 168, 231]. However, a fundamental limitation in much of this literature is that the mappings learned from data are typically only defined for inputs of the same dimension as the ones in the training data, and extension to inputs of different dimensions is handled on a case-by-case basis. In contrast, we wish to learn *algorithms* that should be defined for inputs of any relevant size, i.e., we aim to identify a sequence of solution maps, one for each input size. Identifying a freely-described convex set offers a convenient approach to learning an algorithm specified as linear optimization over convex sets (convex programs). Further, to facilitate numerical search over a family of freely-described convex sets—for example, to fit an element of this family to training data—it is natural to seek *finitely-parametrized* families of such sets, such as the following examples.

Example 1.1.2. *The following are examples of finitely-parametrized families of freely-described convex sets.*

(a) The sequence of subsets

$$\{\alpha_1 \mathrm{diag}(X) + \alpha_2 X \mathbb{1}_n : X \succeq 0,\ \alpha_3 \mathrm{Tr}(X)\mathbb{1}_n + \alpha_4 \mathrm{diag}(X) + \alpha_5 X\mathbb{1}_n = \alpha_6 \mathbb{1}_n\},$$

of $\mathbb{R}^n$ is parametrized by $\alpha \in \mathbb{R}^6$.

(b) Free spectrahedra are sequences of the form

$$\left\{(X_1, \dots, X_d) \in (\mathbb{S}^n)^d : L_0 \otimes I_n + \sum_{i=1}^d L_i \otimes X_i \succeq 0\right\},$$

that are parametrized by $L_0, \dots, L_d \in \mathbb{S}^k$ for $k \in \mathbb{N}$. They arise in the theory of matrix convexity and free convex algebraic geometry [101, 130].

(c) The sequence

$$\begin{aligned} C_n = \Big\{ X \in \mathbb{S}^n : & \frac{\mathbb{1}_n^\top X \mathbb{1}_n}{n^2} L_1 \otimes \mathbb{1}_n\mathbb{1}_n^\top + \frac{\mathrm{Tr}(X)}{n} L_2 \otimes \mathbb{1}_n\mathbb{1}_n^\top \\ & + L_3 \otimes \frac{1}{n}\left(X\mathbb{1}_n\mathbb{1}_n^\top + \mathbb{1}_n\mathbb{1}_n^\top X\right) + L_4 \otimes \left(\mathrm{diag}(X)\mathbb{1}_n^\top + \mathbb{1}_n \mathrm{diag}(X)^\top\right) \\ & + L_5 \otimes X + L_6 \otimes \mathbb{1}_n\mathbb{1}_n^\top + L_7 \otimes (nI_n) \succeq 0 \Big\}, \end{aligned}$$

> *is parametrized by $L_1, \ldots, L_7 \in \mathbb{S}^k$. We obtain this sequence from our framework in Section 2.4 by viewing a matrix $X \in \mathbb{S}^n$ representing a weighted graph as a step* graphon *[146], and deriving a parametric family of dimension-free descriptions extending continuously to more general spaces of graphons.*

Each example here is a parametric family, and constitutes a freely-described sequence of convex sets for each value of its associated parameters.

Note that the parametric dimension-free descriptions in Example 1.1.2 are composed of linear combinations of the dimension-free elements and linear maps that appear in the descriptions of Example 1.1.1. In fact, the parameters above are simply the coefficients in these linear combinations. Thus, we can obtain finitely-parametrized families of freely-described convex sets by deriving *finite-dimensional spaces* of dimension-free elements.

The central thesis of Chapter 2 is that the dimension-free elements which constitute dimension-free descriptions arise from a recently-identified phenomenon in algebraic topology called *representation stability*, which relates invariants across dimensions in a sequence of group representations [47]. We use these relations to formally define the dimension-free elements that constitute our dimension-free descriptions, and investigate various structural properties of such descriptions as well as the associated sequences of optimization problems. Our definitions together with results from representation stability also yield finite-dimensional spaces of dimension-free elements, which we use to derive parametric families of freely-described convex sets adapted to specific applications and to fit elements of these families to data.

1.2 Any-Dimensional Neural Networks (Chapter 3).

Researchers are often interested in learning mappings that are defined for inputs of different sizes. We call such objects "*free*" as they are dimension-agnostic. Free objects are pervasive in a range of scientific and engineering fields. For example, in physics, quantities like electromagnetic fields are defined for any number of particles. In signal processing, problems such as deconvolution are well-defined for signals of any length, while image and text classification tasks are meaningful regardless of the input size. In mathematics, norms of vectors and matrices are defined for any dimension; sorting algorithms handle arrays of any length, and graph parameters, such as the max-cut value, are defined for graphs of any size.

In contrast, most existing supervised learning methods aim to learn mappings by fitting a parametrized function $\widehat{f}\colon \mathbb{R}^d \to \mathbb{R}^k$ to a set of input-output pairs $\mathcal{S} = \{(X_1, y_1), \ldots, (X_n, y_n)\} \subseteq \mathbb{R}^d \times \mathbb{R}^k$ with fixed dimensions d and k. Naturally, the function $\widehat{f}$ is only defined on inputs of dimension d, and *a priori* cannot be applied to inputs in other dimensions. Practitioners use ad-hoc heuristics, such as downsampling, to resize the dimension of new inputs and match that of the learned map [96]. The lack of systematic methodology for applying learned functions to inputs of different sizes motivates the main question of this chapter: How can we learn mappings that accept inputs in any dimension? We answer this question for mappings that are invariant or equivariant under the action of groups. Such mappings are ubiquitous since many application domains exhibit symmetry, e.g., physical quantities are equivariant under translations and rotations because they only depend on the relative position of particles [222], while graph parameters are equivariant under permutations because they only depend on the underlying topology rather than on its labelling [41]. We consider learning such mappings using free equivariant neural networks. Formally, we consider a sequence of neural networks and a sequence of groups, such that a network at any level in the sequence is equivariant with respect to the group at the same level; see Figure 3.1. We train an equivariant network with data $\mathcal{S}$ in a fixed dimension and leverage symmetry to extend it to other dimensions.

1.3 Any-Dimensional Polynomial Optimization (Chapter 4).

Optimization problems specified by multivariate polynomials arise commonly in control, data science, dynamical systems, and theoretical computer science. Such polynomial optimization problems (POPs) are NP-hard to solve in general, and much research effort has focused on producing tractable bounds on their optimal values using sums of squares certificates of polynomial nonnegativity [16, 133, 172]. In many application domains such as game theory and extremal combinatorics, POPs arise naturally as sequences indexed by dimension and it is of interest to simultaneously bound all of the optimal values in the sequence. The next few examples illustrate the ubiquity of these problems.

Example 1.3.1 (Inequalities in moments)**.** *Fix a compact set $\Theta \subseteq \mathbb{R}^d$ and consider the sequence of POPs:*

$$u_n = \inf_{(x_1,\ldots,x_n)\in\Theta^n} f\left(\frac{1}{n}\sum_{i=1}^n x_i^{\alpha_1}, \ldots, \frac{1}{n}\sum_{i=1}^n x_i^{\alpha_m}\right), \tag{1.1}$$

where f is an m-variate real polynomial and $\alpha_1, \ldots, \alpha_m \in \mathbb{N}^d$ are multi-indices. Here $x^\alpha = \prod_{i=1}^d x_i^{\alpha_i}$ if $x \in \mathbb{R}^d$ and $\alpha \in \mathbb{N}^d$. Observing that $\frac{1}{n}\sum_{i=1}^n x_i^\alpha = \mathbb{E}_{X\sim\mu(x)} X^\alpha$ where $\mu(x) = \frac{1}{n}\delta_{x_i}$ is a discrete measure, computing $u_\infty = \inf_n u_n$ amounts to proving polynomial inequalities satisfied by moments of any measure supported on Θ. Such problems also arise in the context of potential mean-field games [36] as well as in operator theory and quantum information [124].

Example 1.3.2 (Inequalities in symmetric functions)**.** *Symmetric functions are certain symmetric polynomials defined for any number of variables, such as power sums $s_k(x) = \sum_i x_i^k$ and elementary symmetric functions $\sum_{i_1<\ldots<i_k} x_{i_1}\cdots x_{i_k}$. These are fundamental objects of study in enumerative combinatorics, representation theory, and analysis [151, 209]. In particular, inequalities between symmetric functions, which hold regardless of the number of variables, are of interest in all of these areas [42, 106, 120, 181]. Since any symmetric function can be written as a polynomial expression $f(s_1, \ldots, s_k)$ in power sums, checking whether a given symmetric function is nonnegative in all dimensions amounts to considering the following sequence of POPs:*

$$u_n = \inf_{x\in\mathbb{R}^n} f\left(\sum_{i=1}^n x_i, \ldots, \sum_{i=1}^n x_i^m\right),$$

and checking whether $u_\infty = \inf_n u_n \geq 0$. The problem of checking $u_\infty \geq 0$ given a polynomial f is in general undecidable [4], and cannot always be done by expressing f as a sum of squares [3].

Example 1.3.3 (Inequalities in graph densities)**.** *For graphs H, G, the* homomorphism density *$t(H; G)$ of H in G is the fraction of maps between their vertex sets that are graph homomorphisms, and this fraction can be written as a polynomial in the entries of the adjacency matrix of G [185]. Many problems in extremal combinatorics can be formulated as proving inequalities between homomorphism densities that hold for all graph sizes [28, 146]. Formally, given $c_1, \ldots, c_m \in \mathbb{R}$ and graphs $H_1, \ldots, H_m$, we consider the sequence of POPs:*

$$u_n = \inf_{\substack{\text{graphs } G,\\ |V(G)|=n}} \sum_{i=1}^m c_i t(H_i; G), \tag{1.2}$$

over graphs on n vertices, and ask whether $u_\infty = \inf_n u_n \geq 0$. This is equivalent to asking whether the inequality $\sum_i c_i t(H_i; G) \geq 0$ holds for graphs G of all sizes. Once again, checking this given c_i, H_i is undecidable in general, and there are nonnegative such expressions which are not sums of squares [97].

We refer to such sequences of polynomial minimization problems indexed by dimension as *any-dimensional* POPs, and in each of the preceding examples, the central question is to compute $u_\infty = \inf_n u_n$ or to prove that $u_\infty \geq 0$. As Examples 1.3.2-1.3.3 illustrate, this is in general undecidable. To address this difficulty, we adopt an optimization perspective and aim to compute a lower bound $\gamma \in \mathbb{R}$ on u_∞, or equivalently, on all the optimal values u_n in the sequence, using a finite-time procedure. Although there is a large literature on computing such bounds for fixed-dimensional POPs [16], producing bounds on one of the optimal values u_n in the sequence does not, in general, yield bounds on the limiting value u_∞. Also, while undecidability of general any-dimensional POPs implies that there are no finite-time procedures to decide, for example, whether u_∞ is nonnegative, this does not preclude the existence of such procedures for producing some finite lower bounds on u_∞.

In Chapter 4, we develop a systematic framework for producing lower bounds on the limiting optimal value u_∞ of any-dimensional POPs $(u_n = \inf_{x\in\Omega_n} p_n(x))_{n\in\mathbb{N}}$ in terms of *finite-dimensional POPs* of the form $\ell_n = \inf_{x\in\Omega_n} q_n(x)$, so that

$$\ell_n \leq u_\infty \leq u_n, \quad \text{for all } n \in \mathbb{N}. \tag{1.3}$$

These lower-bounding problems have the same constraint sets as in the original sequence of problems, but their cost functions q_n are distinct from p_n. Moreover, these lower bounds ℓ_n converge at explicit rates to u_∞. Either solving the finite-dimensional problems defining ℓ_n or applying standard convex relaxations to them yields lower bounds for u_∞. Our lower bounds can be viewed as searching over tractable families of *any-dimensional certificates of nonnegativity*, and hence providing a computational procedure for proving inequalities valid in all dimensions. The following are a few examples of the bounds we obtain using our framework.

Example 1.3.4 (Mean-field games)**.** *Consider the sequence of problems*

$$u_n = \inf_{x\in[-1,1]^n} 5\left(\frac{1}{n}\sum_i x_i\right)^2 - 4\left(\frac{1}{n}\sum_i x_i^2\right)\left(\frac{1}{n}\sum_i x_i\right) - \frac{1}{n}\sum_i x_i,$$

which computes the value of an n-player extension of the non-convex-concave two-player game studied in [174, Ex. 3.2]. The limiting optimal value u_∞ computes the value of a mean-field limit of these games, and we seek lower bounds on this value. Our framework yields the lower bounds for $n \geq 3$

$$u_\infty \geq \ell_n = \inf_{x\in[-1,1]^n} 5\frac{1}{\binom{n}{2}}\sum_{i<j} x_i x_j - 4\frac{1}{n(n-1)}\sum_{i\neq j} x_i^2 x_j - \frac{1}{n}\sum_i x_i.$$

Moreover, we get a nonasymptotic bound for the convergence of our bounds, which reads

$$u_n - \ell_n \leq \frac{60}{n}, \tag{1.4}$$

that is, ℓ_n converges to u_∞ at a rate of $O(1/n)$ by (1.3). We numerically illustrate the convergence of ℓ_n to this asymptotic optimal value, and validate our theoretical rate of convergence, see Figure 1.1 and Example 4.6.1 for more details.

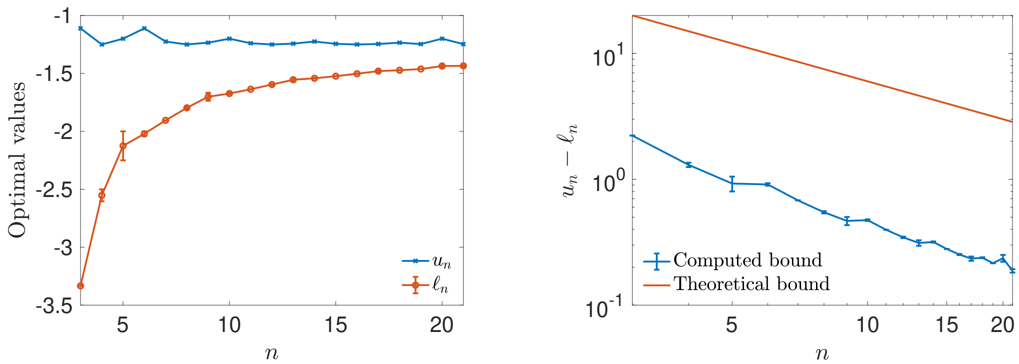


Figure 1.1: Numerical results for mean-field game in Example 1.3.4: (left) Upper bounds on u_n and intervals containing ℓ_n; (right) Comparing $u_n - \ell_n$ and bound from (1.4).

Example 1.3.5 (Symmetric functions)**.** *Our framework allows us to prove new inequalities between symmetric functions that are inaccessible to previous, sums-of-squares-based approaches. Recall from Example 1.3.2 that any symmetric function can be written as a polynomial in power sums $s_k(x) = \sum_i x_i^k$. Consider the symmetric function*

$$\begin{aligned} f(s_1, s_2, s_4, s_6) =& \tfrac{1}{2}s_6 - \tfrac{15}{16}s_4 s_2 + \tfrac{1}{16}s_4 s_1^2 + \tfrac{15}{32}s_2^3 - \tfrac{3}{32}s_2^2 s_1^2 - \tfrac{1}{32}s_2 s_1^4 + \tfrac{1}{32}s_1^6 \\ &+ \tfrac{3}{8}s_4 - \tfrac{9}{16}s_2^2 + \tfrac{3}{8}s_2 s_1^2 - \tfrac{3}{16}s_1^4 + 1. \end{aligned} \tag{1.5}$$

Checking whether or not this function is nonnegative in all dimensions reduces to setting $u_n = \inf_{x\in\mathbb{R}^n} f(s_1(x), s_2(x), s_4(x), s_6(x))$ and checking whether $u_\infty = \inf_n u_n \geq 0$. Notably, the polynomial (1.5) is not a sum of squares in any dimension $n \geq 2$, so its nonnegativity cannot be proved using the relaxations of [3], see Example 4.6.5 for details. We prove its nonnegativity in all dimensions by using our framework to produce the lower bound

$$\ell_2 = \inf_{x\in\mathbb{R}^2} x_1^4 x_2^2 + x_1^2 x_2^4 - 3x_1^2 x_2^2 + 1 \tag{1.6}$$

consisting of minimizing the well-known Motzkin polynomial. The polynomial in (1.6) *is nonnegative by the AM-GM inequality, proving that* $\ell_2 \geq 0$ *and hence* $u_\infty \geq 0$*.*

Example 1.3.6 (Graph densities)**.** *To certify inequalities in graph densities, we apply our framework to the sequence of problems* (1.2) *to produce the lower bounds*

$$\ell_n = \inf_{\substack{\text{graphs } G,\\ |V(G)|=n}} \sum_{i=1}^{m} c_i t_{\text{inj}}(H_i; G), \tag{1.7}$$

which are almost identical to (1.2) *except that the homomorphism densities* $t(H_i; G)$ *are replaced with* injective *homomorphism densities* $t_{\text{inj}}(H_i; G)$*. The latter gives the fraction of injective maps between the vertex sets of* H_i *and of* G *that are graph homomorphisms. Moreover, we get the nonasymptotic convergence bound*

$$u_n - \ell_n \leq \frac{V(V-1)\sum_{i=1}^m |c_i|}{n}, \quad \text{where } V = \max_{i=1,\ldots,m} |V(H_i)|.$$

As an application, we prove that the Ramsey multiplicity of the graph consisting of a triangle with a pendant edge is lower-bounded by 0.0476, see Example 4.6.8.

Example 1.3.7 (Graph numbers)**.** *Unlike inequalities in graph densities, inequalities in their unnormalized counterparts have been less extensively studied. For example, let* $\text{inj}(H; G)$ *be the number of injective graph homomorphisms from* H *to* G*. This extends to a polynomial defined on all weighted graphs as explained in Section 4.6. Consider*

$$u_n = \inf_{X \in \mathbb{S}^n} \text{inj}(P_3; X) - \text{inj}(K_2 \sqcup K_2; X), \quad \text{s.t. } X \geq 0,\ \mathbb{1}^\top X \mathbb{1} = 1. \tag{1.8}$$

The objective is a homogeneous quadratic, so certifying $u_\infty \geq \gamma$ *implies* $\text{inj}(P_3; X) - \text{inj}(K_2 \sqcup K_2; X) \geq \gamma \text{hom}(K_2; X)^2$ *for all* $X \geq 0$ *of all sizes. Our lower bounds take the form*

$$\ell_n = \inf_{X \in \mathbb{S}^n} \frac{n^3}{n(n-1)(n-2)} \left[\text{inj}(P_3; X) - \frac{n+1}{n-3}\text{inj}(K_2 \sqcup K_2; X)\right], \quad \text{s.t. } X \geq 0,\ \mathbb{1}^\top X \mathbb{1} = 1. \tag{1.9}$$

We numerically compute upper bounds on u_n *and intervals containing* ℓ_n *for* $n \in [4, 9]$*, with the results shown in Table 1.1, see Example 4.6.11 for more details.*

We construct and analyze these bounds using new generalizations of de Finetti's theorem from probability [10, 56, 57].

n	4	5	6	7	8	9
u_n	-0.50	-0.50	-0.67	-0.67	-0.75	-0.75
ℓ_n	-6.67	[-3.30, -3.12]	-2.8	[-2.33, -2.18]	-2.06	[-1.88, -1.81]

Table 1.1: Bounds on u_n and ℓ_n for the injective homomorphism numbers in Example 4.6.11.

1.4 Graph Limits via Random Quotients (Chapter 5).

Analyzing the structure of large graphs is a challenge that arises in numerous problem domains such as bioinformatics, network tomography, operations research, and the social sciences [13, 38, 107, 164, 210]. The graphs arising in these domains are, in general, weighted and directed. For example, in traffic networks vertices correspond to locations, and edge weights quantify the amount of traffic between two locations in each direction, see Figure 1.2. In functional brain connectome graphs, vertices correspond to regions of the brain, and edge weights quantify the connectivity between these regions in each direction, see Figure 1.3. A particularly important global structural aspect of such graphs is the presence of hub vertices and connections between them [29, 99, 167, 180, 214]. For graphs with nonnegative edge weights and possible self-loops, a *hub* is a vertex with a significant fraction of the total edge weight incident to it, such as a large city in a traffic network, a well-connected region of the brain, or a particularly interactive protein. As a consequence of their high connectivity, hubs play a central role in the structure and function of large networks. The structure of such large graphs has been studied using suitable notions of graph limits, most prominently graphons [146]; see Section 5.1 for related work. However, these previous approaches are either limited to simple graphs, or summarize large graphs by their subgraph densities. Subgraph densities fail to properly account for global structures such as hubs. Indeed, a network can only contain a relatively small number of hubs, and therefore most subgraphs of a network do not contain any hubs.

Our central thesis is that hub structure in large graphs can be studied using *quotients*. Formally, a quotient of size k of a graph on n vertices is defined by an ordered partition $f_{k,n} : [n] \to [k]$, which partitions the vertex set $[n]$ into k parts that are given by the fibers $f_{k,n}^{-1}(j)$ for $j \in [k]$. In particular, for a graph on n vertices with adjacency matrix $G \in \mathbb{R}_+^{n\times n}$, its quotient by the ordered partition $f_{k,n}$ is given by the graph $\rho(f_{k,n})G \in \mathbb{R}_+^{k\times k}$ on k vertices with edge weights:

$$[\rho(f_{k,n})G]_{i,j} = \sum_{\substack{v\in f_{k,n}^{-1}(i)\\ u\in f_{k,n}^{-1}(j)}} G_{v,u}. \tag{1.10}$$

Here we view $\rho(f_{k,n}) : \mathbb{R}^{n\times n} \to \mathbb{R}^{k\times k}$ as a 'linear representation' of $f_{k,n}$. In words, the vertices of the quotient $\rho(f_{k,n})G$ correspond to the parts in the partition defined by $f_{k,n}$, and the edge weight between parts $i, j \in [k]$ is given by the sum of all edge weights in G between vertices in $f_{k,n}^{-1}(i)$ and $f_{k,n}^{-1}(j)$. Thus, a quotient of G represents the edge weights between each pair of parts in a vertex partition.

As an illustration, we give in Figures 1.2-1.3 examples of small quotients of two networks. The first, depicted in Figure 1.2, is a traffic network representing the fraction of trips starting and ending at each zone of the Gold Coast, Australia [207]. The second, depicted in Figure 1.3, is a brain connectome, the network of connections between neurons of an adult male *C. elegans* roundworm [50, 163]. The structure of each quotient indicates some of the features of the underlying large network. As we will see in Section 5.4, the presence of hubs in the original network leads to nonuniform distribution of edge weight in their random quotients.

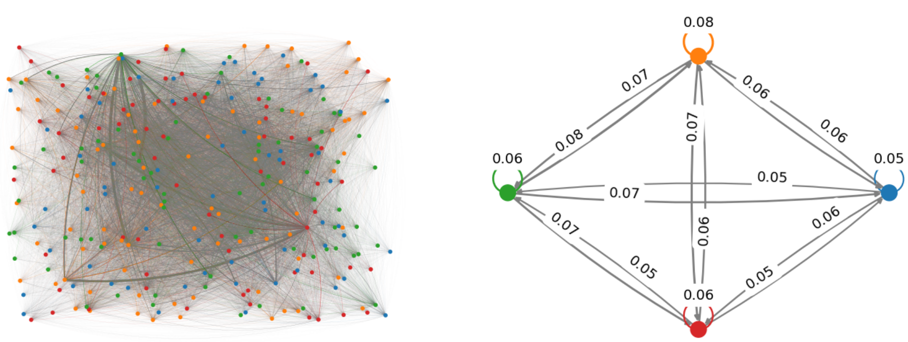


Figure 1.2: Traffic network for the Gold Coast, Australia, and a realization of a random quotient. The city is divided into 1068 zones, and the weight of an edge is the fraction of trips from a dataset with given start and end zones [207]. Here we plot the first 300 of these zones and their interconnections, displaying the width of an edge proportionally to its weight. We color the vertices according to their image in the quotient, and color an edge connecting two vertices with the same image by the color of that image. The nonuniformity in the edge weights of the quotient indicates the presence of hubs that we can see in the plot of the entire network.

In Chapter 5, we define a new notion of convergence for sequences of growing-sized graphs based on convergence of their random quotients. In analogy with

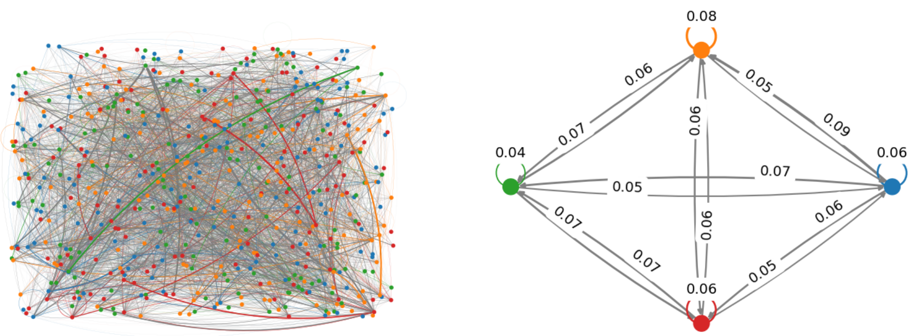


Figure 1.3: Network of chemical connections among the 575 neurons of adult male C. elegans (a species of roundworm), and a realization of a random quotient. Edges are directed from pre-synaptic to post-synaptic cells, and their weight is the total number of EM serial sections of connectivity [50, 163]. The nonuniform distribution of self-loops and the unequal in-degrees and out-degrees among the four vertices in the quotient again indicate the presence of hubs.

graphons [146, §1], we define suitable limit objects in terms of random measures for any quotient-convergent sequence of graphs. Finally, we show that finite (random) graphs sampled from our limit objects are quotient-convergent to the limit almost surely. The key concepts in Chapter 5 are dual to those in the literature on graphons, and we emphasize this duality throughout our development. In particular, we shall see that random quotients and random subgraphs, in terms of which graphon convergence is defined, are random linear maps which are adjoints of each other. Further, a dual analog of the cut metric plays a prominent role in our development, and we derive an edge-based analog of the Szemerédi regularity lemma as a consequence of our results.

1.5 Any-Dimensional Learning (Chapter 6).

Out-of-sample generalization is a core challenge in machine learning, particularly when the test data differs substantially from from the training data. Importantly, such difference arises when the dimensionality of the data can vary from training to test time, as commonly happens in applications involving point clouds containing different numbers of points, graphs containing different numbers of vertices, and sequences of tokens of different lengths. In these and other domains, we aim to learn a function from a data set containing inputs of limited sizes that can be applied at test time to inputs of any size. Such a learned function must therefore exhibit *any-dimensional generalization* in order to perform well on inputs of unseen sizes.

A related problem is one in which we are given a function that is well-defined on inputs of any size, and we wish to evaluate the function on a high-dimensional input. When such a function is expensive to evaluate on high-dimensional inputs, we seek to suitably subsample, or sketch, the input to obtain a low-dimensional object on which the function can be evaluated efficiently, and whose value is close to that of the original input. We call the problem of producing such low-dimensional sketches *any-dimensional sketching*, as we seek to produce small approximations of arbitrarily-sized inputs.

Our objective in Chapter 6 is to show that the above two problems of any-dimensional generalization and sketching are closely related to each other, and to develop a systematic approach to tackle them. The fundamental challenge in addressing these questions is that of comparing inputs of different sizes. In what sense can a small object be an approximation to a large one? In assessing generalization performance across dimension, what is an appropriate notion of a data distribution? As described in Chapter 6, we develop a sampling-based framework to answer these questions. Building on this approach, we then identify classes of any-dimensional functions that can be learned from or approximated on low-dimensional inputs, along with explicit rates. We also discuss applications to a number of domains of contemporary interest.

Chapter 2

DIMENSION-FREE DESCRIPTIONS OF CONVEX SETS

This chapter is based on [139].

2.1 Introduction

This chapter consists of four main contributions. First, we formally define dimension-free descriptions of convex sets by generalizing the insights derived from Examples 1.1.1 and 1.1.2 using representation stability. Second, we prove structural results pertaining to freely-described sequences of convex sets, the relationship between sets in such a sequence in different dimensions, and limits of such sequences. Our proofs combine concepts from convex analysis and representation theory. Third, we derive consequences of our framework for invariant conic programs, showing that sequences of such programs of growing dimension can often be solved in constant time. Our results unify and generalize existing work in the literature. Finally, we apply our framework and its structural results to derive an algorithm for fitting a freely-described convex function to given data in different dimensions, a problem we call dimension-free convex regression. The latter two contributions are aimed at addressing our mathematical and algorithmic motivations above.

Freely-described Convex Sets

To formalize dimension-free descriptions of convex sets, we begin by briefly reviewing *conic descriptions*, which express a convex set as an affine section of a convex cone. Conic descriptions are the most popular approach to specifying convex sets in the optimization literature, and they have played a central role in the development of modern convex optimization [14]. Indeed, we often classify convex sets based on their conic descriptions—polyhedra are affine sections of nonnegative orthants and spectrahedra are affine sections of PSD cones. Formally, if $\mathbb{V}, \mathbb{W}, \mathbb{U}$ are (finite-dimensional) vector spaces and $\mathcal{K} \subseteq \mathbb{U}$ is a convex cone, then a convex subset $C \subseteq \mathbb{V}$ can be described using linear maps $A\colon \mathbb{V} \to \mathbb{U}$, $B\colon \mathbb{W} \to \mathbb{U}$ and a vector $u \in \mathbb{U}$ as follows:

$$C = \{x \in \mathbb{V} : \exists y \in \mathbb{W} \text{ s.t. } Ax + By + u \in \mathcal{K}\}. \tag{Conic}$$

We will refer to the spaces $\mathbb{W}$ and $\mathbb{U}$ as the *description spaces* associated to the conic description. If the cone $\mathcal{K}$ is a nonnegative orthant (resp., PSD cone), then linear optimization over C is a linear (resp., semidefinite) program. The type of the cone as well as its dimension determine the computational complexity of optimization over C.

Each sequence of convex sets $C_n \subseteq \mathbb{V}_n$ in Examples 1.1.1 and 1.1.2 is given by (Conic) for suitable sequences of description spaces $\mathbb{W}_n, \mathbb{U}_n$, of vectors $u_n \in \mathbb{U}_n$, of linear maps $A_n\colon \mathbb{V}_n \to \mathbb{U}_n$, $B_n\colon \mathbb{W}_n \to \mathbb{U}_n$, and of convex cones $\mathcal{K}_n \subseteq \mathbb{U}_n$. The vector spaces and cones are typically fixed in advance and are dictated by the application domain. Having fixed the vector spaces and cones, we seek to formalize the vectors u_n and linear maps A_n, B_n that are defined for any dimension, such as those appearing in Examples 1.1.1 and 1.1.2. Furthermore, we wish to obtain finite-dimensional spaces of these dimension-free elements, which yield finitely-parametrized families of dimension-free descriptions. To that end, we observe that these dimension-free elements are sequences of invariants under sequences of groups related in a particular way across dimensions. For example, the all-1's vector $\mathbb{1}_n$ of length n is invariant under the group of permutations on n letters acting by permuting coordinates. Further, the all-1's vectors of different lengths are related to each other: extracting the first n entries of $\mathbb{1}_{n+1}$ yields $\mathbb{1}_n$. Similarly, the $n\times n$ identity matrix I_n is invariant under the orthogonal group of size n acting by conjugation, and extracting the top left $n \times n$ submatrix of I_{n+1} yields I_n. Thus, to give a formal definition for dimension-free vectors and linear maps, we consider sequences of groups acting on sequences of vector spaces, and we require the spaces in the sequence to be related to each other – specifically, we embed lower-dimensional spaces into higher-dimensional ones and project higher-dimensional spaces onto lower-dimensional ones. Such sequences of group representations are called *consistent sequences* and they were first defined in the seminal paper [47] introducing representation stability.

Definition 2.1.1 (Consistent sequences [47])**.** *Fix a family of compact*[1] *groups* $\mathscr{G} = \{\mathsf{G}_n\}_{n\in\mathbb{N}}$ *such that* $\mathsf{G}_n \subseteq \mathsf{G}_{n+1}$. *A* consistent sequence *of* $\mathscr{G}$*-representations is a sequence* $\mathscr{V} = \{(\mathbb{V}_n, \varphi_n)\}_{n\in\mathbb{N}}$ *satisfying the following properties:*

(a) $\mathbb{V}_n$ *is an orthogonal* G_n*-representation;*

(b) $\varphi_n\colon \mathbb{V}_n \hookrightarrow \mathbb{V}_{n+1}$ *is a linear* G_n*-equivariant isometry.*

[1]Our theory can be extended to reductive groups.

Unless we want to emphasize the embeddings φ_n, we shall identify $\mathbb{V}_n$ with its image inside $\mathbb{V}_{n+1}$. We then write $\mathscr{V} = \{\mathbb{V}_n\}$ and take φ_n to be inclusions $\mathbb{V}_n \subseteq \mathbb{V}_{n+1}$. [2]

As φ_n is an isometry, its adjoint φ_n^* defines the orthogonal projection $\mathcal{P}_{\mathbb{V}_n}$ of higher-dimensional spaces onto lower-dimensional ones. We now formally define dimension-free vectors or linear maps as sequences of invariants projecting onto each other.

Definition 2.1.2 (Freely-described elements). *A* freely-described element *in a consistent sequence $\{(\mathbb{V}_n, \varphi_n)\}$ of $\{\mathsf{G}_n\}$-representations is a sequence $\{v_n \in \mathbb{V}_n^{\mathsf{G}_n}\}$ of invariants satisfying $\varphi_n^*(v_{n+1}) = v_n$ for all n.*

Importantly, the set of freely-described elements in a given consistent sequence $\mathscr{V}$ is naturally a linear space,[3] because if $\{v_n\}$ and $\{v'_n\}$ are freely-described elements then so is $\{\alpha v_n + \beta v'_n\}$ for any $\alpha, \beta \in \mathbb{R}$. Furthermore, fundamental results in representation stability imply that these spaces of freely-described elements are often finite-dimensional. More precisely, when the consistent sequence $\mathscr{V}$ is *finitely-generated*, the restricted projections $\mathcal{P}_{\mathbb{V}_n}|_{\mathbb{V}_{n+1}^{\mathsf{G}_{n+1}}} : \mathbb{V}_{n+1}^{\mathsf{G}_{n+1}} \to \mathbb{V}_n^{\mathsf{G}_n}$ become isomorphisms for all n exceeding the *presentation degree* of the sequence, see Section 2.2 for precise definitions.

Definition 2.1.2 also defines freely-described linear maps. Indeed, if $\mathscr{V} = \{(\mathbb{V}_n, \varphi_n)\}$ and $\mathscr{U} = \{(\mathbb{U}_n, \psi_n)\}$ are consistent sequences of $\{\mathsf{G}_n\}$-representations, then the sequence of spaces of linear maps between them can be naturally identified with $\{(\mathbb{V}_n \otimes \mathbb{U}_n, \varphi_n \otimes \psi_n)\}$ where $\otimes$ is the tensor (or Kronecker) product, and this is also a consistent sequence. A freely-described element of $\mathscr{V} \otimes \mathscr{U}$ is a sequence of equivariant maps $\{A_n \in \mathcal{L}(\mathbb{V}_n, \mathbb{U}_n)^{\mathsf{G}_n}\}$ satisfying $\psi_n^* A_{n+1} \varphi_n = A_n$. When φ_n, ψ_n are inclusions, this simplifies to $\mathcal{P}_{\mathbb{U}_n} A_{n+1}|_{\mathbb{V}_n} = A_n$ where $\mathcal{P}_{\mathbb{U}_n}$ is orthogonal projection onto $\mathbb{U}_n$. From results in representation stability, it follows that for a large class of consistent sequences, the sequence $\mathscr{V} \otimes \mathscr{U}$ is finitely-generated when $\mathscr{V}, \mathscr{U}$ are, and the presentation degree of the former can be bounded in terms of those of the latter; see Theorem 2.2.11. Thus, the space of freely-described linear maps is often finite-dimensional as well.

Example 2.1.3 (Vectors with zero-padding). *Let $V_n = \mathbb{R}^n$ with the standard inner product, and let $\varphi_n(x) = [x^\top, 0]^\top$ correspond to padding a vector with a zero. This*

[2] Formally, we obtain such inclusions inside the direct limit of the sequence.

[3] Formally, the space of freely-described elements is the inverse limit $\varprojlim_n \mathbb{V}_n^{\mathsf{G}_n}$.

is a consistent sequence for many standard sequences $\{\mathsf{G}_n\}$ *of groups, including the groups of* $n \times n$ *orthogonal matrices, and (signed) permutation matrices. Here* G_n *is embedded in* G_{n+1} *by sending* $g \in \mathsf{G}_n$ *represented as an* $n \times n$ *matrix to* $\mathrm{blkdiag}(g, 1)$. *For each of these sequences of groups, we get a consistent sequence* $\mathscr{V} = \{(\mathbb{V}_n, \varphi_n)\}$. *When* $\mathsf{G}_n = \mathsf{S}_n$ *is the group of permutations on n letters, the space of freely-described elements in* $\mathscr{V}$ *is one-dimensional and consists of sequences* $\{\alpha \mathbb{1}_n\}$ *for* $\alpha \in \mathbb{R}$. *The space of freely-described linear maps from* $\mathbb{V}_n$ *to itself is two-dimensional, and consists of sequences* $\{\alpha_1 \mathbb{1}_n \mathbb{1}_n^\top + \alpha_2 I_n\}$ *for* $\alpha \in \mathbb{R}^2$.

Having formalized freely-described vectors and linear maps, we are ready to define dimension-free descriptions of convex sets. These are just sequences of conic descriptions specified using freely-described elements.

Definition 2.1.4 (Dimension-free conic descriptions). *Let* $\mathscr{V} = \{\mathbb{V}_n\}$, $\mathscr{W} = \{\mathbb{W}_n\}$, $\mathscr{U} = \{\mathbb{U}_n\}$ *be consistent sequences of* $\{\mathsf{G}_n\}$*-representations, and* $\{\mathcal{K}_n \subseteq \mathbb{U}_n\}$ *be a sequence of convex cones. A sequence of conic descriptions*

$$C_n = \{x \in \mathbb{V}_n : \exists y \in \mathbb{W}_n \ s.t. \ A_n x + B_n y + u_n \in \mathcal{K}_n\}, \tag{ConicSeq}$$

is called dimension-free *if* $\{A_n\}, \{B_n\}$*, and* $\{u_n\}$ *are freely-described elements of the consistent sequences* $\mathscr{V} \otimes \mathscr{U}$, $\mathscr{W} \otimes \mathscr{U}$*, and* $\mathscr{U}$*, respectively.*

All the descriptions in Examples 1.1.1 and 1.1.2 become dimension-free when the relevant sequences of vector spaces are endowed with natural consistent sequence structure. Moreover, when $\mathscr{V} \otimes \mathscr{U}, \mathscr{W} \otimes \mathscr{U}, \mathscr{U}$ are all finitely-generated, Definition 2.1.4 yields a finitely-parametrized family of dimension-free descriptions, obtained by choosing bases for the spaces of freely-described $\{A_n\}, \{B_n\}, \{u_n\}$ and viewing the coefficients in these bases as the parameters. These parametric families generalize Example 1.1.2. More broadly, they can be adapted to specific applications via the choice of embeddings and group actions involved in the consistent sequences, which formally relate instances of different sizes and their symmetries.

We can use dimension-free descriptions to describe convex *functions* as well as sets, using the many correspondences between convex sets and functions from convex analysis. For example, given a freely-described sequence of convex sets we can consider the sequence of their support and gauge functions, and given a sequence of convex functions we can consider dimension-free descriptions for the sequence of their epigraphs [189, §4]. Thus, Definition 2.1.4 yields finitely-parametrized infinite sequences of convex functions as well.

Example 2.1.5 (Parametric convex graph invariants)**.** *A convex graph invariant is a convex function over symmetric matrices (viewed as adjacency matrices of weighted graphs) that is invariant under conjugation of its argument by permutation matrices (viewed as relabelling the vertices of the graphs); for example, the max-cut value of a (weighted) graph is a convex graph invariant [41]. These examples and others are defined for graphs of any size, and therefore they correspond to a sequence of convex functions* $\{f_n : \mathbb{S}^n \to \mathbb{R}\}$ *such that* f_n *is invariant under conjugation of its argument by permutation matrices for each* n.

Our framework yields parametric families of convex graph invariants as support or gauge functions of parametric freely-described convex sets. Set $\mathbb{V}_n = \mathbb{S}^n$ *with the action of* $\mathsf{G}_n = \mathsf{S}_n$ *by conjugation, since we seek convex subsets of symmetric matrices invariant under this group action. We choose embeddings* $\varphi_n : \mathbb{S}^n \hookrightarrow \mathbb{S}^{n+1}$ *by padding with a zero row and column, which corresponds to appending isolated vertices to a graph. This yields the consistent sequence* $\mathscr{V} = \{(\mathbb{V}_n, \varphi_n)\}$.

To obtain convex sets, we also need to choose description spaces and cones. For simplicity, we choose description spaces $\mathbb{W}_n = 0$ *and* $\mathbb{U}_n = \mathbb{V}_n$ *with the same embeddings and group actions as for* $\mathscr{V}$*, and the positive semidefinite cones* $\mathcal{K}_n = \mathbb{S}^n_+$. *Once the relevant consistent sequences and cones have been chosen, the parametric family of freely-described sets, parametrized in this case by* $\alpha \in \mathbb{R}^{11}$*, are obtained transparently from our framework:*

$$\begin{aligned} C_n = \Big\{ X \in \mathbb{S}^n : &\ \alpha_1(\mathbb{1}_n^\top X \mathbb{1}_n)\mathbb{1}_n\mathbb{1}_n^\top + \alpha_2(\mathbb{1}_n^\top X \mathbb{1}_n) I_n + \alpha_3 \mathrm{Tr}(X)\mathbb{1}_n\mathbb{1}_n^\top + \alpha_4 \mathrm{Tr}(X) I_n \\ &+ \alpha_5\Big(X\mathbb{1}_n\mathbb{1}_n^\top + \mathbb{1}_n\mathbb{1}_n^\top X\Big) + \alpha_6\Big(\mathrm{diag}(X)\mathbb{1}_n^\top + \mathbb{1}_n \mathrm{diag}(X)^\top\Big) + \alpha_7 X + \alpha_8 \mathrm{diag}(X\mathbb{1}_n) \\ &+ \alpha_9 I \odot X + \alpha_{10}\mathbb{1}_n\mathbb{1}_n^\top + \alpha_{11} I_n \succeq 0 \Big\}. \end{aligned} \tag{2.1}$$

We can obtain a larger parametric family by enlarging the description space. For example, taking k *copies of the description spaces and cones above yields convex sets described by* k *linear matrix inequalities of the form* (2.1)*, depending on* $11k$ *parameters.*

Structural Aspects of Dimension-Free Descriptions

We have defined freely-described convex sets in Definition 2.1.4 by relating the *descriptions* of the convex sets in a sequence $\{C_n\}$ across dimensions. However, in many applications it is further desirable to directly relate the sets C_n themselves or the functions that arise from them across dimensions. We study the following two relations (which are dual to each other by [189, Cor. 16.3.2]).

Definition 2.1.6 (Compatibility conditions)**.** *Let $\{\mathbb{V}_n\}$ be a nested sequence of vector spaces and let $\{C_n \subseteq \mathbb{V}_n\}$ be a sequence of convex sets (respectively, let $\{f_n \colon \mathbb{V}_n \to \mathbb{R} \cup \{\infty\}\}$ be a sequence of convex functions). We say that $\{C_n\}$ (resp., $\{f_n\}$) satisfies*

Intersection compatibility *if $C_{n+1} \cap \mathbb{V}_n = C_n$ (resp., $f_{n+1}|_{\mathbb{V}_n} = f_n$);*

Projection compatibility *if $\mathcal{P}_{\mathbb{V}_n} C_{n+1} = C_n$ (resp., $\mathcal{P}_{\mathbb{V}_n} f_{n+1} = f_n$).*

Here $(\mathcal{P}_{\mathbb{V}_n} f_{n+1})(x) = \inf_{x' \in \mathcal{P}_{\mathbb{V}_n}^{-1}(x)} f_{n+1}(x')$ defines a convex function on $\mathbb{V}_n$ [189, Thm. 5.7]. One can check that a sequence of convex functions is intersection or projection compatible if and only if its sequence of epigraphs is correspondingly compatible, and a similar correspondence holds for gauge and support functions of compatible sequences of sets; see Section 2.1. In what follows, the nested sequence of spaces in Definition 2.1.6 will always be a consistent sequence.

The conditions in Definition 2.1.6 are natural in a variety of applications. For example, many graph parameters remain unchanged when an isolated vertex is appended to the graph, such as the max cut value; hence these are intersection-compatible with respect to the embeddings in Example 2.1.5. Other parameters, such as the stability number, are nonincreasing when taking induced subgraphs, and any small graph occurs as an induced subgraph of a larger one that has the same parameter value. Hence such parameters are projection-compatible with respect to the same embeddings. In inverse problems, it is desirable to use regularizers that are both intersection and projection compatible. Indeed, if $\{\mathbb{V}_n\}$ is a consistent sequence and we are given data about a signal $x \in \mathbb{V}_N$ that only depends on its projection $\mathcal{P}_{\mathbb{V}_n} x$ onto $\mathbb{V}_n$ for $n < N$, then compatibility of the regularizer ensures that the recovered signal also lies in $\mathbb{V}_n$, see Section 2.4. Compatibility plays a central role in the analysis of approximations for the copositive cones, and the failure of a variant of intersection compatibility for the natural sums-of-squares relaxation of copositive cones underlies several results in this area [95, 134]. Compatibility conditions also arise in noncommutative geometry, where matrix-convex sets are defined as sequences of sets of matrices of each size related across dimensions by conditions stronger, in general, than Definition 2.1.6 (see Proposition 2.4.5);

in particular, the free spectrahedra in Example 1.1.2 satisfy both intersection and projection compatibility.

A freely-described sequence of convex sets need not satisfy either intersection or projection compatibility; see Example 2.3.1 below. Therefore, we investigate freely-described convex sets that additionally satisfy compatibility conditions, and we obtain three structural results pertaining to such sequences of sets. Each of these results is obtained by combining the relevant concepts from representation stability along with notions from convex analysis. We now describe these three results in more detail.

Our first structural result gives conditions under which dimension-free descriptions *certify* compatibility, i.e., under which the sequence of sets derived from these descriptions is evidently compatible. Consider a freely-described sequence $\{C_n\}$ of convex sets (ConicSeq). Assuming that the sequence of cones $\{\mathcal{K}_n\}$ underlying the description of $\{C_n\}$ is both intersection and projection compatible, we prove in Proposition 2.3.2 that the sequence of convex sets $\{C_n\}$ satisfies compatibility conditions if the underlying freely-described elements in (ConicSeq) lie in a convex cone. This result serves as the foundation for our subsequent developments—we use it to show in Section 2.4 that many sets arising in applications naturally admit dimension-free descriptions certifying their compatibility. We also use this result in Section 2.6 to design an algorithm fitting a freely-described and compatible sequence of sets to data.

Our second structural result is central to the development of our computational framework in Section 2.6 and it addresses the following question. Given a convex set with a conic description (Conic) in a fixed dimension n_0, when does this description *extend* to a dimension-free description of a sequence of sets satisfying compatibility conditions? As a concrete example, suppose we are given a convex relaxation for a combinatorial optimization problem, or a convex regularizer for an inverse problem, in a particular problem dimension; under what conditions can these be extended to problems in any desired dimension? Leveraging our preceding result along with properties of presentation degrees of consistent sequences, we give conditions in Theorem 2.3.5 on the dimension n_0 and the conic description in that dimension ensuring the desired extendability. We use this result in Section 2.6 to computationally extend a set fitted to data in a fixed dimension to any other dimension.

Finally, our third structural result develops a notion of a limiting object for a freely-described convex set. Given a consistent sequence $\{\mathbb{V}_n\}$ of G_n-representations, consider the vector space $\mathbb{V}_\infty = \cup_n \mathbb{V}_n$ viewed as a representation of[4] $\mathsf{G}_\infty = \cup_n \mathsf{G}_n$ and let $\overline{\mathbb{V}_\infty}$ denote the completion of $\mathbb{V}_\infty$ with respect to some norm. Consider now a freely-described sequence of convex sets $\{C_n\}$ that is intersection- and projection-compatible. The set $C_\infty = \cup_n C_n$ is a convex subset of $\mathbb{V}_\infty$, and we would like to describe its continuous limit $\overline{C_\infty}$ inside $\overline{\mathbb{V}_\infty}$. This is natural in many applications where problems of different sizes can be naturally viewed as suitable finite-dimensional discretizations of infinite-dimensional problems. Examples include finite graphs obtained via discretization of continuous graphons, vectors obtained via discretizations of continuous-time signals, and matrices obtained via discretizations of operators between infinite dimensional spaces. We show in Theorem 2.3.6 that if the dimension-free description underlying the sequence $\{C_n\}$ certifies our compatibility conditions (in the sense of Proposition 2.3.2) and if the freely-described elements constituting the description extend continuously to their respective limits, then the dimension-free description extends to a description of a dense subset of $\overline{C_\infty}$. In particular, this result yields an infinite-dimensional conic program for optimizing a continuous linear functional over $\overline{C_\infty}$.

Example 2.1.7 (Convex graphon parameters)**.** *Consider the freely-described sequence in Example 1.1.2(c), which is intersection- and projection-compatible with respect to the graphon consistent sequence (see Section 2.4). The elements of $\mathbb{V}_\infty$ are step functions on $[0,1]^2$ known as step graphons, and we endow them with the L_∞ norm as is common in the literature [21]. The completion $\overline{\mathbb{V}_\infty}$ then consists of certain piecewise-continuous graphons, and Theorem 2.3.6 shows that the limit of the sequence $\{C_n\}$ in Example 1.1.2(c) with $L_7 = 0$ is*

$$\overline{C_\infty} = \Bigg\{ W \in \overline{\mathbb{V}_\infty} : 0 \preceq \Bigg[(L_1)_{i,j} \int_{[0,1]^2} W(s,t)\, ds\, dt + (L_2)_{i,j} \int_{[0,1]} W(t,t)\, dt \\ + (L_3)_{i,j} \int_{[0,1]} [W(x,t) + W(t,y)]\, dt + (L_4)_{i,j} [W(x,x) + W(y,y)] + (L_5)_{i,j} W(x,y) \Bigg]_{i,j=1}^{k} \\ + L_6 \Bigg\}.$$

Here positive-semidefiniteness of a matrix-valued function on $[0,1]^2$ is meant in the sense of matrix-valued kernels [155], see (2.21) for a definition. We require $L_7 = 0$ because the corresponding term does not extend continuously with respect to the L_∞ norm. We derive this limiting description in Proposition 2.4.7.

[4] Formally, these are the direct limits $\varinjlim \mathbb{V}_n$ and $\varinjlim \mathsf{G}_n$.

We also apply Theorem 2.3.6 to permutahedra and Schur–Horn orbitopes in Section 2.4 to obtain descriptions of limits in analogy to Example 2.1.7. The resulting infinite-dimensional conic descriptions yield an infinite-dimensional generalization of the Schur–Horn theorem, which we give in Proposition 2.4.3. In fact, we obtain limiting descriptions and a Schur–Horn theorem more generally in approximately finite-dimensional (AF) algebras in Appendix C.

Invariant Conic Programs

Our framework yields structural results not only for descriptions of convex sets but also for sequences of optimization problems over them. Specifically, we use representation stability to unify and generalize a number of previous results which show that various sequences of invariant conic programs indexed by dimension can be solved in constant time. Such sequences of programs arise in several applications including extremal combinatorics [184, 185] and quantum information [105, 125]. Concretely, the programs that arise in certifying homomorphism density inequalities over graphs are invariant under symmetric groups of increasing sizes that relabel the vertices of the graphs involved. Similarly, optimizing traces of matrix polynomials yields programs that are invariant under the unitary or orthogonal groups of increasing sizes that conjugate the matrices involved. We now recall symmetry reductions of invariant programs and explain how constant-sized reductions arise from representation stability.

Consider optimizing a linear functional over a convex subset $C \subseteq \mathbb{V}$ specified as an affine section of a convex cone $\mathcal{K}$ as in (Conic). If the linear functional, the affine section, and the cone are invariant under the action of a group G, then we can further restrict the constraint set to the invariant subspace $\mathbb{V}^{\mathsf{G}}$, thereby reducing the size of the program [86, §3]. Consider now a freely-described sequence of convex subsets $\{C_n\}$ of a consistent sequence $\{\mathbb{V}_n\}$ of $\{\mathsf{G}_n\}$-representations given by (ConicSeq). The vectors and linear maps in (ConicSeq) are G_n-invariant as they are given by freely-described elements; if in addition the cones $\mathcal{K}_n$ are G_n-invariant, then the convex sets C_n are also G_n-invariant. When $\{\mathbb{V}_n\}$ is finitely-generated, so that the spaces of invariants in the sequence are all eventually isomorphic, optimizing a G_n-invariant linear functional over C_n for sufficiently large n reduces as above to optimization over these constant-dimensional spaces of invariants. However, even though the dimensionality of the variables in the symmetry-reduced programs stabilize, the complexity of the constraints might still grow with n.

To establish that the complexity of the constraints also stabilizes with n, we show that the invariant sections $\{\mathcal{K}_n^{\mathsf{G}_n}\}$ of the cones have a constant-sized description in the following precise sense.

Definition 2.1.8 (Constant-sized descriptions)**.** *Let $\{\mathbb{U}_n\}$ be a sequence of $\{\mathsf{G}_n\}$ representations and $\{\mathcal{K}_n \subseteq \mathbb{U}_n\}$ a sequence of convex cones. For $t \in \mathbb{N}$, we say that the sequence $\{\mathcal{K}_n^{\mathsf{G}_n} \subseteq \mathbb{U}_n^{\mathsf{G}_n}\}$ admits a* constant-sized description *for $n \geq t$ if there exists a single finite-dimensional vector space $\mathbb{U}$ containing a cone $\mathcal{K}$, linear maps $T_n \colon \mathbb{U} \to \mathbb{U}_n^{\mathsf{G}_n}$, and subspaces $\mathbb{L}_n \subseteq \mathbb{U}$ such that $\mathcal{K}_n^{\mathsf{G}_n} = T_n(\mathcal{K} \cap \mathbb{L}_n)$ for all $n \geq t$.*

Note that Definition 2.1.8 does not require $\{\mathbb{U}_n\}$ to be a consistent sequence, nor in particular that $\mathbb{U}_n$ is embedded into $\mathbb{U}_{n+1}$. Proofs of constant-sized symmetry reductions in the literature have implicitly proceeded in a case-by-case manner by showing that the relevant cones, including symmetric PSD and relative entropy cones, have a constant-sized description in the sense of our Definition 2.1.8, see [58, 160, 185, 188]. In Section 2.5, we explain how these constant-sized descriptions can be generalized and derived systematically from an interplay between representation stability and the structure of the cones in question. To that end, we use a stronger form of representation stability known as uniform representation stability, which shows that the whole decomposition into irreducibles of the sequences of representations involved stabilize, rather than only their spaces of invariants; see Section 2.2. Our approach allows us to prove the existence of constant-sized descriptions for invariant sections of the following cones under the actions of the classical Weyl groups. The latter consist of the sequence of permutation groups $\{\mathsf{S}_n\}$, signed permutation groups $\{\mathsf{B}_n\}$, and even-signed permutation groups $\{\mathsf{D}_n\}$, see Table 2.1 for definitions.

Theorem 2.1.9 (Informal, see Theorem 2.5.1)**.** *Consider the consistent sequence $\mathscr{V}_0 = \{\mathbb{R}^n\}$ with embeddings by zero-padding and the standard actions of one of the sequences of classical Weyl groups as in Example 2.1.3. For any consistent sequence $\mathscr{V} = \{\mathbb{V}_n\}$ obtained from $\mathscr{V}_0$ by taking finitely-many direct sums, tensor products, symmetric algebras, or skew-symmetric algebras, the invariant sections of the cones $\{\mathrm{Sym}^2_+(\mathbb{V}_n)\}$ admit constant-sized descriptions for $n \geq d+k$, where d and k are the generation and presentation degrees of $\mathscr{V}$, respectively. These degrees can be bounded using the calculus in Theorem 2.2.11.*

The sequences of classical Weyl groups are the groups of permutations $\{\mathsf{S}_n\}$, signed permutations $\{\mathsf{B}_n\}$, or even signed permutations $\{\mathsf{D}_n\}$. Here $\mathrm{Sym}^2_+(\mathbb{V})$ denotes the collection of nonnegative quadratic forms on $\mathbb{V}$. For example, Theorem 2.1.9 shows that the sections of the PSD cones $\{\mathbb{S}^n_+\}$ invariant under the usual action of the classical Weyl groups by conjugation admit constant-sized descriptions. Theorem 2.1.9 is used to derive constant-sized descriptions for cones of invariant sums-of-squares in Theorem 2.1.10 below. In the following result, we consider polynomials in $\binom{n+k-1}{k}$ variables identified with k-multisets of n letters, and show that the cones of sums of squares of such polynomials modulo any sequence of ideals admit constant-sized descriptions. The proof of the following result is given in Section 2.5.

Theorem 2.1.10. *Let $\{\mathsf{G}_n\}$ be one of the sequences of classical Weyl groups acting as usual on $\mathbb{R}^n$. Let*

$$\mathcal{I}_n \subseteq \bigoplus_{d\geq 0} \mathrm{Sym}^d\left(\mathrm{Sym}^k\mathbb{R}^n\right) \cong \mathbb{R}[x_{i_1,\ldots,i_k}]_{1\leq i_1\leq\cdots\leq i_k\leq n} =: \mathbb{V}_n$$

be G_n-invariant ideals, and let $\mathbb{U}_n = \mathrm{Sym}^{\leq 2d}(\mathrm{Sym}^k\mathbb{R}^n)/\mathcal{I}_n$. Consider the sums-of-squares cones $\mathrm{SOS}_{\mathbb{U}_n} = \{f \in \mathbb{U}_n : f \text{ is a sum of squares mod } \mathcal{I}_n\}$. Then the sequence $\{\mathrm{SOS}_{\mathbb{U}_n}^{\mathsf{G}_n}\}$ admits a constant-sized description for $n \geq 2kd$ if $\mathsf{G}_n = \mathsf{S}_n$ or B_n, and for $n \geq 2kd + 1$ if $\mathsf{G}_n = \mathsf{D}_n$.

Theorem 2.1.10 generalizes a number of results from the literature.

- Setting $k = 1$, $\mathcal{I}_n = (0)$, and $\mathsf{G}_n = \mathsf{S}_n$, we recover [188, Thms. 4.7, 4.10].
- Setting $k = 1$, $\mathcal{I}_n = (0)$, and $\mathsf{G}_n = \mathsf{B}_n$ or D_n, we recover [58, Cor. 3.23].
- For $k \geq 2$, setting $\mathcal{I}_n = (x_I - x_I^2, x_J)_{\substack{I\in\binom{[n]}{k}\\ J\notin\binom{[n]}{k}}}$ where $\binom{[n]}{k}$ denotes the set $\{(i_1 < \ldots < i_k) : i_j \in [n]\}$, and setting $\mathsf{G}_n = \mathsf{S}_n$, we recover [185, Thm. 2.4].

Theorem 2.1.10 generalizes all of these to include any of the classical Weyl groups and any sequence of invariant ideals.

Theorem 2.1.10 can be applied to derive constant-sized SDPs to certify graph homomorphism density inequalities [184, 185]. Many problems in extremal combinatorics can be recast as proving polynomial inequalities between homomorphism densities of graphs, which is the fraction of maps between the vertex sets of two graphs that define graph homomorphisms. A simple example is Mantel's theorem, which states that the maximum number of edges in a triangle-free graph is $\lfloor n^2/4\rfloor$. Razborov proposed a method of certifying such inequalities using *flag algebras* [186], which were shown in [184, 185] to be sums-of-squares certificates of certain symmetric polynomial inequalities. Razborov's flags are interpreted in [185, §3] as "free" spanning sets for spaces of symmetric polynomials. Formally, they are freely-described elements in the sense of Definition 2.1.2. Our framework shows that both the existence of such freely-described spanning sets and the resulting constant-sized SDPs are consequences of representation stability.

We obtain similar results for invariant sections of cones derived from relative entropy inequalities that have recently played a role in polynomial optimization [40, 162]. Our main result for these cones is stated in Theorem 2.5.6, which generalizes [160, Thm. 5.3] from permutation groups to the other classical Weyl groups.

Dimension-Free Convex Regression

As our final contribution, we apply our framework to develop an algorithm for a dimension-free analog of the convex regression problem. In the classic setting of this problem, the objective is to fit a convex function $f\colon \mathbb{V} \to \mathbb{R}$ to a dataset $\{(x_i, y_i)\}_{i=1}^D$ of inputs $x_i \in \mathbb{V}$ and outputs $y_i \in \mathbb{R}$ such that $f(x_i) \approx y_i$. In our dimension-free extension of this problem, we have a sequence $\{\mathbb{V}_n\}$ of vector spaces, usually of growing dimension, and data $\{(x_i, y_i) \in \mathbb{V}_{n_i} \oplus \mathbb{R}\}$ in finitely-many dimensions n_i. We then seek an *infinite* sequence of convex functions $\{f_n\colon \mathbb{V}_n \to \mathbb{R}\}$ satisfying $f_{n_i}(x_i) \approx y_i$ in the dimensions in which data is available and which, moreover, is freely-described.

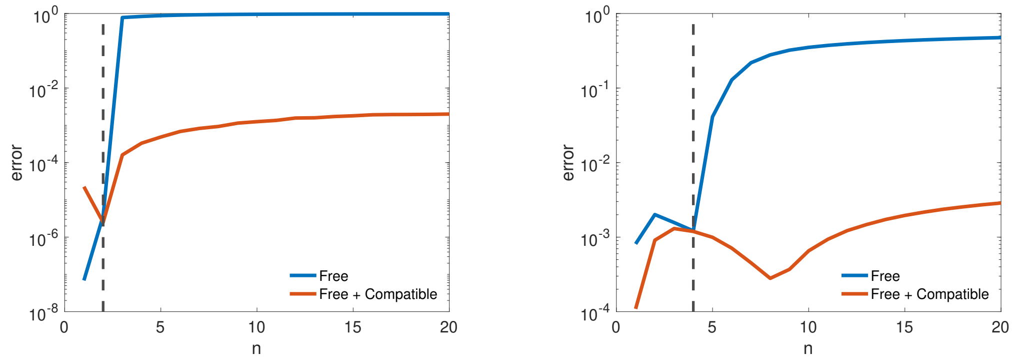
To fit a freely-described sequence of functions, we begin by endowing the vector spaces $\{\mathbb{V}_n\}$ containing the training data with the structure of a consistent sequence based on the symmetries of the problem at hand and the relations between problem instances in different dimensions. We then select description spaces $\{\mathbb{W}_n\}, \{\mathbb{U}_n\}$ and cones $\{\mathcal{K}_n \subseteq \mathbb{U}_n\}$ with respect to which our convex functions are to be described as in (ConicSeq), with freely-described vectors and linear maps parametrizing the desired family of freely-described functions; richer description spaces generally yield more expressive families of convex functions, but fitting a function from such a family is more expensive and requires data in higher dimensions by Theorem 2.3.5.

Having chosen description spaces and cones as above, we have completely specified our family of freely-described functions and turn to the problem of fitting an element of this family to data. We do so in a fully algorithmic fashion, without needing to write down an explicit formula for members of our parametric family, using the following three steps. First, we compute a basis for invariant vectors and linear maps in the dimensions in which we have data using the algorithm of [79]. Second, we numerically identify coefficients in this basis that fit the data using an alternating minimization procedure based on convex duality. Third, we extend the invariant vectors and linear maps we identified in the data dimensions to other dimensions by solving linear systems arising from Definition 2.1.2. Our procedure is summarized in Algorithm 1 and detailed in Section 2.6, and our implementation is publicly available at `https://github.com/eitangl/anyDimCvxSets`. To summarize, once the choice of consistent sequence structure and description spaces is made based on the structure underlying an application, the dimensions of the available data, and the desired richness of the family of functions, the remainder of the procedure is fully computational.

We demonstrate our approach by obtaining semidefinite approximations of two non-semidefinite representable functions: the ℓ_π norm $\|x\|_\pi = (\sum_i |x_i|^\pi)^{1/\pi}$, and the following (nonnegative and positively homogeneous) variant of the quantum entropy function:

$$f(X) = \mathrm{Tr}[(X + \mathrm{Tr}(X)I)\log(X/\mathrm{Tr}(X) + I)]. \tag{2.2}$$

The ℓ_π norm of a vector is defined for vectors of any length, is invariant under signed permutations, remains unchanged under zero-padding of its input, and can only increase if we append any other entry to a given input. It is therefore both intersection and projection compatible with respect to the embeddings and projections of Example 2.1.3. The function (2.2) is well-defined for (symmetric positive-semidefinite) inputs of all sizes, is invariant under conjugation by orthogonal matrices, and its value is unchanged if we zero-pad its input. Thus, it is intersection-compatible on $\{\mathbb{S}^n\}$ with embeddings given by zero-padding. Given evaluations of these two functions on low-dimensional inputs, we use our procedure to fit semidefinite approximations to them by choosing sequences of PSD cones for our descriptions, and fitting a freely-described and compatible sequence of convex functions. Thus, we ensure that our approximations are also group-invariant and satisfy the same compatibility properties as the underlying ground-truth functions. Concretely, we fit an approximation to the ℓ_π norm using its values on 50 random vectors of length at most 2, and fit an approximation to (2.2) using its values on 200 random PSD matrices of size at most 4×4. To evaluate the quality of our approximations, we further generate 10^3 random points in dimensions up to 20 and compare the value of our approximation to the ground truth functions on these points. We detail our description spaces, our fitting algorithm, and our data generation process in Section 2.6. Figure 2.1 shows the error in our semidefinite approximations for different



(a) Learning ℓ_π.

(b) Learning quantum entropy.

Figure 2.1: Errors for learning the ℓ_π norm and the quantum entropy variant (2.2). The dashed vertical lines denote the max n for which data is available.

dimensions when we search over all freely-described sets in our family when fitting to data, and when we search only over the smaller family of compatible sequences. We see that imposing compatibility ensures that the error increases gracefully when extending to higher dimensions.

Obtaining semidefinite approximations for quantum information theoretic quantities such as (2.2) facilitates the use of semidefinite programming solvers in convex optimization problems involving quantum entropy [71–73, 77]. We mention that the authors of [74] derived semidefinite approximations of the quantum entropy and related functions from an analytic perspective. We show in Section 2.4 that their description is dimension-free but does not certify compatibility, in contrast to the description we obtain from data using the preceding procedure.

Related Work

We briefly survey several related areas.

Extended formulations of convex sets: There is a large literature as well as a systematic framework on extended formulations in which conic descriptions of a convex set in a fixed dimension are investigated; see [78] for a review. The goal in this body of work is to express 'complicated' convex sets in $\mathbb{R}^d$ (e.g., polyhedra with many facets) as linear images of affine sections of 'simple' convex cones in a space that is not much larger than d. This framework is also applied to study *equivariant lifts* of group-invariant convex sets, which are descriptions of the form (Conic) consisting of group-invariant cones, vectors, and linear maps, i.e., conic descriptions that make evident the group invariance of the convex set; see [75] and [78, §4.3]. These are precisely the type of descriptions we consider for each of the convex sets in our sequences. Moreover, while the literature on extended formulations is typically articulated in the setting of a convex set in a *fixed* dimension, many results in the area implicitly concern descriptions of a *sequence* $\{C_n\}$ of convex sets and the complexity of these descriptions as a function of n [75, 76, 199, 230]. Thus, there are several points of contact with the present chapter. In fact, many (though not all) descriptions proposed in the literature on extended formulations can in fact be instantiated in any dimension, and are moreover dimension-free in our sense as we show in Section 2.4. More broadly, to the best of our knowledge, descriptions of convex sets that arise from a systematic consideration of relations between dimensions have not been studied previously. By bringing such considerations to the fore, the present effort elucidates the representation-theoretic phenomena and their interaction with the convex-geometric aspects underpinning convex sets that can be instantiated in any dimension.

Free spectrahedra, noncommutative convex algebraic geometry: A broad research program pursued in several areas involves the study of "matrix" or noncommutative analogues of classical "scalar" or commutative objects. Examples include random matrix theory and free probability [157] in which matrix-valued random variables and their limits are the object of study as opposed to scalar-valued ones; and noncommutative algebraic geometry [101], in which polynomials in noncommuting variables and their evaluations on matrices are studied as opposed to standard polynomials in commuting variables that are evaluated on scalars [31, 92, 93, 175]. Applying this program to convex sets yields matrix-convex sets and free spectrahedra as in Example 1.1.2(b). We refer the reader to [101, 130] for surveys and [17, 18] for some applications. In analogy to the present chapter, results in this area explicitly pertain to sequences of sets which are "freely-described", in the sense that their description can be instantiated in any dimension. For example, free spectrahedra are sequences of sets described by a single linear matrix inequality, and free algebraic varieties are defined by the same noncommutative polynomials instantiated on matrices of any size. Another point of contact with our work is the consideration of relations between the sets in the sequence across dimensions, such as matrix-convex combinations which have been formalized and studied in this literature. Our notion of dimension-free descriptions is more general than the ones in this literature however, and it allows us to derive more flexible families of freely-described sets which are adapted to the structure underlying a broader range of applications. Further, the relations between sets in different dimensions we consider in this chapter are less restrictive than matrix convexity, and they yield more general families of sets than free spectrahedra (free spectrahedra may be obtained in our framework via particular instantiations of description spaces, see Section 2.4).

Representation stability: Representation stability arose out of the observation that the decomposition into irreducibles stabilizies for many sequences of representations. This phenomenon has been formalized in [47] using consistent sequences, and it has been further studied in [46, 84, 227] from a categorical perspective and in [194, 195, 197] from a limits-based perspective. We relate the categorical and limits-based formalisms to our setting in appendix A and Section 2.2, respectively, and we refer the reader to [70, 196, 228] for introductions to this area.

Representation stability has been used to study sequences of polyhedral cones and their infinite-dimensional limits [215], as well as sequences of algebraic varieties, their defining equations, and their infinite-dimensional limits [8, 44, 64]. An important distinction between these works and ours is our application of representation stability to *descriptions* of convex sets rather than to their extreme points or rays as in [215]. Thus, we are able to study non-polyhedral sets such as spectrahedra and sets defined by relative entropy programs. Similarly, our study of infinite-dimensional limits in Section 2.3 focuses on limiting descriptions and not just on limits of the sets themselves.

Notation and Basics

We assume familiarity with the basics of representation theory and convex analysis, and we refer the reader to [81, 202] and [189], respectively, for background. In what follows, we review a few basic notions from these areas and introduce notation used throughout the chapter. We list several standard groups and constructions involving vector spaces in Table 2.1.

Basics: We use $[n]$ to denote $\{1, \dots, n\}$. For $i \leq j$ we denote by (i, j) the transposition permuting letters i and j. For real numbers $a < b$, we use $[a, b)$ to denote $\{x \in \mathbb{R} : a \leq x < b\}$. For $i \in [n]$, we denote by $e_i \in \mathbb{R}^n$ the ith standard basis vector with a 1 in the ith entry and zero everywhere else, and we write $e_i^{(n)}$ when we wish to emphasize the dimension. If $x \in \mathbb{R}^n$, we denote by $\mathrm{diag}(x) \in \mathbb{S}^n$ the diagonal matrix with x on the diagonal. If $X \in \mathbb{R}^{n\times n}$, we denote by $\mathrm{diag}(X) \in \mathbb{R}^n$ the vector of its diagonal elements. All vector spaces in this chapter are finite-dimensional real vector spaces equipped with an inner product $\langle \cdot, \cdot \rangle$ unless stated otherwise. We emphasize that some of the inner products we use are nonstandard, so the transpose

Symmetric group	$\mathsf{S}_n = \{g \in \mathbb{R}^{n\times n} : g \text{ is a permutation matrix.}\}$
Signed symmetric group	$\mathsf{B}_n = \{g \in \mathbb{R}^{n\times n} : g \text{ is a signed permutation matrix}\}$
Even-signed symmetric group	$\mathsf{D}_n = \{g \in B_n : g \text{ has an even number of } -1 \text{ entries}\}$
Orthogonal group	$\mathsf{O}_n = \{g \in \mathbb{R}^{n\times n} : g^\top g = I_n\}$
Space of linear maps	$\mathcal{L}(\mathbb{V},\mathbb{U}) = \{A \colon \mathbb{V} \to \mathbb{U} \text{ linear}\}; \qquad \mathcal{L}(\mathbb{V}) = \mathcal{L}(\mathbb{V},\mathbb{V}).$
Direct sum	$\mathbb{V} \oplus \mathbb{U} = \mathbb{V} \times \mathbb{U} = \{(v,u) : v \in \mathbb{V}, u \in \mathbb{U}\}.$
Direct powers	$\mathbb{V}^k = \mathbb{V}^{\oplus k} = \underbrace{\mathbb{V} \oplus \cdots \oplus \mathbb{V}}_{k \text{ times}}.$
Tensor product	$\mathbb{V} \otimes \mathbb{U} = \operatorname{span}\{v \otimes u : v \in \mathbb{V}, u \in \mathbb{U}\} \cong \mathcal{L}(\mathbb{V},\mathbb{U}).$
Tensor power	$\mathbb{V}^{\otimes k} = \underbrace{\mathbb{V} \otimes \cdots \otimes \mathbb{V}}_{k \text{ times}}.$
Symmetric algebra	$\begin{aligned} \operatorname{Sym}^k(\mathbb{V}) &= \operatorname{span}\{v_1 \cdots v_k : v_i \in \mathbb{V}\} \\ &= \{\text{polynomials of degree } = k \text{ on } \mathbb{V}\} \\ &= \{\text{symmetric tensors of order } k \text{ over } \mathbb{V}\} \\ \operatorname{Sym}^{\leq k}(\mathbb{V}) &= \bigoplus_{i=0}^{k} \operatorname{Sym}^i(\mathbb{V}). \end{aligned}$
Alternating algebra	$\begin{aligned} \bigwedge^k \mathbb{V} &= \operatorname{span}\{v_1 \wedge \cdots \wedge v_k : v_i \in \mathbb{V}\} \\ &= \{\text{skew-symmetric tensors of order } k \text{ over } \mathbb{V}\}. \end{aligned}$
Symmetric matrices	$\mathbb{S}^n = \{X \in \mathbb{R}^{n\times n} : X^\top = X\} = \operatorname{Sym}^2(\mathbb{R}^n).$
Skew-symmetric matrices	$\operatorname{Skew}(n) = \{X \in \mathbb{R}^{n\times n} : X^\top = -X\} = \bigwedge^2 \mathbb{R}^n.$
Spaces of invariants	$\mathbb{V}^{\mathsf{G}} = \{v \in \mathbb{V} : g \cdot v = v \text{ for all } g \in \mathsf{G}\},$ $\mathcal{L}(\mathbb{V},\mathbb{U})^{\mathsf{G}} = \{A \in \mathcal{L}(\mathbb{V},\mathbb{U}) : gA = Ag \text{ for all } g \in \mathsf{G}\}.$

Table 2.1: Commonly-used groups and vector spaces. Here $\mathbb{V}, \mathbb{U}$ are finite-dimensional vector spaces.

of a matrix and the adjoint of the linear operator it represents may differ. Given a subspace $\mathbb{W} \subseteq \mathbb{V}$, we denote by $\mathcal{P}_{\mathbb{W}} \colon \mathbb{V} \to \mathbb{W}$ the orthogonal projection onto $\mathbb{W}$. We denote by $\mathbb{R}^n_+$ the cone of entrywise nonnegative length-n vectors, and by $\mathbb{S}^n_+$ the cone of PSD $n \times n$ matrices. If $\mathbb{V}$ is a vector space, we let $\operatorname{Sym}^2_+(\mathbb{V}) \cong \mathbb{S}^{\dim \mathbb{V}}_+$ denote the cone of PSD linear maps in $\mathcal{L}(\mathbb{V})$.

Representation theory: A (linear) action of a group G on a finite-dimensional vector space $\mathbb{V}$ is given by a group homomorphism $\rho \colon \mathsf{G} \to \mathrm{GL}(\mathbb{V})$. Usually ρ is clear from context and we omit it, writing $g \cdot v = \rho(g)v$ for $g \in \mathsf{G}$ and $v \in \mathbb{V}$

instead. All the groups we consider are compact and all group actions are orthogonal, meaning $\langle g \cdot x, g \cdot y\rangle = \langle v, v'\rangle$ for all $x, y \in \mathbb{V}$. The action of G on $\mathbb{V}$ induces an action on $\mathbb{V}^{\otimes k}$ and $\mathrm{Sym}^k(\mathbb{V})$ by setting $g \cdot v_1 \otimes \cdots \otimes v_k = (gv_1) \otimes \cdots \otimes (gv_k)$ and $g \cdot (v_1 \cdots v_k) = (gv_1) \cdots (gv_k)$ and extending by linearity. If $\mathbb{V}$ and $\mathbb{U}$ are both representations of G, we have an action of G on $\mathbb{V} \otimes \mathbb{U}$ by $g \cdot (v \otimes u) = (gv) \otimes (gu)$ (and extending by linearity) and on $\mathcal{L}(\mathbb{V}, \mathbb{U})$ by $g \cdot A = gAg^{-1}$, thus making the representations $\mathbb{V} \otimes \mathbb{U}$ and $\mathcal{L}(\mathbb{V}, \mathbb{U})$ isomorphic. Linear maps invariant under this preceding group action are also called *equivariant* or *intertwining*, since they are precisely the linear maps commuting with the group elements. We denote the group ring of G by $\mathbb{R}[\mathsf{G}] = \mathrm{span}\{e_g\}_{g\in\mathsf{G}}$, where e_g is a basis element indexed by the group element g. Note that a representation of G is the same as a module over the ring $\mathbb{R}[\mathsf{G}]$.

If $\mathsf{H} \subseteq \mathsf{G}$ is a subgroup and $\mathbb{V}$ is a representation of H, the *induced representation* of G from $\mathbb{V}$ is $\mathrm{Ind}_{\mathsf{H}}^{\mathsf{G}}(\mathbb{V}) = \mathbb{R}[\mathsf{G}] \otimes_{\mathbb{R}[\mathsf{H}]} \mathbb{V}$. We have $\dim \mathrm{Ind}_{\mathsf{H}}^{\mathsf{G}}(\mathbb{V}) = |\mathsf{G}/\mathsf{H}| \dim \mathbb{V}$, and we apply this notion only when H has finite index in G. If $g_1 = \mathrm{id}, g_2, \ldots, g_k$ are coset representatives for G/H, we have

$$\mathrm{Ind}_{\mathsf{H}}^{\mathsf{G}}(\mathbb{V}) = \bigoplus_{i=1}^{k} g_i\mathbb{V}, \tag{Ind}$$

together with the following action of G: For each $g \in \mathsf{G}$ and $i \in [k]$, there is a unique $j \in [k]$ and $h \in \mathsf{H}$ satisfying $gg_i = g_jh$, and we let g act on g_iv for $v \in \mathbb{V}$ by $g \cdot g_iv = g_j(h \cdot v)$. This construction is independent of the choice of coset representatives.

As vector spaces, we have an isomorphism $\mathrm{Ind}_{\mathsf{H}}^{\mathsf{G}}(\mathbb{V}) \cong \mathbb{V}^{|\mathsf{G}/\mathsf{H}|}$. Hence, an H-invariant inner product $\langle\cdot,\cdot\rangle$ on $\mathbb{V}$ induces a G-invariant inner product on $\mathbb{V}^{|\mathsf{G}/\mathsf{H}|}$ by setting $\langle g_iv, g_ju\rangle = \delta_{i,j}\langle v, u\rangle$ for $v, u \in \mathbb{V}$ and $i, j \in [|\mathsf{G}/\mathsf{H}|]$. Here $\delta_{i,j} = 1$ if $i = j$ and zero otherwise. We have an isomorphism $(\mathrm{Ind}_{\mathsf{H}}^{\mathsf{G}}\mathbb{V})^{\mathsf{G}} \cong \mathbb{V}^{\mathsf{H}}$ sending $v \in \mathbb{V}^{\mathsf{H}}$ to $\sum_i g_iv \in (\mathrm{Ind}_{\mathsf{H}}^{\mathsf{G}}\mathbb{V})^{\mathsf{H}}$ and $\widetilde{v} \in (\mathrm{Ind}_{\mathsf{H}}^{\mathsf{G}}\mathbb{V})^{\mathsf{H}}$ to $\mathcal{P}_{\mathbb{V}}\widetilde{v} \in \mathbb{V}^{\mathsf{H}}$.

If $\mathsf{H} \subseteq \mathsf{H}'$ and $\mathsf{G} \subseteq \mathsf{G}'$ such that $\mathsf{H}' \cap \mathsf{G} = \mathsf{H}$, then we have the inclusion $\mathsf{G}/\mathsf{H} \hookrightarrow \mathsf{G}'/\mathsf{H}'$ sending $g\mathsf{H} \mapsto g\mathsf{H}'$; this in turn yields an inclusion $\mathrm{Ind}_{\mathsf{H}}^{\mathsf{G}}\mathbb{V} \hookrightarrow \mathrm{Ind}_{\mathsf{H}'}^{\mathsf{G}'}\mathbb{V}$ between induced representations by completing a set of coset representatives for G/H to representatives for G'/H'. Here $\mathbb{V}$ is assumed to be an H'-representation.

If $\mathbb{V}, \mathbb{U}$ are H-representations and $A \in \mathcal{L}(\mathbb{V}, \mathbb{U})^{\mathsf{H}}$, we can extend A to a map $\mathrm{Ind}(A)\colon \mathrm{Ind}_{\mathsf{H}}^{\mathsf{G}}\mathbb{V} \to \mathrm{Ind}_{\mathsf{H}}^{\mathsf{G}}\mathbb{U}$ defined by $\mathrm{Ind}(A)(g_iv) = g_i(Av)$ where g_i is one of the above coset representatives and $v \in \mathbb{V}$. If $\mathbb{V}$ is a G-representation and $\mathbb{W} \subseteq \mathbb{V}$ is an H-subrepresentation, there is a G-equivariant linear map $\mathrm{Ind}_{\mathsf{H}}^{\mathsf{G}}\mathbb{W} \to \mathbb{V}$ sending $g \otimes w \mapsto g \cdot w$ whose image is precisely $\mathbb{R}[\mathsf{G}]\mathbb{W} = \mathrm{span}\{g \cdot w\}_{g\in\mathsf{G}, w\in\mathbb{W}}$.

Convex sets and functions: The epigraph of a convex function $f : \mathbb{V} \to \mathbb{R} \cup \{\infty\}$ is the convex set $\{(x, t) \in \mathbb{V} \oplus \mathbb{R} : f(x) \le t\}$. If $C \subseteq \mathbb{V}$ is a convex set then its gauge function (also called Minkowski functional) is $\gamma_C(x) = \inf\{t : x \in tC\}$ and its support function is $h_C(x) = \sup\{\langle y, x\rangle : y \in C\}$.

Our compatibility conditions for convex sets in Definition 2.1.6 imply compatibility for convex functions derived from them using the above correspondences, and vice versa. Indeed, it is easy to check that a sequence of convex functions is intersection (resp., projection) compatible if and only if the sequence of their epigraphs is intersection (projection) compatible. Similarly, if a sequence of convex sets is intersection (resp., projection) compatible, then the sequence of their gauge functions is intersection (resp., projection) compatible. The correspondences between compatibility conditions between sets and their support functions is a bit subtler. If a sequence of sets is projection-compatible, then the sequence of their support functions is intersection-compatible. If a sequence of *compact* sets is intersection-compatible, then their support functions are projection-compatible.

2.2 Background on Representation Stability

We review some fundamental definitions and results from the representation stability literature, which studies consistent sequences $\{\mathbb{V}_n\}$ of $\{\mathsf{G}_n\}$-representations as in Definition 2.1.1. We further require a notion of maps between consistent sequences, which enables us to define embeddings, quotients, and isomorphisms of consistent sequences.

Definition 2.2.1 (Morphisms of sequences)**.** *If* $\mathscr{V} = \{(\mathbb{V}_n, \varphi_n)\}$ *and* $\mathscr{U} = \{(\mathbb{U}_n, \psi_n)\}$ *are two consistent sequences of* $\{\mathsf{G}_n\}$*-representations, then a* morphism of consistent sequences $\mathscr{A} : \mathscr{V} \to \mathscr{U}$ *is a collection of linear maps* $\mathscr{A} = \{A_n : \mathbb{V}_n \to \mathbb{U}_n\}$ *such that the following hold for each n:*

(a) A_n *is* G_n*-equivariant;*

(b) $A_{n+1}\varphi_n = \psi_n A_n$.

If φ_n and ψ_n are inclusions, condition (b) above becomes $A_{n+1}|_{\mathbb{V}_n} = A_n$. Note that morphisms are freely-described elements (as in Definition 2.1.2) of $\mathscr{V} \otimes \mathscr{U}$, but the converse may not hold. Morphisms of sequences have appeared in the representation stability literature as the natural notion of maps between sequences, see [46, Def. 2.1.1] and [227, §3.2].

Generation Degree

To relate invariants and equivariants across dimensions, we need canonical isomorphisms between spaces of invariants in a consistent sequence. Proposition 2.2.3 below shows that the projections $\mathcal{P}_{\mathbb{V}_n}$ are such isomorphisms, using the following parameter introduced in [46] to control the complexity of a consistent sequence.

Definition 2.2.2 (Generation degree)**.** *A consistent sequence $\mathscr{V} = \{\mathbb{V}_n\}$ of $\{\mathsf{G}_n\}$-representations is* generated in degree d *if $\mathbb{R}[\mathsf{G}_n]\mathbb{V}_d = \mathbb{V}_n$ for all $n \geq d$. The smallest d for which this holds is called the* generation degree *of the sequence. A subset $\mathcal{S} \subseteq \mathbb{V}_d$ is called a* set of generators *for $\mathscr{V}$ if $\mathbb{R}[\mathsf{G}_n]\mathcal{S} = \mathbb{V}_n$ for all $n \geq d$. A sequence is* finitely-generated *if it is generated in degree d for some $d < \infty$.*

Note that $\mathbb{R}[\mathsf{G}_n]\mathcal{S} = \mathrm{span}\{gx\}_{g\in\mathsf{G}_n, x\in\mathcal{S}}$, so that $\mathscr{V}$ is generated in degree d if the span of the G_n-orbit of $\mathbb{V}_d$, when embedded in $\mathbb{V}_n$, is all of $\mathbb{V}_n$ for any $n \geq d$. Note also that if $\mathscr{V}$ is generated in degree d then $\mathbb{V}_d$ is a set of generators for $\mathscr{V}$.

Proposition 2.2.3. *Suppose $\mathscr{V} = \{(\mathbb{V}_n, \varphi_n)\}$ is a consistent sequence of $\{\mathsf{G}_n\}$-representations generated in degree d. Then the restrictions of the projections $\varphi_n^*|_{\mathbb{V}_{n+1}^{\mathsf{G}_{n+1}}} \colon \mathbb{V}_{n+1}^{\mathsf{G}_{n+1}} \to \mathbb{V}_n^{\mathsf{G}_n}$ to spaces of invariants are injective for all $n \geq d$, and are therefore isomorphisms for all large enough n.*

Proof. First, the map φ_n^* is G_n-equivariant because G_n acts orthogonally and $\mathsf{G}_n \subseteq \mathsf{G}_{n+1}$, so it maps G_{n+1}-invariants in $\mathbb{V}_{n+1}$ to G_n-invariants in $\mathbb{V}_n$. Second, suppose $\varphi_n^*(v) = 0$ for some $v \in \mathbb{V}_{n+1}^{\mathsf{G}_{n+1}}$ with $n \geq d$. For any $u \in \mathbb{V}_{n+1}$, write $u = \sum_i g_i\varphi_n(u_i)$ where $u_i \in \mathbb{V}_n$ and $g_i \in \mathsf{G}_{n+1}$. Because v is G_{n+1}-invariant, we have $\langle v, u\rangle = \langle \varphi_n^*(v), \sum_i u_i\rangle = 0$. As $u \in \mathbb{V}_{n+1}$ was arbitrary, we conclude that $v = 0$. Thus, φ_n^* maps $\mathbb{V}_{n+1}^{\mathsf{G}_{n+1}}$ injectively into $\mathbb{V}_n^{\mathsf{G}_n}$, so that $\dim \mathbb{V}_n^{\mathsf{G}_n} \geq \dim \mathbb{V}_{n+1}^{\mathsf{G}_{n+1}}$, for all $n \geq d$. Therefore, the sequence of dimensions $\dim \mathbb{V}_n^{\mathsf{G}_n}$ eventually stabilizes, at which point φ_n^* becomes an isomorphism. $\square$

Note that $\varphi_n^* = \mathcal{P}_{\mathbb{V}_n}$ is precisely the orthogonal projection onto $\mathbb{V}_n$. Proposition 2.2.3 is stated in the representation stability literature in terms of the adjoints of the projections, viewed as maps between *co*invariants, see [46, §3] for example.

Presentation Degree

While boundedness of the generation degree of a consistent sequence ensures that the projections eventually become isomorphisms, providing a precise quantification of this phenomenon requires a more sophisticated concept, namely the *presentation degree*. We describe this concept after giving some preliminary definitions. Our presentation here is brief, and we refer the reader to Appendix B for a more detailed derivation of these notions motivated by our computational considerations.

Definition 2.2.4 (Centralizing subgroups)**.** *Let $\mathscr{V} = \{\mathbb{V}_n\}$ be a consistent sequence of $\{\mathsf{G}_n\}$-representations. For any $d \leq n$, define the* centralizing subgroups *of $\mathbb{V}_d$ by*

$$\mathsf{H}_{n,d} = \{g \in \mathsf{G}_n : g \cdot v = v \text{ for all } v \in \mathbb{V}_d\}.$$

Note that the subgroup generated by G_d and $\mathsf{H}_{n,d}$ inside G_n is the set of products $\mathsf{G}_d\mathsf{H}_{n,d} = \{gh : g \in \mathsf{G}_d,\ h \in \mathsf{H}_{n,d}\}$ since $ghg^{-1} \in \mathsf{H}_{n,d}$ for all $g \in \mathsf{G}_d$ and $h \in \mathsf{H}_{n,d}$.

Definition 2.2.5 ($\mathscr{V}$-modules)**.** *Let $\mathscr{V} = \{\mathbb{V}_n\}$ and $\mathscr{U} = \{\mathbb{U}_n\}$ be consistent sequences of $\{\mathsf{G}_n\}$-representations, and let $\{\mathsf{H}_{n,d}\}_{d\leq n}$ be the centralizing subgroups of $\mathscr{V}$ as in Definition 2.2.4. We say that $\mathscr{U}$ is a $\mathscr{V}$*-module *if $\mathbb{U}_d \subseteq \mathbb{U}_n^{\mathsf{H}_{n,d}}$ for all $d \leq n$.*

Definition 2.2.6 (Induction and algebraically free[5] sequences)**.** *Let $\mathscr{V}$ be a consistent sequence of $\{\mathsf{G}_n\}$-representations, and for $d \leq n$ let $\mathsf{H}_{n,d} \subseteq \mathsf{G}_n$ be its centralizing subgroups.*

(a) Fix $d \in \mathbb{N}$ and a G_d-representation $\mathbb{W}$ on which $\mathsf{H}_{d,d}$ acts trivially. For each $n \geq d$, we view $\mathbb{W}$ as a $\mathsf{G}_d\mathsf{H}_{n,d}$-representation on which $\mathsf{H}_{n,d}$ acts trivially. The associated $\mathscr{V}$-induction sequence *is defined as:*

$$\mathrm{Ind}_{\mathsf{G}_d}(\mathbb{W}) = \left\{\mathrm{Ind}_{\mathsf{G}_d\mathsf{H}_{n,d}}^{\mathsf{G}_n}\mathbb{W}\right\}_n,$$

where the induced representation is taken to be 0 when $n < d$. This is a $\mathscr{V}$-module.

(b) A consistent sequence $\mathscr{F}$ is an algebraically free $\mathscr{V}$-module *if it is a direct sum of $\mathscr{V}$-induction sequences. The sequence $\mathscr{V}$ itself is* algebraically free *if it is an algebraically free $\mathscr{V}$-module.*

[5]Freeness here is meant in the algebraic sense of being generated by generators with no nontrivial relations between them (see Appendix B), in contrast to Definition 2.1.4 where it is meant in the sense of dimension-free descriptions.

Definition 2.2.7 (Relation and presentation degrees)**.** *Let $\mathscr{V}$ be a consistent sequence of $\{\mathsf{G}_n\}$-representations. We say that a $\mathscr{V}$-module $\mathscr{U}$ is* generated in degree d, related in degree r, and presented in degree $k = \max\{d, r\}$ *if there exists an algebraically free $\mathscr{V}$-module $\mathscr{F}$ generated in degree d, and a surjective morphism of sequences $\mathscr{F} \to \mathscr{U}$ whose kernel is generated in degree r. The smallest k for which this holds is called the* presentation degree *of $\mathscr{U}$.*

Note that the presentation degree is at least as large as the generation degree (cf. Definition 2.2.2).

Example 2.2.8. *Let $\mathbb{V}_n = \mathbb{R}^n$ with embeddings by zero-padding as in Example 2.1.3. Recall that this is a consistent sequence for each of the sequences of groups $\mathsf{G}_n = \mathsf{O}_n, \mathsf{B}_n, \mathsf{D}_n, \mathsf{S}_n$ acting by their standard $n \times n$ matrix representations. Here $\mathsf{H}_{n,d}$ is the subgroup of $n \times n$ orthogonal or signed permutation matrices fixing the first d coordinates.*

This sequence is generated in degree 1 for all the groups listed above. Indeed, any of the canonical basis vectors e_i are obtained from the first one e_1 via the action of S_n.

If $\mathsf{G}_n = \mathsf{B}_n$ or S_n then this sequence is algebraically free and hence presented in degree 1 as well, while if $\mathsf{G}_n = \mathsf{D}_n$ then it is not free but presented in degree 2. Indeed, we have $|\mathsf{D}_n/\mathsf{D}_1\mathsf{H}_{n,1}| = 2n$ when $n \geq 2$ with coset representatives $(1, i)s^p$ for $p \in \{0, 1\}$, $i \in [n]$ where $s = \mathrm{diag}(-1, -1, 1 \dots, 1)$. Hence (Ind) yields $\mathrm{Ind}_{\mathsf{D}_1\mathsf{H}_{n,1}}^{\mathsf{D}_n}\mathbb{R} = \mathbb{R}^n \oplus \mathbb{R}^n$ on which $\sigma \in \mathsf{S}_n \subseteq \mathsf{D}_n$ acts by $\sigma(x, y) = (\sigma x, \sigma y)$ and $s(x, y) = ([y_1, y_2, x_3, \dots, x_n]^\top, [x_1, x_2, y_3, \dots, y_n]^\top)$. We have equivariant linear maps $\mathrm{Ind}_{\mathsf{D}_1\mathsf{H}_{n,1}}^{\mathsf{D}_n}\mathbb{R} \to \mathbb{R}^n$ sending $(x, y) \mapsto x - y$ giving a morphism of sequences $\mathrm{Ind}_{\mathsf{D}_1}\mathbb{R} \to \{\mathbb{R}^n\}$ with kernel generated in degree 2.

The presentation degree enables us to strengthen Proposition 2.2.3 and to quantify more precisely when the projections there become isomorphisms.

Proposition 2.2.9. *Let $\mathscr{V}$ be a consistent sequence of $\{\mathsf{G}_n\}$-representations and $\mathscr{U}$ be a $\mathscr{V}$-module presented in degree k. Then the maps $\mathcal{P}_{\mathbb{U}_n} : \mathbb{U}_{n+1}^{\mathsf{G}_{n+1}} \to \mathbb{U}_n^{\mathsf{G}_n}$ are isomorphisms for all $n \geq k$.*

Proof. As $\mathscr{U}$ is presented in degree k, there exists an algebraically free $\mathscr{V}$-module $\mathscr{F} = \{\mathbb{F}_n\}$ and a surjective morphism $\mathscr{F} \to \mathscr{U}$ such that both its kernel $\mathscr{K} = \{\mathbb{K}_n\}$

and $\mathscr{F}$ itself are generated in degree k. Because each map $\mathbb{F}_n \to \mathbb{U}_n$ is a G_n-equivariant surjection with kernel $\mathbb{K}_n$, its restriction to invariants $\mathbb{F}_n^{\mathsf{G}_n} \to \mathbb{U}_n^{\mathsf{G}_n}$ is surjective with kernel $\mathbb{K}_n^{\mathsf{G}_n}$.

As $\mathscr{F}$ is an algebraically free $\mathscr{V}$-module, there exist integers d_j and G_{d_j}-representations $\mathbb{W}_{d_j}$ satisfying $\mathscr{F} = \bigoplus_j \mathrm{Ind}_{\mathsf{G}_{d_j}} \mathbb{W}_{d_j}$. Such $\mathscr{F}$ has generation degree $\max_j d_j \le k$. Therefore, letting $\{\mathsf{H}_{n,d}\}$ be the centralizing subgroups of $\mathscr{V}$, we have for $n \ge k$ (see Section 2.1)

$$\mathbb{F}_n^{\mathsf{G}_n} = \bigoplus_j \left(\mathrm{Ind}_{\mathsf{G}_{d_j}\mathsf{H}_{n,d_j}}^{\mathsf{G}_n}(\mathbb{W}_{d_j})\right)^{\mathsf{G}_n} \cong \bigoplus_j \mathbb{W}_{d_j}^{\mathsf{G}_{d_j}},$$

Thus, $\dim \mathbb{F}_n^{\mathsf{G}_n}$ is constant for $n \ge k$. Moreover, by Proposition 2.2.3 and the fact that $\mathscr{K}$ and $\mathscr{U}$ are generated in degree k, we have $\dim \mathbb{K}_n^{\mathsf{G}_n} \ge \dim \mathbb{K}_{n+1}^{\mathsf{G}_{n+1}}$ and similarly $\dim \mathbb{U}_n^{\mathsf{G}_n} \ge \dim \mathbb{U}_{n+1}^{\mathsf{G}_{n+1}}$ for all $n \ge k$.

By the rank-nullity theorem, we have $\dim \mathbb{U}_n^{\mathsf{G}_n} = \dim \mathbb{F}_n^{\mathsf{G}_n} - \dim \mathbb{K}_n^{\mathsf{G}_n}$. As $\dim \mathbb{F}_n^{\mathsf{G}_n}$ is constant while both $\dim \mathbb{U}_n^{\mathsf{G}_n}$ and $\dim \mathbb{K}_n^{\mathsf{G}_n}$ are nonincreasing for $n \ge k$, we conclude that they are all constant for $n \ge k$. To conclude, we note that $\mathcal{P}_{\mathbb{U}_n}$ is injective when restricted to $\mathbb{U}_{n+1}^{\mathsf{G}_{n+1}}$ for all $n \ge k$ by Proposition 2.2.3. □

Remark 2.2.10 ($\mathscr{V}$-modules vs. centralizing subgroups)**.** *The definition of presentation degree assumes a "base" consistent sequence $\mathscr{V}$. Note however that it depends only on the centralizing subgroups $\{\mathsf{H}_{n,d}\}$ of $\mathscr{V}$. In fact, any sequence of subgroups $\{\mathsf{H}_{n,d} \subseteq \mathsf{G}_n\}_{d\le n}$ satisfying $\mathsf{H}_{n+1,d} \supseteq \mathsf{H}_{n,d}$, $\mathsf{H}_{n,d+1} \subseteq \mathsf{H}_{n,d}$, and $\mathsf{H}_{n+1,d} \cap \mathsf{G}_n = \mathsf{H}_{n,d}$ for $d \le n$ arises as the sequence of centralizing subgroups of such a consistent sequence.*

The centralizing subgroups play a central role because they determine embeddings $\{g\varphi_{n-1}\cdots\varphi_d\}_{g\in\mathsf{G}_n} \cong \mathsf{G}_n/\mathsf{H}_{n,d}$ of $\mathbb{V}_d$ into $\mathbb{V}_n$, and the combinatorics of these embeddings yields Theorem 2.2.11. See the proof of [46, Prop. 2.3.6] and Appendix A for example. We formulate our results in terms of $\mathscr{V}$-modules rather than the subgroups $\{\mathsf{H}_{n,d}\}$ directly because the sequences we use are often constructed from the same base sequence as in Section 2.2 below, easing the application of our results.

Constructions of Consistent Sequences

Expressive families of freely described convex sets require complex description spaces, and in turn complex consistent sequences. In this section, we describe common operations that yield complicated consistent sequences from simpler ones, along with a calculus for bounding the generation and presentation degrees of the resulting sequences.

Fix a family of groups $\mathscr{G} = \{\mathsf{G}_n\}_{n\in\mathbb{N}}$ such that $\mathsf{G}_n \subseteq \mathsf{G}_{n+1}$. Suppose $\mathscr{V} = \{(\mathbb{V}_n, \varphi_n)\}$ and $\mathscr{U} = \{(\mathbb{U}_n, \psi_n)\}$ are consistent sequences of $\mathscr{G}$-representations. Then the following are also consistent sequences of $\mathscr{G}$-representations:

(Sums) The *direct sum* of $\mathscr{V}$ and $\mathscr{U}$ is $\mathscr{V} \oplus \mathscr{U} = \{(\mathbb{V}_n \oplus \mathbb{U}_n, \varphi_n \oplus \psi_n)\}$.

If $\mathbb{W}$ is a fixed vector space, viewed as a trivial G_n-representation for all n, denote $\mathscr{V} \oplus \mathbb{W} = \{(\mathbb{V}_n \oplus \mathbb{W}, \varphi_n \oplus \mathrm{id}_{\mathbb{W}})\}$.

(Tensors) The *tensor product* of $\mathscr{V}$ and $\mathscr{U}$ is $\mathscr{V} \otimes \mathscr{U} = \{(\mathbb{V}_n \otimes \mathbb{U}_n, \varphi_n \otimes \psi_n)\}$.

This is also the sequence of spaces of *linear maps* $\mathcal{L}(\mathbb{V}_n, \mathbb{U}_n) \cong \mathbb{V}_n \otimes \mathbb{U}_n$, where we embed $A_n\colon \mathbb{V}_n \to \mathbb{U}_n$ to $(\varphi_n \otimes \psi_n)A_n = \psi_n A_n \varphi_n^*\colon \mathbb{V}_{n+1} \to \mathbb{U}_{n+1}$.

The *order-k tensors over $\mathscr{V}$* is $\mathscr{V}^{\otimes k}$.

If $\mathbb{W}$ is a fixed vector space, viewed as a trivial G_n-representation for all n, denote $\mathscr{V} \otimes \mathbb{W} = \{(\mathbb{V}_n \otimes \mathbb{W}, \varphi_n \otimes \mathrm{id}_{\mathbb{W}})\}$. Since $\mathbb{V}_n \otimes \mathbb{W} \cong \mathbb{V}_n^{\oplus \dim \mathbb{W}}$ as G_n-representations, we can identify $\mathscr{V} \otimes \mathbb{W}$ with $\mathscr{V}^{\oplus \dim \mathbb{W}}$.

(Polynomials) The *degree-k polynomials over $\mathscr{V}$* is $\mathrm{Sym}^k\mathscr{V} = \{(\mathrm{Sym}^k\mathbb{V}_n, \varphi_n^{\otimes k})\}$, which is also the sequence of order-k symmetric tensors over $\mathscr{V}$. Here we view $\mathrm{Sym}^k\mathbb{V}_n \subseteq \mathbb{V}^{\otimes k}$ and restrict $\varphi_n^{\otimes k}$ to that subspace. The sequence of polynomials of degree at most k is denoted $\mathrm{Sym}^{\leq k}\mathscr{V} = \bigoplus_{j=1}^k \mathrm{Sym}^j\mathscr{V}$.

Similarly, we can form the sequence of kth exterior powers $\bigwedge^k \mathscr{V}$.

(Moments) The sequence of *moment matrices of order k over $\mathscr{V}$* is $\mathrm{Sym}^2(\mathrm{Sym}^{\leq k}\mathscr{V})$. Its elements can be viewed as symmetric matrices whose rows and columns are indexed by monomials of degree at most k in basis elements for $\mathscr{V}$.

(Images and Kernels) If $\mathscr{A} = \{A_n \in \mathcal{L}(\mathbb{V}_n, \mathbb{U}_n)^{\mathsf{G}_n}\}$ is a morphism mapping $\mathscr{V} \to \mathscr{U}$, then the images $\mathrm{Im}\mathscr{A} = \{(A_n(\mathbb{V}_n), \psi_n)\}$ and kernels $\ker\mathscr{A} = \{(\ker A_n, \varphi_n)\}$ form consistent sequences.

If $\mathscr{V}, \mathscr{U}$ are $\mathscr{V}_0$-modules for some common consistent sequence $\mathscr{V}_0$, then all the above are $\mathscr{V}_0$-modules as well.

The group actions above are given in Section 2.1. The following theorem gives a calculus that allows us to bound the presentation degrees of sequences constructed from certain simpler ones with known presentation degrees. The following theorem is a consequence of results in [46, 84, 227] concerning calculus for generation degrees. We combine these results to obtain the following analogous calculus for presentation degrees, whose proof is given in Appendix A.

Theorem 2.2.11 (Calculus for generation and presentation degrees)**.** *Let $\mathscr{V}$ be a consistent sequence of $\{\mathsf{G}_n\}$-representations and let $\mathscr{W}, \mathscr{U}$ be $\mathscr{V}$-modules generated in degrees d_W, d_U and presented in degrees k_W, k_U, respectively. Then*

(Sums) *$\mathscr{W} \oplus \mathscr{U}$ is generated in degree $\max\{d_W, d_U\}$ and presented in degree $\max\{k_W, k_U\}$.*

(Images and kernels) *If $\mathscr{A} : \mathscr{W} \to \mathscr{U}$ is a morphism consisting of linear maps $\mathscr{A} = \{A_n \in \mathcal{L}(\mathbb{W}_n, \mathbb{U}_n)^{\mathsf{G}_n}\}$, and if the sequence of their adjoints $\mathscr{A}^* = \{A_n^* \in \mathcal{L}(\mathbb{U}_n, \mathbb{W}_n)^{\mathsf{G}_n}\}$ is a morphism $\mathscr{U} \to \mathscr{W}$, then $\mathrm{im}\mathscr{A}$ and $\ker \mathscr{A}$ are generated in degree d_W and presented in degree k_W.*

If $\mathscr{V} = \{\mathbb{R}^n\}$ with $\mathsf{G}_n = \mathsf{B}_n, \mathsf{D}_n$, or S_n as in Example 2.1.3, we further have

(Tensors) *$\mathscr{W} \otimes \mathscr{U}$ is generated in degree $d_W + d_U$ and presented in degree $\max\{k_W + d_U, k_U + d_W\}$.*

(Sym and $\bigwedge$) *$\mathrm{Sym}^\ell \mathscr{W}$, $\bigwedge^\ell \mathscr{W}$ are generated in degree ℓd_W and presented in degree $(\ell - 1)d_W + k_W$.*

Proof. This follows from Theorem A.0.13 in the appendix. □

Example 2.2.12. *Suppose $\{\mathbb{R}^n\}$ as in Example 2.1.3 with $\mathsf{G}_n = \mathsf{B}_n, \mathsf{D}_n$, or S_n. Then $(\mathbb{R}^n)^{\otimes k}$ consists of $n \times \cdots \times n$-sized tensors with embeddings by zero padding and $\mathrm{Sym}^k\mathbb{R}^n$ consists of homogeneous polynomials of degree k in n variables. These are generated in degree k and presented in degree k if $\mathsf{G}_n = \mathsf{B}_n, \mathsf{S}_n$ and in degree $k + 1$ if $\mathsf{G}_n = \mathsf{D}_n$.*

Remark 2.2.13 (Other groups)**.** *Note that the last two conclusions in Theorem 2.2.11 fail without the restriction to $\mathsf{G}_n = \mathrm{B}_n, \mathrm{D}_n$, or S_n. For example, consider the sequence of cyclic groups $\mathsf{G}_n = \mathrm{Cyc}_{2^n}$ embedded as usual in each other. Let $\mathscr{V} = \{\mathbb{V}_n = \mathbb{R}^{2^n}\}$ be the consistent sequence with embeddings $\varphi_n(x) = x \otimes [1, 0]^\top$ and the standard action of $\mathsf{G}_n = \mathrm{Cyc}_{2^n}$ by cyclically shifting coordinates. Then $\mathscr{V}$ is generated in degree 1 but $\mathscr{V}^{\otimes 2}$ is not finitely-generated since $\dim \mathcal{L}(\mathbb{R}^{2^n})^{\mathrm{Cyc}_{2^n}} = 2^n$ does not stabilize.*

Permutation Modules

We introduce a class of particularly simple consistent sequences on which the group acts by permuting basis elements. These consistent sequences arise in our study of relative entropy cones and their constant-sized descriptions in Section 2.5. If a group G acts on a (finite) set $\mathcal{A}$, define $\mathbb{R}^{\mathcal{A}} = \bigoplus_{\alpha\in\mathcal{A}} e_\alpha$ to be the vector space with orthonormal basis elements $\{e_\alpha\}_{\alpha\in\mathcal{A}}$, which is a G-representation with action $g \cdot e_\alpha = e_{g\cdot\alpha}$.

Definition 2.2.14 (Permutation modules)**.** *Let $\mathscr{V} = \{\mathbb{V}_n\}$ be a consistent sequence of $\{\mathsf{G}_n\}$-representations. Let $\{\mathcal{A}_n \subseteq \mathbb{V}_n\}$ be finite G_n-invariant sets satisfying $\mathcal{A}_n \subseteq \mathcal{A}_{n+1}$ for all n. Then the* permutation $\mathscr{V}$-module *generated by the sets $\{\mathcal{A}_n\}$ is the $\mathscr{V}$-module $\{\mathbb{R}^{\mathcal{A}_n}\}_n$.*

Permutation modules can be analyzed in terms of the orbits in the sets $\mathcal{A}_n$. In particular, indicators of orbits form a basis for the space of invariants in a permutation module.

Proposition 2.2.15. *Let $\mathscr{V} = \{\mathbb{V}_n\}$ be a consistent sequence of $\{\mathsf{G}_n\}$-representations, let $\{\mathcal{A}_n \subseteq \mathbb{V}_n\}$ be a nested sequence of finite group-invariant sets, and let $\mathscr{U} = \{\mathbb{U}_n = \mathbb{R}^{\mathcal{A}_n}\}$ be the associated permutation $\mathscr{V}$-module.*

(a) $\mathscr{U}$ is generated in degree d if and only if $\mathcal{A}_n = \bigcup_{g\in\mathsf{G}_n} g\mathcal{A}_d$ for all $n \geq d$.

(b) The projections $\mathcal{P}_{\mathbb{U}_n}\colon \mathbb{U}_{n+1}^{\mathsf{G}_{n+1}} \to \mathbb{U}_n^{\mathsf{G}_n}$ are isomorphisms for all $n \geq d$ if and only if either of the equivalent conditions in (a) holds and the number of orbits in $\mathcal{A}_n$, which equals $\dim \mathbb{U}_n^{\mathsf{G}_n}$, is constant for all $n \geq d$.

(c) Suppose $\mathscr{V} = \{\mathbb{R}^n\}$ and $\mathsf{G}_n = \mathsf{B}_n, \mathsf{D}_n$, or S_n as in Example 2.1.3 and $\mathscr{U}$ is generated in degree d. Then $\mathscr{U}$ is free if $\mathsf{G}_n = \mathsf{B}_n, \mathsf{S}_n$, and agrees with a free module starting from degree $d+1$ if $\mathsf{G}_n = \mathsf{D}_n$.

Here we say $\mathscr{U}$ agrees with a free module starting from degree d if there is a free $\mathscr{V}$-module $\mathscr{F} = \{\mathbb{F}_n\}$ and a morphism of sequences $\mathscr{F} \to \mathscr{U}$ such that $\mathbb{F}_n \to \mathbb{U}_n$ is an isomorphism for all $n \geq d$.

Proof. (a) We have $\mathbb{R}[\mathsf{G}_n]\mathbb{V}_d = \sum_{\substack{\alpha\in\mathcal{A}_d\\ g\in\mathsf{G}_n}} \mathbb{R}e_{g\alpha} = \bigoplus_{\alpha\in\bigcup_{g\in\mathsf{G}_n} g\mathcal{A}_d} \mathbb{R}e_\alpha$ which equals $\mathbb{V}_n = \bigoplus_{\alpha\in\mathcal{A}_n} \mathbb{R}e_\alpha$ precisely under the stated condition.

(b) Let $O_1^{(n)}, \dots, O_{D(n)}^{(n)}$ be the G_n-orbits in $\mathcal{A}_n$ and set $O_i^{(n+1)} = \mathsf{G}_{n+1}O_i^{(n)}$ for $i \in [D(n)]$. Note that $\mathbb{U}_n^{\mathsf{G}_n} \cong \mathbb{R}^{D(n)}$ has a basis consisting of orbit indicators $\mathbb{1}_{O_i^{(n)}} = \sum_{\alpha\in O_i^{(n)}} e_\alpha$. Moreover, $\mathcal{P}_{\mathbb{U}_n}\mathbb{1}_{O_i^{(n+1)}} = \mathbb{1}_{O_i^{(n+1)}\cap\mathcal{A}_n} = \sum_{\alpha\in O_i^{(n+1)}\cap\mathcal{A}_n} e_\alpha$. If $\mathcal{A}_{n+1} = \bigcup_{g\in\mathsf{G}_{n+1}} g\mathcal{A}_n$ and the number of orbits in $\mathcal{A}_n$ and $\mathcal{A}_{n+1}$ are equal, then $O_i^{(n+1)}$ are precisely the orbits in $\mathcal{A}_{n+1}$. This implies that $O_i^{(n+1)} \cap \mathcal{A}_n = O_i^{(n)}$ for $i \in [D(n)]$, and hence thta $\mathcal{P}_{\mathbb{U}_n}$ is an isomorphism of invariants. Indeed, the inclusion $\supseteq$ is trivial while if $\alpha \in \left(O_i^{(n+1)} \cap \mathcal{A}_n\right) \setminus O_i^{(n)}$ then $\alpha \in O_j^{(n)}$ for some $j \neq i$ and $\mathsf{G}_{n+1}O_i^{(n)} = \mathsf{G}_{n+1}O_j^{(n)}$ since these two G_{n+1}-orbits are not disjoint (they both contain α), and hence must be equal. This contradicts the fact that $O_i^{(n+1)}$ are distinct. Conversely, suppose $\mathcal{P}_{\mathbb{U}_n}$ is an isomorphism of invariants. If $\alpha \in \mathcal{A}_{n+1} \setminus \bigcup_{g\in\mathsf{G}_{n+1}} g\mathcal{A}_n$ then $\mathsf{G}_{n+1}\alpha \cap \mathcal{A}_n = \emptyset$ and $\mathcal{P}_{\mathbb{U}_n}\mathbb{1}_{\mathsf{G}_{n+1}\alpha} = 0$, a contradiction. Therefore, we have $\mathcal{A}_{n+1} = \bigcup_{g\in\mathsf{G}_{n+1}} g\mathcal{A}_n$. Furthermore, the number of orbits in $\mathcal{A}_n$ and $\mathcal{A}_{n+1}$ are equal, since these are the dimensions of $\mathbb{U}_n^{\mathsf{G}_n}$ and $\mathbb{U}_{n+1}^{\mathsf{G}_{n+1}}$, respectively.

(c) Let $d_0 = 0$ if $\mathsf{G}_n = \mathsf{S}_n, \mathsf{B}_n$ and $d_0 = 1$ if $\mathsf{G}_n = \mathsf{D}_n$. Let $\hat{\mathcal{A}} \subseteq \mathcal{A}_d$ be a set of minimal degree G_{d+d_0}-orbit representatives in $\mathcal{A}_{d+d_0}$. We argue that these are also the G_n-orbit representatives of $\mathcal{A}_n$ for all $n \geq d + d_0$. Indeed, since $\mathcal{A}_n = \bigcup_{g\in\mathsf{G}_n} g\mathcal{A}_d = \bigcup_{g\in\mathsf{G}_n} g\hat{\mathcal{A}}$, it suffices to show that distinct elements in $\hat{\mathcal{A}}$ lie in distinct G_n-orbits. To that end, observe that $\mathscr{V}$ satisfies

$$g \cdot \alpha \in \mathbb{V}_d \text{ for } \alpha \in \mathbb{V}_d,\ g \in \mathsf{G}_n \implies \exists \tilde{g} \in \mathsf{G}_{d+d_0} \text{ s.t. } g \cdot \alpha = \tilde{g} \cdot \alpha, \quad (2.3)$$

since if $\mathsf{G}_n = \mathsf{S}_n, \mathsf{B}_n$ define $\tilde{g}$ to act as g on the coordinates $I = \{i \in [d] : \alpha_i \neq 0\}$ and act trivially on the others, and if $\mathsf{G}_n = \mathsf{D}_n$ and if g flips oddly-many signs of coordinates in I then in addition have $\tilde{g}$ flip the sign of coordinate $d+1$. Therefore, if $\alpha, \alpha' \in \mathcal{A}_d$ lie in the same G_n-orbit for $n > d + d_0$, then they also lie in the same G_{d+d_0}-orbit. Thus, we have $\mathbb{U}_n = \bigoplus_{\alpha\in\hat{\mathcal{A}}} \mathbb{R}[\mathsf{G}_n]e_\alpha$ for all $n \geq d + d_0$.

Next, we argue that $\mathscr{U}$ is free or agrees with a free module starting from degree $d + d_0$. Observe that $\mathscr{V}$ satisfies the additional property

$$\alpha \in \mathbb{V}_d\setminus\mathbb{V}_{d-1} \text{ has min. degree in its orbit} \implies \mathsf{Stab}_{\mathsf{G}_n}(\alpha) = \mathsf{Stab}_{\mathsf{G}_d}(\alpha)\mathsf{H}_{n,d}, \quad (2.4)$$

Indeed, if $\alpha \in \mathbb{V}_d \setminus \mathbb{V}_{d-1}$ has minimal degree then all its d entries are nonzero, hence any $g \in \mathsf{G}_n$ fixing α must fix the first d coordinates. Therefore, if $\alpha \in \mathbb{V}_d\setminus\mathbb{V}_{d-1}$ has minimal degree and $g_1, g_2, \dots, g_M$ are coset representatives for $\mathsf{G}_n/\mathsf{Stab}_{\mathsf{G}_d}(\alpha)\mathsf{H}_{n,d}$, then

$$\begin{aligned}\mathbb{R}[\mathsf{G}_n]e_\alpha = \bigoplus_{m=1}^{M}\mathbb{R}e_{g_m\cdot\alpha} = \bigoplus_{m=1}^{M} g_m\cdot\mathbb{R}e_\alpha = \mathrm{Ind}_{\mathsf{Stab}_{\mathsf{G}_d}(\alpha)\mathsf{H}_{n,d}}^{\mathsf{G}_n}(\mathbb{R}e_\alpha)\\ = \mathrm{Ind}_{\mathsf{G}_d\mathsf{H}_{n,d}}^{\mathsf{G}_n}\left(\mathrm{Ind}_{\mathsf{Stab}_{\mathsf{G}_d}(\alpha)}^{\mathsf{G}_d}\mathbb{R}e_\alpha\right),\end{aligned}$$

where the first equality follows by (2.4), the second by the definition of a permutation representation (Definition 2.2.14), and the third equality follows by (Ind), and the last equality follows because

$$\mathsf{G}_d/\mathsf{Stab}_{\mathsf{G}_d}(\alpha) \cong (\mathsf{G}_d\mathsf{H}_{n,d})/(\mathsf{Stab}_{\mathsf{G}_d}(\alpha)\mathsf{H}_{n,d}).$$

Thus, if $\alpha\in\hat{\mathcal{A}}$ has degree d_α and we define $\mathbb{W}_\alpha = \mathrm{Ind}_{\mathsf{Stab}_{\mathsf{G}_{d_\alpha}}(\alpha)}^{\mathsf{G}_{d_\alpha}}\mathbb{R}e_\alpha$, then the map $\bigoplus_{\alpha\in\hat{\mathcal{A}}}\mathrm{Ind}_{\mathsf{G}_{d_\alpha}}(\mathbb{W}_\alpha)_n \to \mathbb{U}_n$ sending e_α to itself is an isomorphism for all $n\geq d+d_0$. Condition 2.3 shows that $\hat{\mathcal{A}}\cap\mathcal{A}_j$ is a set of minimal degree G_j-orbit representatives for each $j\leq d$ if $\mathsf{G}_n=\mathsf{B}_n,\mathsf{S}_n$, hence the above map is an isomorphism for all n. □

We remark that proposition 2.2.15(c) applies to any $\mathscr{V}$ satisfying (2.3) and (2.4).

Uniform Representation Stability

Proposition 2.2.3 shows that $\dim\mathbb{V}_n^{\mathsf{G}_n}$ stabilizes whenever $\{\mathbb{V}_n\}$ is a finitely-generated consistent sequence of $\{\mathsf{G}_n\}$-representations. In fact, the theory of [46, 197, 227] and others shows that for many standard families $\{\mathsf{G}_n\}$ of groups, the entire decomposition of $\mathbb{V}_n$ into irreducibles stabilizes. This phenomenon was called *uniform representation stability* in [47], and we use it to derive constant-sized descriptions for many sequences of PSD cones in Section 2.5. The following is a concrete instance of this phenomenon that we shall use there.

Theorem 2.2.16 ([46, Thm. 1.13],[227, Thm. 4.27])**.** *Let $\mathscr{V}_0=\{\mathbb{R}^n\}$ with $\mathsf{G}_n=\mathsf{B}_n,\mathsf{D}_n$, or S_n be the consistent sequence from Example 2.1.3 and let $\mathscr{V}=\{\mathbb{V}_n\}$ be a $\mathscr{V}_0$-module generated in degree d and presented in degree k. Then there exists a finite set Λ and integers $(m_\lambda)_{\lambda\in\Lambda}$, together with an assignment $\lambda\mapsto\mathbb{W}_{\lambda[n]}$ of a distinct G_n-irreducible $\mathbb{W}_{\lambda[n]}$ to each $\lambda\in\Lambda$ for $n\geq k+d$ such that $\mathbb{V}_n\cong\bigoplus_{\lambda\in\Lambda}\mathbb{W}_{\lambda[n]}^{m_\lambda}$ as G_n-representations.*

More precisely, there is a standard indexing for irreducibles of $\mathsf{S}_n,\mathsf{B}_n$, and D_n in terms of partitions of n, and a particular way to pad these partitions to form $\lambda[n]$, see [227, §2].

Proof. The irreducibles of the groups $\mathsf{S}_n, \mathsf{D}_n, \mathsf{B}_n$ are indexed as in [227, §2.1], and the consistent labelling of irreducibles for different n is given in [227, §2.2]. Under this labelling, the $\mathscr{V}_0$-module $\mathscr{V}$ is uniformly representation stable with stable range $n \geq k+d$ by [227, Thms. 4.4, 4.27], which precisely says that the set of irreducibles appearing in the decomposition of the $\mathbb{V}_n$ and their multiplicities become constant for $n \geq k+d$ by [47, Def. 2.6]. $\square$

Example 2.2.17. *Irreducibles of* S_n *are indexed by partitions of* n*. If* $\lambda_1[n] = (n)$ *is the trivial partition and* $\lambda_2[n] = (n-1, 1)$*, then* $\mathbb{R}^n = \mathbb{W}_{\lambda_1[n]} \oplus \mathbb{W}_{\lambda_2[n]}$ *for all* $n \geq 1$*, where* $\mathbb{W}_{\lambda_1[n]} = \mathrm{span}\{\mathbb{1}_n\}$ *and* $\mathbb{W}_{\lambda_2[n]} = \{x \in \mathbb{R}^n : \mathbb{1}_n^\top x = 0\}$ *are distinct irreducible representations of* S_n*.*

Stabilization of Shifted Sequences

Many of the phenomena in the representation stability literature, including Theorem 2.2.16, can be derived from properties of *shifted* consistent sequences, which are sequences with group actions restricted to centralizing subgroups, see [85] and references therein. In particular, we shall need the following result for such shifted sequences to derive constant-sized descriptions for relative entropy cones in Section 2.5.

Proposition 2.2.18 ([227, Lemma 4.19])**.** *Let* $\mathscr{V} = \{\mathbb{R}^n\}$ *and* $\mathsf{G}_n = \mathsf{B}_n, \mathsf{D}_n$*, or* S_n *as in Example 2.1.3 and let* $\mathscr{U}$ *be a* $\mathscr{V}$*-module presented in degree* k*. If* $\{\mathsf{H}_{n,d}\}$ *are the centralizing subgroups of* $\mathscr{V}$ *and* $\ell \in \mathbb{N}$*, then the projections* $\mathcal{P}_{\mathbb{U}_n} : \mathbb{U}_{n+1}^{\mathsf{H}_{n+1,\ell}} \to \mathbb{U}_n^{\mathsf{H}_{n,\ell}}$ *are isomorphisms for all* $n \geq \ell + k$*.*

Corollary 2.2.19. *Let* $\mathscr{V} = \{\mathbb{R}^n\}$ *and* $\mathsf{G}_n = \mathsf{B}_n, \mathsf{D}_n$*, or* S_n *as in Example 2.1.3, and let* $\mathscr{U}$ *be a* $\mathscr{V}$*-module presented in degree* k*. If* $\beta \in \mathbb{R}^d \setminus \mathbb{R}^{d-1}$ *has minimal degree in its* G_d*-orbit, then the projections* $\mathcal{P}_{\mathbb{U}_n} : \mathbb{U}_{n+1}^{\mathsf{Stab}_{\mathsf{G}_{n+1}}(\beta)} \to \mathbb{U}_n^{\mathsf{Stab}_{\mathsf{G}_n}(\beta)}$ *are isomorphisms for all* $n \geq d + k$*.*

Proof. As shown in (2.4), we have $\mathsf{Stab}_{\mathsf{G}_n}(\beta) = \mathsf{Stab}_{\mathsf{G}_d}(\beta)\mathsf{H}_{n,d}$ for all $n \geq d$, hence

$$\mathbb{U}_n^{\mathsf{Stab}_{\mathsf{G}_n}(\beta)} = \mathbb{U}_n^{\mathsf{H}_{n,d}} \cap \mathbb{U}_n^{\mathsf{Stab}_{\mathsf{G}_d}(\beta)}.$$

By Proposition 2.2.18, the projections $\mathcal{P}_{\mathbb{U}_n} : \mathbb{U}_{n+1}^{\mathsf{H}_{n+1,d}} \to \mathbb{U}_n^{\mathsf{H}_{n,d}}$ are isomorphisms for all $n \geq d+k$. It thus suffices to show that if $u \in \mathbb{U}_{n+1}^{\mathsf{H}_{n+1,d}}$ satisfies $\mathcal{P}_{\mathbb{U}_n} u \in \mathbb{U}_n^{\mathsf{Stab}_{\mathsf{G}_d}(\beta)}$, then $u \in \mathbb{U}_{n+1}^{\mathsf{Stab}_{\mathsf{G}_d}(\beta)}$. For any such u and $g \in \mathsf{Stab}_{\mathsf{G}_d}(\beta)$ we have $g \cdot u \in \mathbb{U}_{n+1}^{\mathsf{H}_{n+1,d}}$ because $\mathsf{H}_{n+1,d}$ and $\mathsf{Stab}_{\mathsf{G}_d}(\beta) \subseteq \mathsf{G}_d$ commute for our specific $\mathscr{V}$. As $\mathcal{P}_{\mathbb{U}_n}(u - g \cdot u) = 0$ and $\mathcal{P}_{\mathbb{U}_n}$ is injective on $\mathbb{U}_{n+1}^{\mathsf{H}_{n+1,d}}$, we get $u = g \cdot u$. $\square$

Limits of Consistent Sequences

Lastly, we consider limits of consistent sequences. We shall use the following definitions and results to describe limits of freely-described sequences of convex sets in Section 2.3. We define limits of consistent sequences, and interpret our definitions in terms of these limits.

Definition 2.2.20. *For a consistent sequence $\mathscr{V} = \{\mathbb{V}_n\}$ of $\{\mathsf{G}_n\}$-representations, define its* limiting representation *as the vector space $\mathbb{V}_\infty = \bigcup_n \mathbb{V}_n$, viewed as a representation of $\mathsf{G}_\infty = \bigcup_n \mathsf{G}_n$.*

There is an approach to representation stability studying limiting representations of limiting groups as above, instead of representations of categories as in Appendix A. For example, the authors of [197] analyze representations of five standard infinite groups, including $\mathsf{O}_\infty, \mathsf{S}_\infty$, that occur as quotients or subrepresentations of tensor powers of $\mathbb{R}^\infty$ and its dual.

We remark that freely-described elements and morphisms of sequences can equivalently be defined in terms of limits. Indeed, given two consistent sequences $\{\mathbb{V}_n\}, \{\mathbb{U}_n\}$ of $\{\mathsf{G}_n\}$-representations, a sequence of equivariant linear maps $\{A_n \in \mathcal{L}(\mathbb{V}_n, \mathbb{U}_n)^{\mathsf{G}_n}\}$ is a morphism of sequences if and only if there exists $A_\infty \in \mathcal{L}(\mathbb{V}_\infty, \mathbb{U}_\infty)^{\mathsf{G}_\infty}$ satisfying $A_\infty|_{\mathbb{V}_n} = A_n$ for all n, i.e., iff $\{A_n\}$ extends to the limit. Similarly, a sequence of invariants $\{v_n \in \mathbb{V}_n^{\mathsf{G}_n}\}$ defines a sequence of invariant linear functionals $\ell_n(x) = \langle v_n, x\rangle \colon \mathbb{V}_n \to \mathbb{R}$, and these extend to a G_∞-invariant linear functional on $\mathbb{V}_\infty$ if and only if $\{v_n\}$ is a freely-described element. Every invariant functional on $\mathbb{V}_\infty$ arises in this way, so freely-described elements are in one-to-one correspondence with invariant linear functionals on the limit of a consistent sequence.

We also consider continuous limits of consistent sequences by taking the completion of $\mathbb{V}_\infty$ with respect to some norm. It is then natural to consider sequences of linear maps which extend to *continuous* maps between these limits. The inner products on each $\mathbb{V}_n$ extend to the limit $\mathbb{V}_\infty$, so we can always complete with respect to the induced norm and obtain a Hilbert space. For many of the examples we consider, however, we obtain more meaningful completions with respect to other norms. For the purpose of obtaining limiting descriptions of convex sets in this completion, we consider the following class of norms.

Definition 2.2.21 (Continuous limits)**.** *Let $\mathscr{V} = \{\mathbb{V}_n\}$ be a consistent sequence of $\{\mathsf{G}_n\}$-representations, let $\mathbb{V}_\infty = \bigcup_n \mathbb{V}_n$, and let $\mathcal{P}_n \colon \mathbb{V}_\infty \to \mathbb{V}_n$ be the orthogonal projection. Fix a norm $\|\cdot\|_c$ on $\mathbb{V}_\infty$ (not necessarily induced by the inner product)* *that satisfies* $\|\mathcal{P}_n x\|_c \leq C\|x\|_c$ *for some* $C > 0$ *and all* n *and* $x \in \mathbb{V}_\infty$*. We call the completion* $\overline{\mathbb{V}_\infty}$ *of* $\mathbb{V}_\infty$ *with respect to* $\|\cdot\|_c$ *a* continuous limit *of the sequence* $\mathscr{V}$*.*

Let $\{\mathbb{V}_n\}$ *and* $\{\mathbb{U}_n\}$ *be consistent sequences of* $\{\mathsf{G}_n\}$*-representations with associated continuous limits* $\overline{\mathbb{V}_\infty}$ *and* $\overline{\mathbb{U}_\infty}$*. A sequence of maps* $\{A_n : \mathbb{V}_n \to \mathbb{U}_n\}$ extends continuously to the limit *if there exists a bounded linear operator* $\overline{A_\infty} : \overline{\mathbb{V}_\infty} \to \overline{\mathbb{U}_\infty}$ *such that* $\overline{A_\infty}|_{\mathbb{V}_n} = A_n$ *for all* n*. A freely-described element* $\{u_n \in \mathbb{U}_n^{\mathsf{G}_n}\}$ extends continuously to the limit *if there exists* $u_\infty \in \overline{\mathbb{U}_\infty}$ *satisfying* $\mathcal{P}_n u_\infty = u_n$ *for all* n*.*

Note that a morphism of sequences $\{A_n : \mathbb{V}_n \to \mathbb{U}_n\}$ extends to the continuous limit if and only if the sequence of operator norms $\{\|A_n\|_{\mathrm{op},c}\}$ with respect to $\|\cdot\|_c$ is bounded.

Our condition on the norm $\|\cdot\|_c$ ensures that each projection $\mathcal{P}_n$ extends to a bounded linear map on the continuous limit, and that the $\mathcal{P}_n$ converge to the identity in the strong operator topology, as we show in the following lemma.

Lemma 2.2.22. *In the notation of Definition 2.2.21, there exists* $C > 0$ *such that* $\|\mathcal{P}_n x\|_c \leq C\|x\|_c$ *for all* n *and all* $x \in \overline{\mathbb{V}_\infty}$ *if and only if* $\mathcal{P}_n$ *extends to a bounded linear map* $\overline{\mathbb{V}_\infty} \to \mathbb{V}_n$ *for all* n *and* $\|x - \mathcal{P}_n x\|_c \to 0$ *as* $n \to \infty$ *for all* $x \in \overline{\mathbb{V}_\infty}$*.*

Proof. Suppose $\|\mathcal{P}_n x\|_c \leq C\|x\|_c$ for all n and x. Then $\|\mathcal{P}_n\|_{\mathrm{op},c} \leq C$ so that $\mathcal{P}_n$ extends to a bounded linear map on $\overline{\mathbb{V}_\infty}$ for all n. Furthermore, for each $x \in \overline{\mathbb{V}_\infty}$ there exists $x_n \in \mathbb{V}_n$ such that $\|x - x_n\|_c \to 0$. Since $x_n = \mathcal{P}_n x_n$, we have

$$\|x - \mathcal{P}_n x\|_c \leq \|x - x_n\|_c + \|\mathcal{P}_n x_n - \mathcal{P}_n x\|_c \leq (1 + C)\|x - x_n\|_c \to 0.$$

Conversely, if $\|\mathcal{P}_n\|_{\mathrm{op},c} < \infty$ for all n and $\mathcal{P}_n$ converge strongly to the identity, then for any $x \in \overline{\mathbb{V}_\infty}$ we have $\|\mathcal{P}_n x\|_c < \infty$ for all n and $\|\mathcal{P}_n x\|_c \to \|x\|_c$ as $n \to \infty$, hence $\sup_n \|\mathcal{P}_n x\|_c < \infty$. By the uniform boundedness principle, we have $\sup_n \|\mathcal{P}_n\|_{\mathrm{op},c} = C < \infty$. □

We close this section by noting that, more generally, if a sequence of convex functions $\{f_n : \mathbb{V}_n \to \mathbb{R} \cup \{\infty\}\}$ is intersection-compatible, then it extends to $f_\infty : \mathbb{V}_\infty \to \mathbb{R} \cup \{\infty\}$. By taking the closure of its epigraph it can further be extended to $\overline{f_\infty} : \overline{\mathbb{V}_\infty} \to \mathbb{R} \cup \{\infty\}$, in which case $\overline{\gamma_{C_\infty}} = \gamma_{\overline{C_\infty}}$ and $\overline{h_{C_\infty}} = h_{\overline{C_\infty}}$. Thus, the conic descriptions we obtain in Theorem 2.3.6 below for continuous limits of convex sets yield conic descriptions for limits of functions as well.

2.3 Structural Results on Dimension-Free Descriptions

With the relevant background from representation stability in hand, we now proceed to prove the structural results discussed in Section 2.1 pertaining to freely-described convex sets. We leverage these results to provide a range of examples of dimension-free descriptions arising in various applications in Section 2.4. We begin with the following instructive example, which shows that freely-described sequences of convex sets need not satisfy either of our compatibility conditions. It also shows that a fixed convex set can extend to different freely-described sequences, depending on its conic description.

Example 2.3.1 ((In)compatible sequence of hypercubes)**.** *In this example, we construct two freely-described sequences of hypercubes with respect to the same description spaces, one of which is both intersection and projection compatible and the other is neither. Let $\mathscr{V} = \{\mathbb{R}^n\}$ with embeddings by zero-padding and the action of $\mathsf{G}_n = \mathsf{B}_n$ as in Example 2.1.3. Let $s = \mathrm{diag}(-1, 1, \dots, 1) \in \mathsf{B}_n$, which together with S_n generates B_n.*

We define two consistent sequences that will be used to construct our description spaces. Let $\mathscr{W}^{(1)} = \{\mathbb{R}^n\}$ with embeddings by zero-padding but on which s acts trivially and S_n acts as usual. Let $\mathscr{W}^{(2)} = \{\mathbb{R}^n \oplus \mathbb{R}^n\}$ with embeddings by zero-padding each of the two summands, on which B_n acts as follows. We set $s \cdot (x, y) = ([y_1, x_2, \dots, x_n]^\top, [x_1, y_2, \dots, y_n]^\top)$ and $\sigma \cdot (x, y) = (\sigma x, \sigma y)$ for $\sigma \in \mathsf{S}_n$. We set our description spaces to

$$\mathscr{U} = \mathscr{W}^{(1)} \oplus \mathscr{W}^{(2)} \oplus \mathbb{R} = \{(\mathbb{R}^n)^{\oplus 3} \oplus \mathbb{R}\}, \quad \mathscr{K} = \{0 \oplus (\mathbb{R}^n_+)^{\oplus 2} \oplus \mathbb{R}_+\},$$
$$\mathscr{W} = \mathscr{W}^{(1)} \oplus \mathbb{R} = \{\mathbb{R}^n \oplus \mathbb{R}\}.$$

Consider the following two sequences of convex sets, both freely-described with respect to these description spaces. The first is the sequence of unit hypercubes $\{C_n^{(1)} = [-1, 1]^n\}$ given by (ConicSeq) with

$$A_n^{(1)} x = (0, x, -x, 0), \qquad B_n^{(1)}(y, \beta) = (y, y, y, 0), \qquad u_n^{(1)} = (-\mathbb{1}_n, 0, 0, 0), \tag{2.5}$$

where $x, y \in \mathbb{R}^n$ and $\beta \in \mathbb{R}$. The second is the sequence of scaled hypercubes $\{C_n^{(2)} = \frac{3}{2n-1}[-1, 1]^n\}$ given by (ConicSeq) with

$$A_n^{(2)} = A_n^{(1)}, \qquad B_n^{(2)}(y, \beta) = (y - \mathbb{1}_n \mathbb{1}_n^\top y + \beta \mathbb{1}_n, y, y, -\mathbb{1}_n^\top y - \beta), \qquad u_n = (0, 0, 0, 3). \tag{2.6}$$

It is straightforward to check that both sequences are freely-described. In particular, all the relevant spaces of invariants stabilize starting at $n = 2$, so that the extension *of either description of* $[-1,1]^2$ *to a freely-described sequence is unique. However, the two sequences of sets have notably different properties. The first sequence is both intersection and projection compatible while the second is neither. Furthermore, note that* $C_2^{(1)} = C_2^{(2)} = [-1,1]^2$ *but* $C_n^{(1)} \neq C_n^{(2)}$ *for all* $n \neq 2$*. Thus, the extension of the unit square* $[-1,1]^2$ *to a freely-described sequence is not unique, but depends on its conic description.*

Motivated by this example, we give conditions on dimension-free descriptions ensuring that the corresponding sequence of sets satisfies compatibility conditions. In turn, these yield conditions on a description of a fixed set ensuring that it extends to a freely-described and compatible sequence of sets as well as conditions on the existence of limiting descriptions.

Dimension-Free Descriptions Certifying Compatibility

Example 2.3.1 shows that a freely-described sequence of convex sets need not satisfy either compatibility condition in Definition 2.1.6. In this section, we therefore give conditions on dimension-free descriptions ensuring that our compatibility conditions are satisfied.

Proposition 2.3.2. *Let* $\{\mathbb{V}_n\}, \{\mathbb{W}_n\}, \{\mathbb{U}_n\}$ *be consistent sequences of* $\{\mathsf{G}_n\}$*-representations, let* $\{\mathcal{K}_n \subseteq \mathbb{U}_n\}$ *be an intersection- and projection-compatible sequence of convex cones, and let* $\mathscr{C} = \{C_n \subseteq \mathbb{V}_n\}$ *be described by linear maps* $\{A_n\colon \mathbb{V}_n \to \mathbb{U}_n\}, \{B_n\colon \mathbb{W}_n \to \mathbb{U}_n\}$ *and elements* $\{u_n \in \mathbb{U}_n^{\mathsf{G}_n}\}$ *as in (ConicSeq).*

- *(a) If* $\{A_n\}, \{B_n\}, \{B_n^*\}$ *are morphisms,* $\{u_n\}$ *is freely-described, and* $u_{n+1} - u_n \in \mathcal{K}_{n+1}$ *for all n, then* $\mathscr{C}$ *is intersection-compatible. If, in addition,* $\{A_n^*\}$ *is a morphism, then* $\mathscr{C}$ *is also projection-compatible.*
- *(b) If* $\{A_n\}, \{A_n^*\}\{B_n\}, \{B_n^*\}$ *are morphisms,* $\{u_n\}$ *is freely-described, and* $u_{n+1} - u_n \in \mathcal{K}_{n+1} + A_{n+1}(\mathbb{V}_n^{\perp}) + B_{n+1}(\mathbb{W}_n)$*, then* $\mathscr{C}$ *is projection-compatible.*

Proof. (a) We show $C_n \subseteq C_{n+1}$. If $x \in C_n$ then there is $y \in \mathbb{W}_n$ satisfying $A_n x + B_n y + u_n \in \mathcal{K}_n$. Then

$$A_{n+1}x + B_{n+1}y + u_{n+1} = A_n x + B_n y + u_n + (u_{n+1} - u_n) \in \mathcal{K}_{n+1},$$

where we used the facts that $\{A_n\}, \{B_n\}$ are morphisms, that $\mathcal{K}_n \subseteq \mathcal{K}_{n+1}$ by intersection-compatiblity, and that $u_{n+1} - u_n \in \mathcal{K}_{n+1}$. Thus, $x \in C_{n+1}$.

Next, we show $C_{n+1} \cap \mathbb{V}_n \subseteq C_n$. If $x \in C_{n+1} \cap \mathbb{V}_n$, there exists $y \in \mathbb{W}_{n+1}$ satisfying $A_{n+1}x + B_{n+1}y + u_{n+1} \in \mathcal{K}_{n+1}$. Because $\{A_n\}$ is a morphism, we have $A_{n+1}x = A_n x$ and hence $\mathcal{P}_{\mathbb{U}_n} A_{n+1}x = A_n x$. Because $\{B_n^*\}$ is a morphism, we have $\mathcal{P}_{\mathbb{U}_n} B_{n+1} = B_n \mathcal{P}_{\mathbb{W}_n}$, hence $\mathcal{P}_{\mathbb{U}_n} B_{n+1} y = B_{n+1}(\mathcal{P}_{\mathbb{W}_n} y)$. Finally, we have $\mathcal{P}_{\mathbb{U}_n} u_{n+1} = u_n$ because $\{u_n\}$ is freely-described, and $\mathcal{P}_{\mathbb{U}_n} \mathcal{K}_{n+1} \subseteq \mathcal{K}_n$ because $\{\mathcal{K}_n\}$ is projection-compatible. Thus, applying $\mathcal{P}_{\mathbb{U}_n}$ we obtain $A_n x + B_n(\mathcal{P}_{\mathbb{W}_n} y) + u_n \in \mathcal{K}_n$, thus yielding that $x \in C_n$. We conclude that $\mathscr{C}$ is intersection-compatible.

Because $\mathscr{C}$ is intersection-compatible, we have $\mathcal{P}_{\mathbb{V}_n} C_{n+1} \supseteq C_n$. Conversely, if $x \in C_{n+1}$ then $A_{n+1}x + B_{n+1}y + u_{n+1} \in \mathcal{K}_{n+1}$ for some $y \in \mathbb{W}_{n+1}$. If $\{A_n^*\}$ is a morphism, then $\mathcal{P}_{\mathbb{U}_n} A_{n+1} = A_n \mathcal{P}_{\mathbb{V}_n}$. Applying $\mathcal{P}_{\mathbb{U}_n}$ to both sides we obtain $A_n \mathcal{P}_{\mathbb{V}_n} x + B_n \mathcal{P}_{\mathbb{W}_n} y + u_n \in \mathcal{K}_n$ and hence $\mathcal{P}_{\mathbb{V}_n} x \in C_n$, which implies that $\mathscr{C}$ is projection-compatible.

(b) First, we show $\mathcal{P}_{\mathbb{V}_n} C_{n+1} \subseteq C_n$. If $x \in C_{n+1}$ then there is $y \in \mathbb{W}_{n+1}$ satisfying $A_{n+1}x + B_{n+1}y + u_{n+1} \in \mathcal{K}_{n+1}$. Applying $\mathcal{P}_{\mathbb{U}_n}$ to both sides and using the facts that $\{A_n^*\}, \{B_n^*\}$ are morphisms, that $\{u_n\}$ is freely-described, and that $\{\mathcal{K}_n\}$ is projection-compatible, we obtain $A_n(\mathcal{P}_{\mathbb{V}_n} x) + B_n(\mathcal{P}_{\mathbb{W}_n} y) + u_n \in \mathcal{K}_n$, showing that $\mathcal{P}_{\mathbb{V}_n} x \in C_n$. Second, we show $C_n \subseteq \mathcal{P}_{\mathbb{V}_n} C_{n+1}$. Suppose $A_n x + B_n y + u_n \in \mathcal{K}_n$ for $x \in \mathbb{V}_n$. Let $x_\perp \in \mathbb{V}_n^\perp$ and $y' \in \mathbb{W}_n$ satisfy $u_{n+1} - u_n + A_{n+1}x_\perp + B_{n+1}y' \in \mathcal{K}_{n+1}$. As $\{A_n\}, \{B_n\}$ are morphisms,

$$A_{n+1}(x+x_\perp)+B_{n+1}(y+y')+u_{n+1} = A_n x+B_n y+u_n+(u_{n+1}-u_n+A_{n+1}x_\perp+B_{n+1}y') \in \mathcal{K}_{n+1},$$

hence $x + x_\perp \in C_{n+1}$ and $\mathcal{P}_{\mathbb{V}_n}(x + x_\perp) = x$. This shows $\mathscr{C}$ is projection-compatible. $\square$

We say that a sequence of conic descriptions *certifies compatibility* when it satisfies the hypotheses of Proposition 2.3.2.

Remark 2.3.3. *We make a number of remarks about the conditions in Proposition 2.3.2.*

- *Standard sequences of cones such as nonnegative orthants and PSD cones satisfy both intersection and projection compatibility.*
- *The set of linear maps $\{A_n\}$ and $\{B_n\}$ satisfying the hypotheses of Proposition 2.3.2(a) form* linear subspaces *of the corresponding spaces of freely-described elements. The set of sequences $\{u_n\}$ satisfying those hypotheses*

form a convex cone. *Similarly, we can parametrize descriptions satisfying the hypotheses of Proposition 2.3.2(b) by a convex cone, by considering $\{u_n\}$ of the form $u_n = A_n(v_n) + B_n(w_n) + z_n$ where $\{v_n \in \mathbb{V}_n^{\mathsf{G}_n}\}$, $\{w_n \in \mathbb{W}_n^{\mathsf{G}_n}\}$, and $\{z_n \in \mathcal{K}_n^{\mathsf{G}_n}\}$ are freely-described.*

- *Freely-described sets satisfying the hypotheses of Proposition 2.3.2(b) need not be intersection-compatible, as the elliptope in (2.17) studied below demonstrates.*

- *Dimension-free descriptions certifying compatibility often extend to descriptions of infinite-dimensional limits, see Theorem 2.3.6 below.*

Returning to Example 2.3.1, we note that the description (2.5) satisfies all the hypotheses of Proposition 2.3.2(a) and hence certifies the intersection and projection compatibility of the sequence of hypercubes it describes. On the other hand, the description (2.6) does not satisfy the hypotheses of either part of the above theorem because neither $\{B_n^{(2)}\}$ nor $\{(B_n^{(2)})^*\}$ are morphisms.

Extending a Convex Set to a Freely-Described Sequence

Let $\mathscr{V} = \{\mathbb{V}_n\}$ be a consistent sequence of $\{\mathsf{G}_n\}$-representations. In this section, we fix n_0 and seek to extend a convex subset $C_{n_0} \subseteq \mathbb{V}_{n_0}$ to a freely-described and compatible sequence. As we saw in Example 2.3.1, different extensions may be obtained from different fixed-dimensional conic descriptions of C_{n_0}. We therefore further fix consistent sequences $\mathscr{U} = \{\mathbb{U}_n\}, \mathscr{W} = \{\mathbb{W}_n\}$ and cones $\{\mathcal{K}_n \subseteq \mathbb{U}_n\}$ satisfying both intersection and projection compatibility, and assume C_{n_0} is described by (ConicSeq). We now ask when the description of C_{n_0} can be extended to a dimension-free description that yields a compatible sequence of convex sets.

If n_0 exceeds the presentation degrees of $\mathscr{W}, \mathscr{U}, \mathscr{V} \otimes \mathscr{U}$ and $\mathscr{W} \otimes \mathscr{U}$, then Proposition 2.2.9 shows that we can uniquely extend all the elements in the description of C_{n_0} to freely-described elements, and hence extend C_{n_0} to a freely-described sequence. To ensure that this unique extension satisfies our compatibility conditions, we give conditions on these invariants guaranteeing that their extensions to freely-described elements satisfy the conditions of Proposition 2.3.2.

We begin by characterizing when a fixed equivariant map $A_{n_0} \in \mathcal{L}(\mathbb{V}_{n_0}, \mathbb{U}_{n_0})^{\mathsf{G}_{n_0}}$ extends to a morphism of sequences $\{A_n \in \mathcal{L}(\mathbb{V}_n, \mathbb{U}_n)^{\mathsf{G}_n}\}$.

Theorem 2.3.4. *Let $\mathscr{V}_0$ be a consistent sequence of $\{\mathsf{G}_n\}$-representations and let $\mathscr{V} = \{\mathbb{V}_n\}, \mathscr{U} = \{\mathbb{U}_n\}$ be $\mathscr{V}_0$-modules. Assume $\mathscr{V}$ is generated in degree d and* *presented in degree k, and fix $n_0 \geq k$. Then A_{n_0} extends to a morphism of sequences if and only if $A_{n_0} \in \mathcal{L}(\mathbb{V}_{n_0}, \mathbb{U}_{n_0})^{\mathsf{G}_{n_0}}$ satisfies $A_{n_0}(\mathbb{V}_j) \subseteq \mathbb{U}_j$ for $j \leq d$.*

Proof. If A_{n_0} extends to a morphism then $A_{n_0}(\mathbb{V}_j) = A_j(\mathbb{V}_j) \subseteq \mathbb{U}_j$ for all $j \leq n_0$. Conversely, assume $A_{n_0}(\mathbb{V}_j) \subseteq \mathbb{U}_j$ for $j \leq d$, and let $\{\mathsf{H}_{n,d}\}$ be the centralizing subgroups of $\mathscr{V}_0$. Suppose first that $\mathscr{V} = \mathscr{F} = \bigoplus_j \mathrm{Ind}_{\mathsf{G}_{d_j}} \mathbb{W}_{d_j}$ is free. Note that it is generated in degree $\max_j d_j \leq d$.

Let $A_{d_j} = A_{n_0}|_{\mathbb{W}_{d_j}}$ and fix $n \geq d_j$. Because $\mathscr{U}$ is a $\mathscr{V}_0$-module, we have $\mathbb{U}_{d_j} \subseteq \mathbb{U}_n^{\mathsf{H}_{n,d_j}}$, so we can view $\mathbb{U}_{d_j}$ as a representation of $\mathsf{G}_{d_j}\mathsf{H}_{n,d_j}$ on which H_{n,d_j} acts trivially. As $A_{d_j}(\mathbb{W}_{d_j}) \subseteq \mathbb{U}_{d_j}$ and is $\mathsf{G}_{d_j}\mathsf{H}_{n,d_j}$-equivariant, the following composition defines an equivariant map

$$A_{n,j} \colon \mathrm{Ind}_{\mathsf{G}_{d_j}\mathsf{H}_{n,d_j}}^{\mathsf{G}_n}(\mathbb{W}_{d_j}) \xrightarrow{\mathrm{Ind}(A_{d_j})} \mathrm{Ind}_{\mathsf{G}_{d_j}\mathsf{H}_{n,d_j}}^{\mathsf{G}_n} \mathbb{U}_{d_j} \xrightarrow{g\otimes u \mapsto g\cdot u} \mathbb{U}_n,$$

where the induced map $\mathrm{Ind}(A_{d_j})$ was defined in Section 2.1. Note that $A_{n_0,j} = A_{n_0}|_{\mathrm{Ind}_{\mathsf{G}_{d_j}}(\mathbb{W}_{d_j})_{n_0}}$, since $A_{n_0,j}(g \otimes w) = g \cdot A_{n_0} w$ for all $g \in \mathsf{G}_n$ and $w \in \mathbb{W}_{d_j}$. Also, $\{A_{n,j}\}$ defines a morphism $\mathrm{Ind}_{\mathsf{G}_{d_j}}(\mathbb{W}_{d_j}) \to \mathscr{U}$. Therefore, the desired extension of A_{n_0} to a morphism $\{A_n\}$ is given by $A_n = \bigoplus_j A_{n,j} \colon \mathbb{V}_n \to \mathbb{U}_n$.

Now suppose $\mathscr{F} = \{\mathbb{F}_n\}$ is an algebraically free $\mathscr{V}$-module as above with a surjection $\mathscr{F} \to \mathscr{V}$ whose kernel $\mathscr{K} = \{\mathbb{K}_n\}$ is generated in degree k. Define the composition

$$\widetilde{A}_{n_0} \colon \mathbb{F}_{n_0} \to \mathbb{V}_{n_0} \xrightarrow{A_{n_0}} \mathbb{U}_{n_0},$$

which satisfies $\widetilde{A}_{n_0}(\mathbb{F}_j) \subseteq \mathbb{U}_j$ for all $j \leq d$ by assumption and $\widetilde{A}_{n_0}(\mathbb{K}_{n_0}) = 0$ by its definition. By the previous paragraph, it extends to a morphism $\{\widetilde{A}_n \colon \mathbb{F}_n \to \mathbb{U}_n\}$. Because $\mathscr{K}$ is generated in degree k and $n_0 \geq k$, we have $\mathbb{K}_n = \mathbb{R}[\mathsf{G}_n]\mathbb{K}_{n_0}$. Because $\widetilde{A}_n$ is equivariant, we have $\widetilde{A}_n(\mathbb{K}_n) = 0$. Therefore, $\widetilde{A}_n$ can be factored as $\mathbb{F}_n \to \mathbb{F}_n/\mathbb{K}_n = \mathbb{V}_n \xrightarrow{A_n} \mathbb{U}_n$, where the maps A_n in this factorization give the desired extension of A_{n_0} to a morphism $\mathscr{V} \to \mathscr{U}$. $\square$

To satisfy the conditions in Proposition 2.3.2, we also use Theorem 2.3.4 to ensure $\{A_n^*\}$ defines a morphism. To that end, note that $A_{n_0}^*(\mathbb{U}_j) \subseteq \mathbb{V}_j$ if and only if $A_{n_0}(\mathbb{V}_j^\perp) \subseteq \mathbb{U}_j^\perp$, where orthogonal complements are taken inside $\mathbb{V}_{n_0}$ and $\mathbb{U}_{n_0}$. We can now give conditions guaranteeing extendability of a convex set to a freely-described and compatible sequence.

Theorem 2.3.5 (Parametrizing freely-described and compatible sequences)**.** *Let $\mathscr{V}_0$ be a consistent sequence of $\{\mathsf{G}_n\}$ representations and let $\mathscr{V} = \{\mathbb{V}_n\}$, $\mathscr{W} = \{\mathbb{W}_n\}$,* *and $\mathscr{U} = \{\mathbb{U}_n\}$ be $\mathscr{V}_0$-modules generated in degrees d_V, d_U, d_W, respectively, and presented in degree k. Let $\{\mathcal{K}_n \subseteq \mathbb{U}_n\}$ be an intersection and projection-compatible sequence of convex cones. Fix $n_0 \geq k$.*

Let $u_{n_0} \in \mathbb{U}_{n_0}^{\mathsf{G}_{n_0}}$ and let $A_{n_0} \in \mathcal{L}(\mathbb{V}_{n_0}, \mathbb{U}_{n_0})^{\mathsf{G}_{n_0}}$, $B_{n_0} \in \mathcal{L}(\mathbb{W}_{n_0}, \mathbb{U}_{n_0})^{\mathsf{G}_{n_0}}$ such that

$$A_{n_0}(\mathbb{V}_j) \subseteq \mathbb{U}_j \text{ for } j \leq d_V, \quad B_{n_0}(\mathbb{W}_j) \subseteq \mathbb{U}_j \text{ for } j \leq d_W, \quad B_{n_0}(\mathbb{W}_j^{\perp}) \subseteq \mathbb{U}_j^{\perp} \text{ for } j \leq d_U. \tag{2.7}$$

Then there are unique extensions of A_{n_0} and B_{n_0} to morphisms $\{A_n \in \mathcal{L}(\mathbb{V}_n, \mathbb{U}_n)^{\mathsf{G}_n}\}$ and $\{B_n \in \mathcal{L}(\mathbb{W}_n, \mathbb{U}_n)^{\mathsf{G}_n}\}$, where $\{B_n^*\}$ is a morphism as well. Furthermore, there is a unique extension of u_{n_0} to a freely-described element $\{u_n \in \mathbb{U}_n^{\mathsf{G}_n}\}$. Let $\mathscr{C} = \{C_n\}$ be the freely-described sequence of convex sets given by (ConicSeq).

(a) *If $u_{n+1} - u_n \in \mathcal{K}_{n+1}$ for all n, then $\mathscr{C}$ is intersection-compatible. If, in addition, we have $A_{n_0}(\mathbb{V}_j^{\perp}) \subseteq \mathbb{U}_j^{\perp}$ for $j \leq d_U$, then $\mathscr{C}$ is also projection-compatible.*

(b) *If $u_{n+1} - u_n \in \mathcal{K}_{n+1} + A_{n+1}(\mathbb{V}_n^{\perp}) + B_{n+1}(\mathbb{U}_{n+1})$ for all n, then $\mathscr{C}$ is projection-compatible.*

Proof. Theorem 2.3.4 shows that A_{n_0} and B_{n_0} uniquely extend to morphisms $\{A_n\}, \{B_n\}$ such that $\{B_n^*\}$ is also a morphism. Proposition 2.2.9 shows that u_{n_0} extends to a freely-described element $\{u_n\}$. Under the stated conditions on these extensions, Proposition 2.3.2 yields the claimed compatibility conditions for the sequence of convex sets $\{C_n\}$ given by (ConicSeq). □

In Section 2.6, we use Theorem 2.3.5 to computationally parametrize and search over dimension-free descriptions certifying compatibility.

Limits of Freely-Described Convex Sets

Our last structural result gives conditions under which dimension-free descriptions extend to descriptions of continuous limits of convex sets. The certificates of compatibility in Proposition 2.3.2 play a major role once again in the existence of these limiting descriptions. If $\mathscr{C} = \{C_n \subseteq \mathbb{V}_n\}$ is an intersection-compatible sequence of convex subsets of a consistent sequence $\{\mathbb{V}_n\}$, define $C_\infty = \bigcup_n C_n$, which is a convex subset of $\mathbb{V}_\infty$.

Theorem 2.3.6 (Descriptions of limits)**.** *Suppose $\mathscr{C} = \{C_n \subseteq \mathbb{V}_n\}$ is given by (ConicSeq). If all the hypotheses of Proposition 2.3.2(a) are satisfied (so $\mathscr{C}$ is intersection and* *projection compatible) and if* $\{A_n\}, \{B_n\}, \{u_n\}$ *extend continuously to the limit, then*

$$\left\{x \in \overline{\mathbb{V}_\infty} : \exists y \in \overline{\mathbb{W}_\infty} \text{ s.t. } \overline{A_\infty}x + \overline{B_\infty}y + u_\infty \in \overline{\mathcal{K}_\infty}\right\}, \tag{2.8}$$

contains C_∞ *and is dense in its closure. If* $B_\infty = 0$*, then* (2.8) *equals* $\overline{C_\infty}$.

Proof. To prove that C_∞ is contained in (2.8), observe that $u_\infty - u_n = \lim_{N\to\infty}(u_N - u_n) \in \overline{\mathcal{K}_\infty}$ for all n by Lemma 2.2.22. Therefore, if $x \in C_n$ and $y \in \mathbb{W}_n$ satisfy $A_n x + B_n y + u_n \in \mathcal{K}_n$, then $\overline{A_\infty}x + \overline{B_\infty}y + u_\infty = A_n x + B_n y + u_n + (u_\infty - u_n) \in \overline{\mathcal{K}_\infty}$, proving that x is in (2.8).

To prove that (2.8) is contained in $\overline{C_\infty}$, suppose $x \in \overline{\mathbb{V}_\infty}$ and $y \in \overline{\mathbb{W}_\infty}$ satisfy $\overline{A_\infty}x + \overline{B_\infty}y + u_\infty \in \overline{\mathcal{K}_\infty}$. Because $\{A_n^*\}, \{B_n^*\}$ are morphisms and $\{\mathcal{K}_n\}$ is projection-compatible, applying $\mathcal{P}_{\mathbb{U}_n}$ we obtain $A_n(\mathcal{P}_{\mathbb{V}_n}x) + B_n(\mathcal{P}_{\mathbb{W}_n}y) + u_n \in \mathcal{K}_n$, hence $\mathcal{P}_{\mathbb{V}_n}x \in C_n$ for all n and $x = \lim_n \mathcal{P}_{\mathbb{V}_n}x \in \overline{C_\infty}$ by Lemma 2.2.22.

If $B_\infty = 0$ then (2.8) is the preimage under the continuous map $x \mapsto \overline{A_\infty}x + u_\infty$ of the closed cone $\overline{\mathcal{K}_\infty}$, hence (2.8) is closed and must equal $\overline{C_\infty}$. □

If the set (2.8) is dense in $\overline{C_\infty}$, then optimizing a continuous function over (2.8) and over $\overline{C_\infty}$ are equivalent, yielding a finitely-parametrized conic program in $\overline{\mathbb{V}_\infty}$.

2.4 Examples

In this section, we present examples of freely-described and compatible sequences of sets and functions arising in a variety of applications. For several of these, we derive finitely-parametrized families of freely-described sets, and we also give limiting descriptions and some related consequences.

Simplices and Norms

Let $\mathscr{V} = \{\mathbb{V}_n = \mathbb{R}^n\}$ with embedding by zero-padding and the action of $\mathsf{G}_n = \mathsf{S}_n$ as in Example 2.1.3.

Simplex: The sequence of simplices $\Delta^{n-1} = \{x \in \mathbb{R}^n : x \geq 0,\ \mathbb{1}_n^\top x = 1\}$ is freely-described, as the associated description is of the form (ConicSeq) with description spaces $\mathscr{U} = \mathscr{V} \oplus \mathbb{R} = \{\mathbb{U}_n = \mathbb{R}^{n+1}\}$, $\mathscr{W} = \{\mathbb{W}_n = 0\}$ and cones $\mathscr{K} = \{\mathcal{K}_n = \mathbb{R}^n_+ \oplus \{0\}\}$, and with $A_n = [I_n, -\mathbb{1}_n]^\top$, $B_n = 0$, and $u_n = [0, 1]^\top$. Moreover, the simplices $\{\Delta^{n-1}\}$ are intersection-compatible (but not projection-compatible); the above description satisfies the hypotheses of Proposition 2.3.2(a) and hence certifies this compatibility.

Solid simplex: In contrast, consider the sequence of solid simplices $\Delta_s^n = \{x \in \mathbb{R}^n : x \geq 0, \mathbb{1}_n^\top x \leq 1\}$. This sequence of conic descriptions is the same as above except that here $\mathcal{K}_n = \mathbb{R}_+^{n+1}$, and in particular is dimension-free and certifies intersection compatibility of $\{\Delta_s^n\}$. However, the sequence $\{\Delta_s^n\}$ is also projection-compatible, but the above dimension-free description of it does *not* certify this compatibility. That is because $\{A_n^*\}$ is not a morphism, hence the hypotheses of the second part of Proposition 2.3.2(a) are not satisfied. Indeed, the above description of the solid simplex does not make its projection compatibility apparent, since $\mathbb{1}_n^\top \mathcal{P}_{\mathbb{V}_n} x$ can be arbitrary compared to $\mathbb{1}_{n+1}^\top x$ for general $x \in \mathbb{R}^{n+1}$, but must be smaller if $x \geq 0$.

Instead, the following is a dimension-free *semidefinite* description of the solid simplices that certifies both intersection and projection compatibilities:

$$\Delta_s^n = \left\{ x \in \mathbb{R}^n : \begin{bmatrix} 1 & x^\top \\ x & \mathrm{diag}(x) \end{bmatrix} = A_n x + u_n \succeq 0 \right\}, \tag{2.9}$$

which is of the form (ConicSeq) with $\mathscr{U} = \mathrm{Sym}^2(\mathrm{Sym}^{\leq 1}\mathscr{V}) = \{\mathbb{S}^{n+1}\}$, $\mathscr{K} = \{\mathbb{S}_+^{n+1}\}$, and $\mathscr{W} = 0$. Here both $\{A_n\}$ and $\{A_n^*\}$ are morphisms. It would be interesting to find a dimension-free linear programming description of the solid simplices which certifies both compatibilities, or to show that none exists.

Norm balls: Several related dimension-free descriptions are derived from the above. The sequence of ℓ_1 unit balls $\mathcal{B}_{\ell_1}^n = \{x \in \mathbb{R}^n : \|x\|_1 \leq 1\}$ is intersection and projection compatible. It can be written as $\mathcal{B}_{\ell_1}^n = \{x \in \mathbb{R}^n : \exists y \in \Delta_s^n \text{ s.t. } -y \leq x \leq y\}$, which combined with (2.9) yields a dimension-free description certifying both compatibilities. Similarly, the sequence of ℓ_2 unit balls $\mathcal{B}_{\ell_2}^n = \{x \in \mathbb{R}^n : \|x\|_2 \leq 1\}$ is intersection and projection compatible, and its standard SDP description $\mathcal{B}_{\ell_2}^n = \left\{ x \in \mathbb{R}^n : \begin{bmatrix} 1 & x^\top \\ x & I_n \end{bmatrix} = A_n x + u_n \succeq 0 \right\}$ is dimension-free and certifies these compatibilities similarly to (2.9).

Notice that both sequences of norm balls are invariant under the larger group $\mathsf{G}_n = \mathsf{B}_n$. All of the above descriptions are in fact equivariant under this larger group, when $\mathscr{U}, \mathscr{W}$ are endowed with the corresponding group action.

Regular Polygons

The following example illustrates a natural sequence of convex sets that is freely-described but satisfies neither intersection nor projection compatibility. Let $\mathscr{V} = \{(\mathbb{V}_n, \varphi_n)\}$ be the consistent sequence $\mathbb{V}_n = \mathbb{R}^2$ with $\varphi_n = \mathrm{id}_{\mathbb{R}^2}$ and the standard action of the dihedral group $\mathsf{G}_n = \mathrm{Dih}_{2^n}$. Consider the sequence of regular 2^n-gons $\mathscr{C} = \{C_n \subseteq \mathbb{R}^2\}$ defined by

$$C_n = \mathrm{conv}\left\{\begin{bmatrix}\cos\theta_i \\ \sin\theta_i\end{bmatrix}\right\}, \quad \theta_i = \frac{2\pi i}{2^n}, \quad i \in \{0, \dots, 2^n - 1\}. \tag{2.10}$$

Because $\mathbb{V}_n = \mathbb{V}_{n+1}$ while $C_n \neq C_{n+1}$, the sequence $\mathscr{C}$ satisfies neither intersection nor projection compatibility. Nevertheless, it admits the dimension-free description

$$C_n = \left\{x \in \mathbb{R}^2 : \exists y \in \mathbb{R}^{2^n} \text{ s.t. } \begin{bmatrix}-I \\ 0 \\ 0\end{bmatrix} x + \begin{bmatrix}\begin{bmatrix}\cdots & \cos(2\pi i/2^n) & \cdots \\ & \sin(2\pi i/2^n) & \end{bmatrix} \\ I_{2^n} \\ \mathbb{1}_{2^n}^\top\end{bmatrix} y + \begin{bmatrix}0 \\ 0 \\ -1\end{bmatrix} \in 0 \oplus \mathbb{R}_+^{2^n} \oplus 0\right\},$$

where $\mathscr{W} = \{\mathbb{R}^{2^n}\}_n$ with embeddings $y \mapsto y \otimes [1, 0]^\top$, and $\mathscr{U} = \mathscr{V} \oplus \mathscr{W} \oplus \mathbb{R}$. We put the standard inner products on $\mathbb{R}^{2^n}$. The group permutes the 2^n vertices of C_n, defining a permutation action on $[2^n]$, and it acts on $\mathbb{R}^{2^n}$ by applying these permutations to coordinates.

If $\theta_i = \pi(2i+1)/2^n$ in (2.10) instead, we mention that the semidefinite description of C_n given in [76] is also dimension-free when $\mathscr{U}, \mathscr{W}, \mathscr{K}$ are chosen appropriately.

Permutahedra, Schur–Horn Orbitopes, and Limits

Permutahedra: Consider the sequence of standard permutahedra

$$\mathrm{Perm}_n = \mathrm{conv}\left\{g \cdot [1, 2, \dots, n]^\top\right\} = \left\{M[1, 2, \dots, n]^\top : M \in \mathbb{R}_+^{n\times n},\ M\mathbb{1}_n = M^\top\mathbb{1}_n = \mathbb{1}_n\right\}, \tag{2.11}$$

where the second equality follows by the Birkhoff–von Neumann theorem. The sequence $\{\mathrm{Perm}_n\}$, viewed as subsets of the consistent sequence in Example 2.1.3 with $\mathsf{G}_n = \mathsf{S}_n$, is neither intersection- nor projection-compatible. Furthermore, their description (2.11) is not dimension-free because the map $M \mapsto M[1, 2, \dots, n]^\top$ is not G_n-equivariant. The smaller descriptions of these permutahedra in [88, 143] are also not dimension-free because they are not equivariant. However, there is a sequence of permutahedra arising naturally from a limiting perspective that is both intersection- and projection-compatible and whose description certifies this compatibility.

Fix $q, m \in \mathbb{N}$ and a vector $\lambda \in \mathbb{R}^q$ with distinct entries, and define $\widetilde{\lambda} = (\lambda_1, \dots, \lambda_1, \dots, \lambda_q, \dots, \lambda_q) \in \mathbb{R}^m$ in which λ_i appears m_i times (so $\sum_i m_i = m$). Let $\mathsf{G}_n = \mathsf{S}_{m2^n}$ embedded in G_{n+1} by sending a $m2^n \times m2^n$ matrix $g \in \mathsf{G}_n$ to $g \otimes I_2$. Let $\mathscr{V} = \{\mathbb{R}^{m2^n}\}_n$ with embeddings $x \mapsto x \otimes \mathbb{1}_2$, the normalized inner product $\langle x, y\rangle = (m2^n)^{-1}x^\top y$, and the standard action of G_n. We consider convex hulls of all the vectors in $\mathscr{V}$ containing λ_i in a fraction m_i/m of its entries, which are given using the Birkhoff–von Neumann theorem by

$$\begin{aligned}\mathrm{Perm}(\lambda)_n &= \mathrm{conv}\{g \cdot (\widetilde{\lambda} \otimes \mathbb{1}_{m2^n})\}_{g \in \mathsf{S}_{m2^n}} \\ &= \left\{M\lambda : M \in \mathbb{R}_+^{m2^n \times q},\ M\mathbb{1}_q = \mathbb{1}_{m2^n},\ M^\top \mathbb{1}_{m2^n} = 2^n[m_1, \dots, m_q]^\top\right\}.\end{aligned} \tag{2.12}$$

This is an intersection- and projection-compatible sequence of subsets of $\mathscr{V}$. Moreover, the description (2.12) is dimension-free and certifies intersection and projection compatibility. Indeed, let $\mathscr{W} = \mathscr{V}^{\oplus q} = \{\mathbb{R}^{m2^n \times q}\}_n$ and $\mathscr{U} = \mathscr{W} \oplus \mathscr{V}^{\oplus 2} \oplus \mathbb{R}^q$ containing cones $\{\mathbb{R}_+^{m2^n \times q} \oplus 0 \oplus 0\}$. Then (2.12) is of the form (ConicSeq) with

$$A_n x = (0, -x, 0, 0), \quad B_n M = (M,\ M\lambda,\ M\mathbb{1}_q,\ (m2^n)^{-1}M^\top \mathbb{1}_{m2^n}),$$

and $u_n = (0, 0, -\mathbb{1}_{m2^n}, -\left[\frac{m_1}{m}, \dots, \frac{m_q}{m}\right]^\top)$. Since $\{A_n\}, \{A_n^*\}, \{B_n\}, \{B_n^*\}$ are morphisms and $u_n = u_{n+1}$ under our embedding, Proposition 2.3.2(a) applies (note that the inner product here is nonstandard, so $A_n^* \neq A_n^\top$ and similarly for B_n).

The insight behind this construction comes from the limiting perspective of Section 2.3. Note that $\mathbb{V}_\infty$ can be viewed as a space of piecewise-constant functions on $[0, 1)$. Indeed, define intervals $I_i^{(n)} = [(i-1)/m2^n, i/m2^n)$ and associate to each $v \in \mathbb{V}_n$ the piecewise-constant function $f_v(x) = \sum_{i=1}^{m2^n} v_i \mathbb{1}_{I_i^{(n)}}(x)$ where $\mathbb{1}_S(x) = 1$ if $x \in S$ and 0 otherwise. Note that $v \in \mathbb{V}_n$ and $v \otimes \mathbb{1}_2 \in \mathbb{V}_{n+1}$ define the same function in this way, and that $\langle v, w\rangle = \langle f_v, f_w\rangle_{L^2([0,1))}$. Also, we have for $f \in \mathbb{V}_\infty$

$$(\mathcal{P}_n f)(x) = m2^n \int_{I_i^{(n)}} f \quad \text{if } x \in I_i^{(n)}, \tag{2.13}$$

hence Jensen's inequality implies $\|\mathcal{P}_n f\|_{L^p} \leq \|f\|_{L^p}$ for all $p \in [1, \infty]$. Thus, we can identify $\mathbb{V}_\infty$ with the space of such step functions, which is contained in $L^p([0, 1))$ for all $p \in [1, \infty]$, and we can consider the completion of $\mathbb{V}_\infty$ with respect to any of the L^p norms. If $p < \infty$ we get $\overline{\mathbb{V}_\infty} = L^p([0, 1))$ while if $p = \infty$ then $\overline{\mathbb{V}_\infty}$ is the space of functions continuous on $[0, 1) \setminus \{m/2^n : m, n \in \mathbb{N}\}$ and having left and right limits everywhere [62, §7.6]. Then $\overline{\mathrm{Perm}(\lambda)_\infty}$ is the closed convex hull of functions in $\overline{\mathbb{V}_\infty}$ that take values λ_i on a subset of $[0, 1]$ of measure m_i/m. Furthermore, Theorem 2.3.6 applies to the description in (2.12) and yields the following dense subset of $\overline{\mathrm{Perm}(\lambda)_\infty}$.

Proposition 2.4.1. *Endowing $\mathbb{V}_\infty$ with the $L^p([0,1))$ norm as above for any $p \in [1,\infty]$, we obtain*

$$\begin{aligned}\overline{\mathrm{Perm}(\lambda)_\infty} &= \overline{\mathrm{conv}}\left\{f \in \overline{\mathbb{V}_\infty} : f([0,1)) = \{\lambda_1,\dots,\lambda_q\},\ |f^{-1}(\lambda_i)| = \frac{m_i}{m}\right\} \\ &= \overline{\left\{\sum_{i=1}^q \lambda_i f_i : f_i \in \overline{\mathbb{V}_\infty},\ f_i \geq 0 \ a.e.\ , \sum_{i=1}^q f_i = 1,\ \int_{[0,1)} f_i = \frac{m_i}{m}\right\}}.\end{aligned} \tag{2.14}$$

Proof. For the first equality, the inclusion $\subseteq$ is clear. For the reverse inclusion, if $f \in \overline{\mathbb{V}_\infty}$ takes values λ_i on sets $\Omega_i = f^{-1}(\lambda_i)$ of measure m_i/m partitioning $[0,1)$, we can write $f = \sum_{i=1}^q \lambda_i \mathbb{1}_{\Omega_i}$. Then $\mathcal{P}_n f = \sum_{i=1}^q \lambda_i(\mathcal{P}_n \mathbb{1}_{\Omega_i})$, and under the above identification of $\mathbb{R}^{m2^n}$ with piecewise-constant functions, we can view $\mathcal{P}_n \mathbb{1}_{\Omega_i} \in \mathbb{R}_+^{m2^n}$ as nonnegative vectors. The matrix $M = [\mathcal{P}_n \mathbb{1}_{\Omega_1},\dots,\mathcal{P}_n \mathbb{1}_{\Omega_q}] \in \mathbb{R}_+^{m2^n\times q}$ then satisfies the conditions in (2.12), hence $\mathcal{P}_n f \in \mathrm{Perm}(\lambda)_n$ and $f = \lim_n \mathcal{P}_n f \in \overline{\mathrm{Perm}(\lambda)_\infty}$ by Lemma 2.2.22, giving the reverse inclusion.

The second equality follows from Theorem 2.3.6. Indeed, endow $\mathbb{W}_\infty = \mathbb{V}_\infty^q$ with the norm $\|[f_1,\dots,f_q]\| = \max_i \|f_i\|_p$ and $\mathbb{U}_\infty$ with the norm $([f_1,\dots,f_q], g_1, g_2, \mu) = \max\{\|f_i\|_p, \|g_j\|_p, \|\mu\|_\infty\}$. Then $\overline{\mathbb{W}_\infty} = \overline{\mathbb{V}_\infty}^q$ and $\overline{\mathbb{U}_\infty} = \overline{\mathbb{W}_\infty} \oplus \overline{\mathbb{V}_\infty}^2 \oplus \mathbb{R}^q$, and we have $\|A_n\|_{\mathrm{op}} = 1$, $\|B_n\|_{\mathrm{op}} \leq \max\{\sum_i |\lambda_i|, q\}$, and $u_n = u_{n+1}$ for all n. □

Schur–Horn orbitopes: We consider the matrix analogs of the above permutahedra, which are convex hulls of all matrices with a given spectrum. Let $\mathsf{G}_n = \mathsf{O}_{m2^n}$ embed in G_{n+1} by sending a $m2^n \times m2^n$ matrix g to $g \otimes I_2$, let $\mathbb{V}_n = \mathbb{S}^{m2^n}$ embed in $\mathbb{V}_{n+1}$ by $X \mapsto X \otimes I_2$ with normalized inner product $\langle X, Y\rangle = (m2^n)^{-1}\mathrm{Tr}(X^\top Y)$, and finally let G_n act by conjugation on $\mathbb{V}_n$. Consider the sequence of Schur–Horn orbitopes [35, Eq. (19)]

$$\begin{aligned}\mathrm{SH}(\lambda)_n &= \mathrm{conv}\{g\cdot \mathrm{diag}(\lambda\otimes \mathbb{1}_{2^n})\}_{g\in\mathsf{O}_{m2^n}} \\ &= \left\{\sum_{i=1}^q \lambda_i Y_i : Y_1,\dots,Y_q \in \mathbb{V}_n \text{ s.t. } \sum_{i=1}^q Y_i = I,\ Y_i \succeq 0,\ \mathrm{Tr}(Y_i) = m_i 2^n \text{ for } i = 1,\dots,q\right\},\end{aligned} \tag{2.15}$$

which is the matrix analog of (2.12). This is again a dimension-free description certifying both intersection- and projection-compatibility. Indeed, let $\mathscr{W} = \mathscr{V}^{\oplus q}$ and $\mathscr{U} = \mathscr{W} \oplus \mathscr{V}^{\oplus 2} \oplus \mathbb{R}^q$ containing the cones $\{\mathcal{K}_n = (\mathbb{S}_+^{m2^n})^{\oplus q} \oplus 0 \oplus 0\}$. Then (2.15) is of the form (ConicSeq) with

$$A_n X = (0, -X, 0, 0), \quad B_n[Y_i]_{i=1}^q = \left([Y_i]_{i=1}^q,\ \sum_i \lambda_i Y_i,\ \sum_i Y_i,\ (m2^n)^{-1}[\mathrm{Tr}(Y_1),\dots,\mathrm{Tr}(Y_q)]\right),$$

and $u_n = (0, 0, -I_{m2^n}, -[\frac{m_1}{m}, \dots, \frac{m_q}{m}]^\top)$. Here $\{A_n\}, \{A_n^*\}, \{B_n\}, \{B_n^*\}$ are all morphisms and $u_n = u_{n+1}$ under the above embedding, hence Proposition 2.3.2(a) applies.

Again, we can use Theorem 2.3.6 to describe an infinite-dimensional limit $\overline{\mathrm{SH}(\lambda)_\infty}$ of these Schur–Horn orbitopes. We would like to interpret the elements of $\mathrm{SH}(\lambda)_\infty$ as operators with spectral measure $\sum_{i=1}^q \frac{m_i}{m}\delta_{\lambda_i}$, then describe the convex hull of all such operators in $\overline{\mathbb{V}_\infty}$. To that end, we complete $\mathbb{V}_\infty$ with respect to the operator norm and consider spectral measures with respect to the normalized trace τ on $\mathbb{V}_\infty$, which also extends to the limit. Such spectral measures and associated orbitopes exist in more general algebras. In fact, our framework yields descriptions for Schur–Horn orbitopes, as well as a generalization of the Schur–Horn theorem, to self-adjoint elements in so-called approximately finite-dimensional (AF) algebras. These orbitopes generalize both the permutahedra and Schur–Horn orbitopes in this section. We construct these algebras and apply our framework to the resulting orbitopes in Appendix C. In particular, we obtain the following description for the limit in our present setting, proved in Proposition C.0.2 of that appendix.

Proposition 2.4.2. *Endow $\mathbb{V}_\infty$ with the operator norm and let $\tau\colon \mathbb{V}_\infty \to \mathbb{R}$ be the normalized trace given by $\tau(X) = (m2^n)^{-1}\mathrm{Tr}(X)$ for $X \in \mathbb{V}_n$. Then τ extends to $\overline{\mathbb{V}_\infty}$, and each $X \in \overline{\mathbb{V}_\infty}$ has a spectral measure μ_X^τ with respect to τ. Furthermore,*

$$\overline{\mathrm{SH}(\lambda)_\infty} = \overline{\mathrm{conv}}\left\{X \in \overline{\mathbb{V}_\infty} : \mu_X^\tau = \sum_{i=1}^q \frac{m_i}{m}\delta_{\lambda_i}\right\} = \overline{\left\{\sum_{i=1}^q \lambda Y_i : Y_i \succeq 0,\ \sum_i Y_i = I,\ \tau(Y_i) = \frac{m_i}{m}\right\}}.$$

Furthermore, the diagonal map $\mathrm{diag}_n\colon \mathbb{S}^{m2^n} \to \mathbb{R}^{m2^n}$ satisfies $\mathrm{diag}_n(\mathrm{SH}(\lambda)_n) = \mathrm{Perm}(\lambda)_n$ by the Schur–Horn theorem. Our limiting descriptions yield the following limiting version of this theorem, proved in greater generality in Proposition C.0.3 of the above appendix.

Proposition 2.4.3. *In the setting of Proposition 2.4.2, we have $\mathrm{diag}(\overline{\mathrm{SH}(\lambda)_\infty}) = \overline{\mathrm{Perm}(\lambda)_\infty}$ where we complete the permutahedra as in Proposition 2.4.1 with respect to the L^∞ norm.*

Schur–Horn orbitopes are special cases of so-called *spectral polyhedra* studied in [198]. It would be interesting to identify further examples of freely-described sequences of spectral polyhedra arising in applications and to consider their limits.

Remark 2.4.4. *We remark that the above constructions can be extended to handle vectors and matrices of all sizes, not just of sizes $(m2^n)_{n\in\mathbb{N}}$. This can be done by* *indexing our consistent sequences by posets. In particular, our results generalize to the following, more complicated, sequences of group representations. If $\mathcal{N}$ is a strict poset, an $\mathcal{N}$-indexed consistent sequence of $\{\mathsf{G}_n\}_{n\in\mathcal{N}}$-representations is a sequence $\{(V_n, \varphi_{N,n})\}_{n<N\in\mathcal{N}}$ of G_n-representations and embeddings $\varphi_{N,n}\colon V_n \hookrightarrow V_N$ for each $n < N$ such that $\varphi_{N,n}$ is G_n-equivariant, and $\varphi_{M,N}\circ\varphi_{N,n} = \varphi_{M,n}$ whenever $n < N < M$. All our results apply in this setting after replacing all occurrences of $n+1$ by $N > n$. To handle permutahedra and Schur–Horn orbitopes of any sizes we let $\mathcal{N} = \mathbb{N}$ with the divisibility partial order, whereby $n \leq m$ iff $m = nk$ for some $k \in \mathbb{N}$, with embeddings $\varphi_{kn,n}\colon \mathbb{R}^n \hookrightarrow \mathbb{R}^{nk}$ sending $x \mapsto x \otimes \mathbb{1}_k$ or $\varphi_{kn,n}\colon \mathbb{S}^n \hookrightarrow \mathbb{S}^{nk}$ sending $X \mapsto X \otimes I_k$.*

Free Spectrahedra

The family of free spectrahedra in Example 1.1.2(b) may be obtained as a special case of our framework with appropriate selection of description spaces and by imposing compatibility.

Let $\mathscr{V}_0 = \{\mathbb{S}^n\}$ with embeddings by zero-padding and the action of $\mathsf{G}_n = \mathsf{O}_n$ by conjugation. Fix $d, k \in \mathbb{N}$, and let $\mathscr{V} = \mathscr{V}_0^{\oplus d}$, $\mathscr{U} = \mathbb{S}^k \otimes \mathscr{V}_0$, and $\mathscr{W} = \{\mathbb{W}_n = 0\}$.

As the only morphisms $\mathscr{V}_0 \to \mathscr{V}_0$ are multiples of the identity, and elements of $\mathbb{S}^k \otimes \mathbb{S}^n$ can be viewed as $k\times k$ symmetric block matrices with symmetric $n \times n$ blocks of $n\times n$ symmetric matrices, we conclude that the morphisms $\mathscr{V} \to \mathscr{U}$ are precisely maps of the form $(X_1,\dots,X_d) \mapsto \sum_i L_i \otimes X_i$ for some $L_1,\dots,L_d \in \mathbb{S}^k$. Note that sequences of adjoints of such maps are also morphisms in this case. As the only G_n-invariants in $\mathbb{S}^n$ are multiples of I_n, the space of freely-described elements in $\mathscr{U}$ is $\{\{L_0 \otimes I_n\}_n : L_0 \in \mathbb{S}^k\}$, which satisfy the condition $L_0 \otimes (I_{n+1} - I_n) \succeq 0$ from Proposition 2.3.2(a) if and only if $L_0 \succeq 0$. Thus, the parametric family of dimension-free descriptions certifying compatibility as in Proposition 2.3.2(a) is

$$(\mathcal{D}_{\mathcal{L}})_n = \left\{(X_1,\dots,X_d) \in (\mathbb{S}^n)^d : L_0 \otimes I_n + \sum_{i=1}^d L_i \otimes X_i \succeq 0\right\}, \quad L_0 \succeq 0.$$

which are free spectrahedra parametrized by $\mathcal{L} = (L_0,\dots,L_d)$. It is common to take either $L_0 = I_k$ (the *monic* case) or $L_0 = 0$ (the *homogeneous* case) [130]. As discussed in Section 2.1, free spectrahedra are fundamental objects in noncommutative free convex and algebraic geometry, see [101, 130] for an introduction. In particular, they satisfy both intersection and projection compatibility, and more generally closure under so-called *matrix-convex combinations*.

We remark that orthogonal invariance and compatibility alone do not yield closure under matrix-convex combinations. For example, the nuclear norm balls $\{X \in \mathbb{S}^n : \|X\|_* \leq 1\}$ are O_n-invariant and form a compatible sequence which is not matrix-convex. However, a compatible sequence of orthogonally-invariant convex *cones* is indeed matrix convex, as the following result shows.

Proposition 2.4.5. *Let $\{\mathbb{V}_n = (\mathbb{S}^n)^d\}$ be the consistent sequence of $\{\mathsf{O}_n\}$-representations above, and suppose $\{C_n \subseteq \mathbb{V}_n\}$ is an intersection- and projection-compatible sequence of convex cones such that C_n is O_n-invariant for all n. Then $\{C_n\}$ is matrix-convex.*

Proof. By [100, §2.3], it suffices to show that if $(X_1, \ldots, X_d) \in C_n$ and $(Y_1, \ldots, Y_d) \in C_m$ then $(X_1 \oplus Y_1, \ldots, X_d \oplus Y_d) \in C_{n+m}$, and that $(V^\top X_1 V, \ldots, V^\top X_d V) \in C_k$ for any isometry $V \in \mathbb{R}^{n\times k}$. For the first, intersection compatibility shows that $(X_1 \oplus 0_m, \ldots, X_d \oplus 0_m) \in C_{n+m}$ and $(Y_1 \oplus 0_n, \ldots, Y_d \oplus 0_n) \in C_{n+m}$. Conjugating the latter tuple by appropriate permutation matrices, we conclude that $(0_n \oplus Y_1, \ldots, 0_n \oplus Y_d) \in C_{n+m}$. Finally, since C_{n+m} is a convex cone we get $(X_1 \oplus Y_1, \ldots, X_d \oplus Y_d) \in C_{n+m}$, so $\{C_n\}$ is closed under direct sums. For the second, any isometry $V \in \mathbb{R}^{n\times k}$ can be written as $V = U\iota_{n,k}$ where $U \in \mathsf{O}_n$ and $\iota_{n,k} : \mathbb{R}^k \to \mathbb{R}^n$ is the zero-padding embedding. Observe that $(\iota_{n,k}^\top X_1 \iota_{n,k}, \ldots, \iota_{n,k}^\top X_d \iota_{n,k}) = \mathcal{P}_k(X_1, \ldots, X_d)$ is the orthogonal projection of $(X_1, \ldots, X_d)$ onto $\mathbb{V}_k$ when embedded in $\mathbb{V}_n$. Thus, $(V^\top X_1 V, \ldots, V^\top X_d V) = \mathcal{P}_k(U^\top X_1 U, \ldots, U^\top X_d U) \in C_k$ since C_n is O_n-invariant and since $\{C_n\}$ is projection-compatible. □

To recap, orthogonal invariance and compatibility of a sequence of convex sets do not yield matrix convexity in general, but they do so for convex cones. With additional choices of description spaces, our framework yields a particular parametric family of matrix-convex sets, namely free spectrahedra.

Spectral Functions, (Quantum) Entropy, and Variants

Let $\mathscr{V} = \{\mathbb{S}^n\}$ with embeddings by zero-padding and with the action of O_n by conjugation, and let $\mathscr{V}' = \{\mathbb{R}^n\}$ with embeddings by zero-padding and the standard action of S_n. Recall (e.g., [201]) that a convex function $F_n : \mathbb{S}^n \to \mathbb{R}$ is O_n-invariant if and only if there exists an S_n-invariant convex function $f_n : \mathbb{R}^n \to \mathbb{R}$ satisfying $F_n(X) = f_n(\lambda(X))$ where $\lambda(X) \in \mathbb{R}^n$ is the vector of eigenvalues of $X \in \mathbb{S}^n$. Also, the sequence $\{F_n : \mathbb{S}^n \to \mathbb{R}\}$ is intersection-compatible if and only if the sequence $\{f_n : \mathbb{R}^n \to \mathbb{R}\}$ is.

Examples of such sequences of functions $\mathfrak{F} = \{F_n\}$ and $\mathfrak{f} = \{f_n\}$ arise in (quantum) information theory, where $\mathfrak{F}$ is the quantum analog of classical information-theoretic parameters $\mathfrak{f}$. These are often intersection-compatible as distributions on n states can be viewed as distributions on $n+1$ states with zero probability on the last state. For example, the negative entropy and relative entropy and their quantum variants are given by

$$\begin{aligned} h_n(x) &= \sum_i x_i \log x_i, \qquad \mathsf{H}_n(X) = h_n(\lambda(X)) = \mathrm{Tr}(X \log X),\\ D_n(x,y) &= \sum_i x_i \log \frac{x_i}{y_i}, \quad S_n(X,Y) = D_n(\lambda(X),\lambda(Y)) = \mathrm{Tr}(X(\log X - \log Y)). \end{aligned} \tag{2.16}$$

Here $\mathrm{dom}(h_n) = \Delta^{n-1}$ and $\mathrm{dom}(D_n) = (\mathbb{R}^n_+)^2$, while $\mathrm{dom}(\mathsf{H}_n) = \mathcal{D}^{n-1}$ is the spectraplex from Example 1.1.1(b) and $\mathrm{dom}(S_n) = (\mathbb{S}^n_+)^2$. We use the standard convention that $0 \log \frac{0}{y} = 0$ even if $y = 0$, and $x \log \frac{x}{0} = \infty$ when $x \neq 0$ [52, §2.3]. These sequences of functions are intersection-compatible but not projection-compatible (e.g., their domains are not projection-compatible).

Semidefinite approximations: The functions (2.16) are not semidefinite-representable (i.e., cannot be evaluated using semidefinite programming), though semidefinite approximations of them have been proposed in the literature [74]. We show that these approximations are freely-described, but that these descriptions do not certify intersection compatibility. The family of approximations of [74] to the negative quantum entropy is parametrized by $m, k \in \mathbb{N}$ via their epigraphs

$$\begin{aligned} E_n^{(m,k)} = \Bigg\{ (X,t) \in \mathbb{S}^n \oplus \mathbb{R} \,\Bigg|\, &\exists T_0,\dots,T_m, Z_0,\dots,Z_k \in \mathbb{S}^n \text{ s.t. } Z_0 = I_n,\ \sum_{j=1}^m w_j T_j = -2^{-k} T_0,\\ &\begin{bmatrix} Z_i & Z_{i+1} \\ Z_{i+1} & X \end{bmatrix} \succeq 0, \text{ for } i = 0,\dots,k-1,\ \begin{bmatrix} Z_k - X - T_j & -\sqrt{s_j} T_j \\ -\sqrt{s_j} T_j & X - s_j T_j \end{bmatrix} \succeq 0,\\ &\text{for } j = 1,\dots,m,\ \mathrm{Tr}(T_0) \le t \Bigg\}, \end{aligned}$$

where $s, w \in \mathbb{R}^m$ are the nodes and weights for Gauss-Legendre quadrature.

This is a dimension-free description of the form (ConicSeq) which almost, but not quite, satisfies the conditions of Proposition 2.3.2. Indeed, let $\mathscr{W} = \mathscr{V}^{\oplus(m+k+1)}$ and $\mathscr{U} = \mathscr{V} \oplus (\mathbb{S}^2 \otimes \mathscr{V})^{\oplus(m+k)} \oplus \mathbb{R}$ containing the cones $\{\mathcal{K}_n = \{0\} \oplus (\mathbb{S}^2 \otimes \mathbb{S}^n)_+^{\oplus(m+k)} \oplus \mathbb{R}_+\}$.

Define

$$A_n(X,t) = \left(0, \left(\begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix} \otimes X\right)^{\oplus k}, \left(\begin{bmatrix}-1 & 0\\ 0 & 1\end{bmatrix} \otimes X\right)^{\oplus m}, t\right),$$

$$B_n(T_0,\dots,T_m,Z_1,\dots,Z_k) = \left(2^{-k}T_0 + \sum_{j=1}^m w_jT_j, \begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix} \otimes Z_1, \bigoplus_{i=1}^{k-1}\left(\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} \otimes Z_i + \begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix} \otimes Z_{i+1}\right),\right.$$

$$\left.\bigoplus_{j=1}^m\left(\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} \otimes Z_k - \begin{bmatrix}1 & \sqrt{s_j}\\ \sqrt{s_j} & s_j\end{bmatrix} \otimes T_j\right), -\mathrm{Tr}(T_0)\right),$$

$$u_n = \left(0, \begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} \otimes I_n, 0, \dots, 0\right).$$

Note that $\{A_n\}$ and $\{B_n\}$ are morphisms, that $\{u_n\}$ is a freely-described element of $\mathscr{U}$ satisfying $u_{n+1} - u_n \in \mathcal{K}_{n+1}$, but that $\{B_n^*\}$ is *not* a morphism because of the $\mathrm{Tr}(T_0)$ term in B_n. By the proof of Proposition 2.3.2, this description certifies that $E_n^{(m,k)} \subseteq E_{n+1}^{(m,k)}$ but not $E_{n+1}^{(m,k)} \cap (\mathbb{S}^n \oplus \mathbb{R}) \subseteq E_n^{(m,k)}$. Analogously to the naïve description of the solid simplex in Section 2.4, this dimension-free description does not make the intersection compatibility of $\{E_n^{(m,k)}\}$ obvious.

Parametric families: We can use the description spaces of [74] above to derive parametric families of freely-described sets. When compatibility is not required, the resulting family includes the approximation $\{E_n^{(m,k)}\}_n$ of [74]; when compatibility is imposed, we obtain a smaller family excluding $\{E_n^{(m,k)}\}_n$.

Note that $\mathbb{S}^2 \otimes \mathbb{S}^n \cong (\mathbb{S}^n)^3$ as O_n-representations, and these isomorphisms commute with zero-padding, so that $\mathscr{U} \cong \mathscr{V}^{1+3(m+k)} \oplus \mathbb{R}$ as consistent sequences. As $\dim(\mathbb{S}^n)^{\mathsf{O}_n} = 1$ and $\dim \mathcal{L}(\mathbb{S}^n)^{\mathsf{O}_n} = 2$, the dimension of invariants parametrizing dimension-free descriptions are

$$\dim \mathcal{L}(\mathbb{V}_n, \mathbb{U}_n)^{\mathsf{O}_n} = 3[2(m+k)+1], \quad \dim \mathcal{L}(\mathbb{W}_n, \mathbb{U}_n)^{\mathsf{O}_n} = 3(m+k+1)[2(m+k)+1],$$
$$\dim \mathbb{U}_n^{\mathsf{O}_n} = 2 + 3(m+k).$$

When $m = k = 3$ (the default values in the implementation of [74]), we get

$$\dim \mathcal{L}(\mathbb{V}_n, \mathbb{U}_n)^{\mathsf{O}_n} = 39, \quad \dim \mathcal{L}(\mathbb{W}_n, \mathbb{U}_n)^{\mathsf{O}_n} = 273, \quad \dim \mathbb{U}_n^{\mathsf{O}_n} = 20.$$

As the only morphisms of sequences $\mathscr{V} \to \mathscr{V}$ are multiples of the identity, and the only morphisms $\mathscr{V} \to \mathbb{R}$ are multiples of the trace, the dimensions of $\{A_n\}, \{B_n\}$ satisfying the conditions of Proposition 2.3.2(a) are

$$\dim\left\{\{A_n\colon \mathbb{V}_n \to \mathbb{U}_n\} \text{ morphism}\right\} = 3(m+k)+2 = 20,$$
$$\dim\left\{\{B_n\colon \mathbb{W}_n \to \mathbb{U}_n\} : \text{both } \{B_n\} \text{ and } \{B_n^*\} \text{ morphisms}\right\} = (m+k+1)[3(m+k)+1] = 133.$$

Graph Parameters

Let $\mathscr{V} = \{\mathbb{S}^n\}$ with embeddings by zero-padding and the action of $\mathsf{G}_n = \mathsf{S}_n$ by conjugation. As discussed in Section 2.1 and Example 2.1.5, graph parameters are sequences $\{f_n\colon \mathbb{S}^n \to \mathbb{R}\}$ of G_n-invariant functions, and many standard graph parameters are convex (or concave) and satisfy either intersection or projection compatibility. Furthermore, it is desirable to find relaxations for standard graph parameters satisfying the same compatibility conditions. We consider here the examples of max-cut and inverse stability number, and we show that their usual relaxations are freely-described with descriptions certifying their compatibility. We then derive parametric families of dimension-free descriptions from our framework that may be used to fit graph parameters to data.

Max-cut: Computing the max-cut value of a weighted undirected graph amounts to evaluating the support function of the *cut polytope* $\mathrm{CUT}_n = \mathrm{conv}\{xx^\top : x \in \{\pm 1\}^n\}$. The sequence of cut polytopes $\{\mathrm{CUT}_n\}$ viewed as subsets of $\mathscr{V}$ is projection-compatible and compact, hence the sequence of their support functions and the max-cut value itself are intersection-compatible (see Section 2.1). Approximation of the max-cut value reduces to approximation of the cut polytopes. A standard outer approximation of the sequence of cut polytopes is the sequence of elliptopes

$$\mathcal{E}_n = \{X \in \mathbb{S}^n : X \succeq 0, \mathrm{diag}(X) = \mathbb{1}_n\}. \tag{2.17}$$

The sequence $\{\mathcal{E}_n\}$ also satisfies projection compatibility, and the above is a dimension-free description certifying this compatibility. Indeed, let $\mathscr{W} = \{0\}$ and $\mathscr{U} = \mathscr{V} \oplus \{\mathbb{R}^n\}$ where the latter sequence is the usual one from Example 2.1.3, with cones $\mathscr{K} = \{\mathbb{S}^n_+ \oplus \{0\}\}$. Then (2.17) is of the form (ConicSeq) with $A_nX = (X, \mathrm{diag}(X))$, $B_n = 0$, and $u_n = (0, -\mathbb{1}_n)$. Note that $\{A_n\}, \{A_n^*\}$ are morphisms and $u_{n+1} - u_n = (0, -e_{n+1}) = (e_{n+1}e_{n+1}^\top, 0) - A_{n+1}e_{n+1}e_{n+1}^\top$, hence Proposition 2.3.2(b) applies. Neither the cut polytopes nor the elliptopes is intersection-compatible, as zero-padding a matrix with all-1's diagonal does not yield such a matrix. The sequences $\{\mathrm{CUT}_n - I_n\}$ and $\{\mathcal{E}_n - I_n\}$ are, however, both intersection- and projection-compatible, and the shifted elliptopes admit dimension-free descriptions certifying their compatibility.

Inverse stability number: Computing the inverse stability number reduces to evaluating the support functions of $\mathcal{D}_n = \mathrm{conv}\{xx^\top : x \in \Delta^{n-1}\}$, see [159]. A natural SDP relaxation for this problem is evaluating the support function of

$$\widetilde{\mathcal{D}}_n = \{X \in \mathbb{S}^n : X \succeq 0, X \geq 0, \mathbb{1}_n^\top X \mathbb{1}_n = 1\},$$

where $X \geq 0$ denotes an entrywise nonnegative matrix. Both $\{\mathcal{D}_n\}$ and $\{\widetilde{\mathcal{D}}_n\}$ are intersection-compatible. Moreover, the above description of $\{\widetilde{\mathcal{D}}_n\}$ is dimension-free and certifies this compatibility as in Proposition 2.3.2(b). Indeed, let $\mathscr{W} = \{0\}$ and $\mathscr{U} = \mathscr{V} \oplus \mathbb{R}$ with cones $\mathcal{K}_n = (\mathbb{S}^n_+ \cap \mathbb{R}^{n\times n}_+) \oplus 0$. Then the above description of $\widetilde{\mathcal{D}}_n$ is of the form (ConicSeq) with $A_nX = (X, \mathbb{1}_n^\top X\mathbb{1}_n)$, $B_n = 0$, and $u_n = (0,-1)$. Note that $\{A_n\}$ is a morphism and $u_n = u_{n+1}$, hence Proposition 2.3.2(a) applies. Neither $\mathcal{D}_n$ nor $\widetilde{\mathcal{D}}_n$ is projection-compatible since their support functions, which are the inverse stability number and its semidefinite approximation above, are not intersection compatible (see Section 2.1). Indeed, appending isolated vertices to a graph increases its stability number.

Parametric families: Beyond the two particular preceding examples, we derive here an expressive parametric family of convex graph parameters; in addition to above examples, our family also includes several other examples from [39]. To that end, let $\mathscr{W} = \mathscr{U} = \mathrm{Sym}^2(\mathrm{Sym}^{\leq 2}\{\mathbb{R}^n\}) = \{\mathbb{S}^{\binom{n+2}{2}}\}$. The dimensions of invariants parametrizing dimension-free descriptions in this case is too large for us to write explicit bases for them as functions of n. Instead, we compute these dimensions using the algorithm in Section 2.6 below, see Example 2.6.2(b):

$$\dim \mathcal{L}(\mathbb{V}_n,\mathbb{U}_n)^{\mathsf{G}_n} = 93, \quad \dim \mathcal{L}(\mathbb{W}_n,\mathbb{U}_n)^{\mathsf{G}_n} = 1068, \quad \dim \mathbb{W}_n^{\mathsf{G}_n} = 17, \quad \text{for all } n \geq 8.$$

Using the same algorithm, the dimensions of sequences $\{A_n\}, \{B_n\}$ certifying intersection compatibility as in Proposition 2.3.2(a) are

$$\begin{aligned} &\dim \Big\{\{A_n\colon \mathbb{V}_n \to \mathbb{U}_n\} \text{ morphism}\Big\} = 19, \\ &\dim \Big\{\{B_n\colon \mathbb{W}_n \to \mathbb{U}_n\} : \text{both } \{B_n\} \text{ and } \{B_n^*\} \text{ morphisms}\Big\} = 104. \end{aligned} \tag{2.18}$$

The algorithm we present in Section 2.6 further allows us to compute bases for these invariants in a fixed dimension and extend them to any other; in turn, these are useful for fitting graph parameters defined for graphs of all sizes given data.

Graphon Parameters

A different embedding between graphs arises in the theory of graphons [146], where a weighted graph $X \in \mathbb{S}^{2^n}$ is viewed as a step function $W_X\colon [0,1]^2 \to \mathbb{R}$ defined by $W_X(x,y) = X_{i,j}$ if $(x,y) \in [(i-1)/2^n, i/2^n) \times [(j-1)/2^n, j/2^n)$; see Figure 2.2. Note that X and $X \otimes \mathbb{1}_2\mathbb{1}_2^\top \in \mathbb{S}^{2^{n+1}}$ correspond to the same step function, and that the inner product of two such step functions W_X, W_Y in $L^2([0,1]^2)$ equals the normalized Frobenius inner product $\langle X, Y\rangle = 2^{-2n}\mathrm{Tr}(X^\top Y)$. We therefore define the

$$X = \begin{bmatrix} 0 & 0.75 & 0 & 0 & 0.25 & 0 \\ 0.75 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0.25 & 1 & 0 \\ 0 & 0 & 0.25 & 0 & 0 & 0.75 \\ 0.25 & 0 & 1 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 0.75 & 0.5 & 0 \end{bmatrix}$$

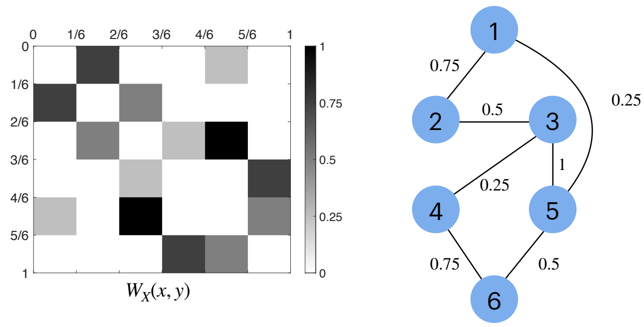


Figure 2.2: Weighted undirected graph represented as a graph, an adjacency matrix X, and a symmetric function (graphon) W_X on $[0, 1]^2$.

graphon consistent sequence $\mathscr{V} = \{\mathbb{V}_n = \mathbb{S}^{2^n}\}$ with embeddings $\varphi_n(X) = X \otimes \mathbb{1}_{2\times 2}$, the above normalized inner products, and the action of $\mathsf{G}_n = \mathsf{S}_{2^n}$ by conjugation. Here G_n is embedded into G_{n+1} by sending a permutation matrix g to $g \otimes I_2$. We can extend our consistent sequence to include symmetric matrices of any size by having our consistent sequence be indexed by the poset $\mathbb{N}$ with the divisibility partial order, see Remark 2.4.4. The graphon sequence is finitely-generated, as the following computer-assisted proof shows.

Proposition 2.4.6. *The graphon sequence $\{\mathbb{V}_n = \mathbb{S}^{2^n}\}$ is generated in degree* 2.

Proof. Define $E_1^{(n)} = e_1^{(2^n)}(e_1^{(2^n)})^\top$ and $E_2^{(n)} = e_1^{(2^n)}(e_2^{(2^n)})^\top + e_2^{(2^n)}(e_1^{(2^n)})^\top$, whose G_n-orbits span $\mathbb{V}_n$. We verify computationally that $\dim \sum_{i=1}^2 \mathbb{R}[\mathsf{S}_{2^3}]\varphi_2(E_i^{(2)}) = \dim \mathbb{V}_3$, see the GitHub repository. Therefore,

$$E_i^{(3)} = \sum\nolimits_{j=1}^{2} r_{i,j}\varphi_2(E_j^{(2)}), \qquad \text{for } i \in [2],\ r_{i,j} \in \mathbb{R}[S_{2^3}]. \tag{2.19}$$

Let $\psi_n(X) = X \oplus 0$ be an embedding of $\mathbb{V}_n$ into $\mathbb{V}_{n+1}$ by zero-padding, and note that $E_i^{(n+1)} = \psi_n(E_i^{(n)})$ and that ψ_n commutes with a different embedding of S_{2^n} into $\mathsf{S}_{2^{n+1}}$, namely, one that sends $g \mapsto g \oplus I_{2^n}$. Applying ψ_n to (2.19), we conclude that $E_i^{(n+1)}$ can be written as $\mathbb{R}[\mathsf{S}_{2^{n+1}}]$-linear combinations of $\varphi_n(E_j^{(n)})$ for all $n \geq 2$, hence that $\mathbb{R}[\mathsf{S}_{2^{n+1}}]\varphi_n(\mathbb{V}_n) = \mathbb{V}_{n+1}$ for all $n \geq 2$. $\square$

Graphon parameters: A permutation-invariant and intersection-compatible sequence of functions $\mathfrak{f} = \{f_n : \mathbb{V}_n \to \mathbb{R}\}$ is called a *graphon parameter*, since these are precisely the functions of graphs that do not depend on their inputs' vertex labels, and that only depend on their input graphs via their associated graphons. A family of graphon parameters that plays a central role in the theory of graphons and in extremal combinatorics are graph homomorphism densities [146]. Their convexity is related to weakly-norming graphs and Sidorenko's conjecture, a major open problem in extremal combinatorics [136, 205].

We can obtain parametric families of graphon parameters by taking the gauge functions of parametric families of intersection-compatible and freely-described convex sets. For example, let $\mathscr{U} = \mathbb{S}^k \otimes \mathscr{V}$ with cones $\mathscr{K} = \{\mathcal{K}_n = (\mathbb{S}^k \otimes \mathbb{S}^{2^n})_+\}$ and $\mathscr{W} = \{\mathbb{W}_n = 0\}$. Using Proposition 2.3.2(a), we get the following parametric family of freely-described and intersection-compatible sets, parametrized by $L_1, \ldots, L_7 \in \mathbb{S}^k$:

$$C_n = \left\{ X \in \mathbb{S}^{2^n} : \frac{\mathbb{1}^\top X \mathbb{1}}{2^{2n}} L_1 \otimes \mathbb{1}\mathbb{1}^\top + \frac{\mathrm{Tr}(X)}{2^n} L_2 \otimes \mathbb{1}\mathbb{1}^\top + L_3 \otimes \frac{1}{2^n}\left(X\mathbb{1}\mathbb{1}^\top + \mathbb{1}\mathbb{1}^\top X\right) \right.$$
$$\left. + L_4 \otimes \left(\mathrm{diag}(X)\mathbb{1}^\top + \mathbb{1}\mathrm{diag}(X)^\top\right) + L_5 \otimes X + L_6 \otimes \mathbb{1}\mathbb{1}^\top + L_7 \otimes (2^n I_{2^n}) \succeq 0 \right\}, \tag{2.20}$$

where $L_7 \succeq 0$. Note that all the functions of X appearing in the above description can be interpreted as natural functions of the associated step graphon W_X. For example, $\frac{\mathbb{1}^\top X \mathbb{1}}{2^{2n}} = \int_{[0,1]^2} W_X(t,s)\,\mathrm{d}t\,\mathrm{d}s$ and $\frac{\mathrm{Tr}(X)}{2^n} = \int_0^1 W_X(t,t)\,\mathrm{d}t$. Similarly, $\frac{1}{2^n}(X\mathbb{1})_i = \int_{(i-1)/2^n}^{i/2^n} W_X((i-1/2)/2^n, y)\,dy$ gives the values of the degree function of the graphon W_X, and $\mathrm{diag}(X)$ gives the values of the diagonal function $W_X(x,x)$ as x ranges over $[0,1]$, see [146, Chap. 7].

Expressing the above dimension-free description in terms of graphons yields a description of a limiting convex set by Theorem 2.3.6. Its gauge function is, in turn, a convex graphon parameter extending continuously to the corresponding limit. Endow $\mathbb{V}_\infty$ with the $L^\infty([0,1]^2)$-norm, noting that $\|\mathcal{P}_n W\|_\infty \leq \|W\|_\infty$ for all $W \in L^\infty([0,1]^2)$ by Jensen's inequality. We view elements $[W_{i,j}(x,y)]_{i,j=1}^k \in \mathbb{S}^k \otimes L^\infty([0,1]^2)$ as symmetric matrices with entries in $L^\infty([0,1]^2)$ (so $W_{j,i} = W_{i,j}$), and denote

$$[W_{i,j}(x,y)]_{i,j=1}^k \succeq 0 \quad \text{if} \quad \int_{[0,1]^2} \sum_{i,j=1}^k W_{i,j}(x,y) f_i(x) f_j(y)\,dx\,dy \geq 0 \quad \text{for all} \quad f_1, \ldots, f_k \in L^2([0,1]). \tag{2.21}$$

Remarkably, convex graphon parameters that extend continuously to this limit are also projection-compatible by [63, Thm. 3.17].

Proposition 2.4.7. *If $L_7 = 0$ in (2.20) then*

$$\overline{C_\infty} = \Bigg\{ W \in \overline{\mathbb{V}_\infty} : \overline{A_\infty}(W) + u_\infty := \Big[(L_1)_{i,j} \int_{[0,1]^2} W(s,t)\, ds\, dt + (L_2)_{i,j} \int_{[0,1]} W(t,t)\, dt$$
$$+ (L_3)_{i,j} \int_{[0,1]} [W(x,t) + W(t,y)]\, dt + (L_4)_{i,j} [W(x,x) + W(y,y)] + (L_5)_{i,j} W(x,y) \Big]_{i,j=1}^k$$
$$+ L_6 \succeq 0 \Bigg\},$$

where $\overline{\mathbb{V}_\infty} \subseteq L^\infty([0,1]^2)$ is a subspace of symmetric bounded measurable functions on $[0,1]^2$.

Proof. Endow $\mathbb{U}_\infty = \mathbb{S}^k \otimes \mathbb{V}_\infty$ with the norm $\|[W_{i,j}]_{i,j=1}^k\| = \max_{i,j} \|W_{i,j}\|_\infty$, so $\overline{\mathbb{U}_\infty} = \mathbb{S}^k \otimes \overline{\mathbb{V}_\infty}$. Then $\overline{\mathcal{K}_\infty}$ is the cone of PSD matrices $[W_{i,j}(x,y)]_{i,j=1}^k \succeq 0$ in $\overline{\mathbb{U}_\infty}$. Furthermore, for any $W \in \mathbb{V}_\infty$ we have

$$\|A_\infty\| \leq \max_{i,j} \Big(|(L_1)_{i,j}| + |(L_2)_{i,j}| + 2|(L_3)_{i,j}| + 2|(L_4)_{i,j}| + |(L_5)_{i,j}| \Big) \|W\|_\infty,$$

so A_∞ extends continuously to the limit, and $u_\infty = L_6 \in U_1 \subseteq \overline{\mathbb{U}_\infty}$ satisfies $\mathcal{P}_n u_\infty = L_6$ for all n. Thus, Theorem 2.3.6 yields the claim. □

We require $L_7 = 0$ since there is no $u_\infty \in \overline{\mathbb{V}_\infty}$ satisfying $\mathcal{P}_n u_\infty = 2^n I_{2^n}$ for all n.

Compatibility in Inverse Problems

Our compatibility conditions naturally arise in the context of inverse problems, where we demonstrate that it is desirable to use regularizers which are both intersection- and projection-compatible.

Consider a consistent sequence $\mathscr{V} = \{\mathbb{V}_n\}$ of $\{\mathsf{G}_n\}$-representations. A popular approach to recover $x \in \mathbb{V}_n$ from $m \in \mathbb{N}$ linear observations takes as input a forward map $A\colon \mathbb{V}_n \to \mathbb{R}^m$ and data $y \in \mathbb{R}^m$ and outputs

$$F_{m,n}(A,y) = \operatorname{argmin} x \in \mathbb{V}_n\ f_n(x) + \lambda \|Ax - y\|_2^2, \tag{2.22}$$

where $f_n\colon \mathbb{V}_n \to \mathbb{R}$ is a convex regularizer promoting desired structure in the solution. It is desirable for the maps $F_{m,n}$ defined in (2.22)—which can be instantiated for any $(A,y) \in \mathcal{L}(\mathbb{V}_n, \mathbb{R}^m) \oplus \mathbb{R}^m$ and for any $n, m \in \mathbb{N}$—to satisfy

$$F_{m,n+1}(A\mathcal{P}_n, y) = F_{m,n}(A,y), \tag{2.23}$$

whenever the corresponding minimizers are unique. Indeed, condition (2.23) requires the recovered solution to lie in $\mathbb{V}_n$ if the data only depends on the component of $x \in \mathbb{V}_{n+1}$ in $\mathbb{V}_n$, as this property avoids overfitting. Condition (2.23) holds when the sequence of regularizers is both intersection and projection-compatible. Indeed, if the sequence of regularizers $\mathfrak{f} = \{f_n\}$ is *projection-compatible*, then

$$\min_{\widetilde{x}\in\mathbb{V}_{n+1}} f_{n+1}(\widetilde{x})+\lambda\|A\mathcal{P}_n\widetilde{x}-b\|_2^2 = \min_{x\in\mathbb{V}_n}\min_{\substack{\widetilde{x}\in\mathbb{V}_{n+1}\\ \mathcal{P}_n\widetilde{x}=x}} f_{n+1}(\widetilde{x})+\lambda\|Ax-y\|_2^2 = \min_{x\in\mathbb{V}_n} f_n(x)+\lambda\|Ax-y\|_2^2$$

Moreover, if $x_* = F_{m,n}(A,y)$ minimizes $f_n(x) + \lambda\|Ax - y\|_2^2$ and $\mathfrak{f}$ is *intersection-compatible*, then $f_n(x_*) + \lambda\|Ax_* - y\|_2^2 = f_{n+1}(x_*) + \lambda\|A\mathcal{P}_n x_* - y\|_2^2$ and hence $x_* = F_{m,n+1}(A\mathcal{P}_n, y)$, showing (2.23).

2.5 Constant-Sized Invariant Conic Programs

In the previous sections, we studied freely-described sequences of convex sets $\{C_n\}$ contained in a consistent sequence $\mathscr{V}$. These convex sets are group-invariant whenever the cones $\mathcal{K}_n$ in their descriptions (ConicSeq) are group-invariant, which is the case for all the standard sequences of cones we consider. In this section, we further consider optimizing sequences of invariant linear functionals over such sequences of sets. Each program in the sequence can be simplified by restricting its domain to invariant vectors [86, §3]. As we have seen in Proposition 2.2.3, when $\mathscr{V}$ is finitely-generated the dimensions of its spaces of invariants stabilize, so the size of the variables in such programs stabilizes as well. However, the size of the constraints may not stabilize, because the invariant sections of the cones $\{\mathcal{K}_n^{\mathsf{G}_n}\}$ may grow in complexity. For example, if $\mathcal{K}_n$ is the cone of n-variate degree k polynomials that are nonnegative over all of $\mathbb{R}^n$ and $\mathsf{G}_n = \mathsf{S}_n$, the best-known description of $\mathcal{K}_n^{\mathsf{G}_n}$ has complexity which is a (nonconstant) polynomial in n [5, 187, 211]. We therefore seek conditions for the existence of *constant-sized* descriptions for $\{\mathcal{K}_n^{\mathsf{G}_n}\}$, and bounds on the value of n after which the size stabilizes in the sense of Definition 2.1.8. Constant-sized descriptions for symmetric PSD and relative entropy cones have been obtained on a case-by-case basis in the literature [58, 160, 185, 188]. In this section, we explain how these results can be generalized and derived systematically from an interplay between representation stability and the structure of the cones in question.

The PSD Cone and Variants

We begin by giving constant-sized descriptions for certain sequences of PSD cones. We do so by deriving constant-sized bases for spaces of invariants in terms of which membership in the cones is simply expressed. The following is a precise statement of Theorem 2.1.9 stated informally in Section 3.1.

Theorem 2.5.1. *Let* $\mathscr{V}_0 = \{\mathbb{R}^n\}$ *with* $\mathsf{G}_n = \mathsf{B}_n, \mathsf{D}_n$*, or* S_n *as in Example 2.1.3, and let* $\mathscr{V} = \{\mathbb{V}_n\}$ *be a* $\mathscr{V}_0$*-module generated in degree* d *and presented in degree* k*. Then the cones* $\{\mathrm{Sym}^2_+(\mathbb{V}_n)^{\mathsf{G}_n}\}$ *admit constant-sized descriptions for* $n \geq k + d$*.*

Proof. By Theorem 2.2.16, there exists a finite set Λ satisfying

$$\mathbb{V}_n = \bigoplus_{\lambda \in \Lambda} \underbrace{\mathbb{W}_{\lambda[n]}^{m_\lambda}}_{=:\mathbb{V}_{\lambda[n]}}$$

where $\mathbb{W}_{\lambda[n]}$ is a G_n-irreducible and $\mathbb{V}_{\lambda[n]}$ is the corresponding isotypic component. Invariant elements of $\mathbb{U}_n = \mathrm{Sym}^2(\mathbb{V}_n)$ are equivariant and self-adjoint endomorphisms of $\mathbb{V}_n$. If $X \in \mathbb{U}_n^{\mathsf{G}_n}$ is such an endomorphism and $\lambda \neq \mu \in \Lambda$ index distinct irreducibles, then $\mathcal{P}_{\mathbb{V}_{\mu[n]}} X|_{\mathbb{V}_{\lambda[n]}} = 0$ by Schur's Lemma [81, §1.2]. Because the irreducibles of $\mathsf{G}_n = \mathsf{B}_n, \mathsf{D}_n$, and S_n are of real type [224] (meaning they remain irreducible when complexified), Schur's Lemma also implies that $\mathcal{P}_{\mathbb{W}_{\lambda[n]}} X|_{\mathbb{W}_{\lambda[n]}}$ is a multiple of the identity for each $\lambda \in \Lambda$, hence $\mathcal{P}_{\mathbb{V}_{\lambda[n]}} X|_{\mathbb{V}_{\lambda[n]}} = X_\lambda \otimes I_{\dim \mathbb{W}_{\lambda[n]}}$ for some $X_\lambda \in \mathbb{S}^{m_\lambda}$. We conclude that there exists an orthogonal matrix Q_n depending on the irreducible decomposition of $\mathbb{V}_n$ satisfying

$$\mathbb{U}_n^{\mathsf{G}_n} = \left\{ Q_n \left(\bigoplus_{\lambda \in \Lambda} (X_\lambda \otimes I_{\dim \mathbb{W}_{\lambda[n]}}) \right) Q_n^* : X_\lambda \in \mathbb{S}^{m_\lambda} \right\}, \tag{2.24}$$

hence

$$\mathrm{Sym}^2_+(\mathbb{V}_n)^{\mathsf{G}_n} = \left\{ X \in \mathbb{U}_n^{\mathsf{G}_n} : X \succeq 0 \right\} = \left\{ Q_n \left(\bigoplus_{\lambda \in \Lambda_n} (X_\lambda \otimes I_{\dim \mathbb{W}_\lambda}) \right) Q_n^* : X_\lambda \in \mathbb{S}_+^{m_\lambda} \right\}. \tag{2.25}$$

Thus, we obtain constant-sized descriptions by defining $\mathbb{U} = \bigoplus_{\lambda \in \Lambda} \mathbb{S}^{m_\lambda}$ and $T_n : \mathbb{U} \to \mathbb{U}_n^{\mathsf{G}_n}$ sending $(X_\lambda)_{\lambda \in \Lambda}$ to $Q_n \bigoplus_{\lambda \in \Lambda} (X_\lambda \otimes I_{\dim \mathbb{W}_\lambda}) Q_n^*$, which maps $\mathcal{K} = \bigoplus_{\lambda \in \Lambda} \mathbb{S}_+^{m_\lambda}$ onto $\mathrm{Sym}^2_+(\mathbb{V}_n)^{\mathsf{G}_n}$. ◻

We now instantiate $\mathscr{V}$ to obtain more concrete corollaries.

Corollary 2.5.2. *If* $\mathsf{G}_n = \mathsf{S}_n, \mathsf{D}_n$ *or* B_n *acts on* $\mathbb{R}^n$ *as in Example 2.1.3, then the cones* $\mathrm{Sym}^2_+(\mathrm{Sym}^{\leq d}\mathbb{R}^n)^{\mathsf{G}_n} \cong \left(\mathbb{S}_+^{\binom{n+k}{k}}\right)^{\mathsf{G}_n}$ *admit constant-sized descriptions by* (2.25) *for* $n \geq 2d$ *if* $\mathsf{G}_n = \mathsf{S}_n, \mathsf{B}_n$ *and* $n \geq 2d + 1$ *if* $\mathsf{G}_n = \mathsf{D}_n$*.*

Proof. The sequence $\mathscr{V} = \mathrm{Sym}^{\leq d}\mathscr{V}_0$ is generated in degree d and presented in degree d if $\mathsf{G}_n = \mathsf{S}_n, \mathsf{B}_n$ or in degree $d+1$ if $\mathsf{G}_n = \mathsf{D}_n$, by Theorem 2.2.11 and Example 2.2.8. □

To obtain constant-sized descriptions for cones of invariant sums-of-squares, we consider equivariant images of the PSD cones above.

Proposition 2.5.3. *Let $\mathscr{U} = \{\mathbb{U}_n\}$ and $\mathscr{W} = \{\mathbb{W}_n\}$ be sequences of $\{\mathsf{G}_n\}$-representations (not necessarily consistent). Let $\{\mathcal{K}_n \subseteq \mathbb{U}_n\}$ be a sequence of convex cones such that $\mathcal{K}_n$ is G_n-invariant for each n. If $\{\mathcal{K}_n^{\mathsf{G}_n}\}$ admits constant-sized descriptions for $n \geq t$, then so does $\{\pi_n(\mathcal{K}_n \cap \mathbb{L}_n)^{\mathsf{G}_n} \subseteq \mathbb{W}_n\}$ for any sequence $\{\pi_n \in \mathcal{L}(\mathbb{U}_n, \mathbb{W}_n)^{\mathsf{G}_n}\}$ and any sequence of G_n-invariant subspaces $\mathbb{L}_n \subseteq \mathbb{U}_n$.*

Proof. Suppose $\mathcal{K}_n^{\mathsf{G}_n} = T_n(\mathcal{K} \cap \mathbb{L}_n')$ for $n \geq t$ where $T_n \colon \mathbb{U} \to \mathbb{U}_n^{\mathsf{G}_n}$ and $\mathbb{L}_n' \subseteq \mathbb{U}$ are subspaces as in Definition 2.1.8. Because π_n is G_n-equivariant,

$$\pi_n(\mathcal{K}_n \cap \mathbb{L}_n)^{\mathsf{G}_n} = \pi_n(\mathcal{K}_n^{\mathsf{G}_n} \cap \mathbb{L}_n^{\mathsf{G}_n}) = (\pi_n \circ T_n)(\mathcal{K} \cap \mathbb{L}_n' \cap T_n^{-1}(\mathbb{L}_n^{\mathsf{G}_n})).$$

Noting that $\mathbb{L}_n' \cap T_n^{-1}(\mathbb{L}_n^{\mathsf{G}_n})$ is a subspace of $\mathbb{U}$, we get the desired constant-sized descriptions. □

We now prove Theorem 2.1.10 giving constant-sized descriptions for invariant sums-of-squares. s the vector whose coordinates are all the monomials in

Proof (Theorem 2.1.10). If $v(x)$ is the vector whose coordinates are all the monomials of degree at most d in the variables $x_{i_1,\ldots,i_k}$ with $1 \leq i_1 \leq \ldots \leq i_k \leq n$, then [173, Thm. 3.39] yields

$$\mathrm{SOS}_{\mathbb{U}_n} = \pi_n\left(\mathrm{Sym}_+^2\left(\mathrm{Sym}^{\leq d}\left(\mathrm{Sym}^k\mathbb{R}^n\right)\right)\right), \quad \pi_n(M) = v(x)^\top M v(x) + \mathcal{I}_n.$$

The map $\pi_n \colon \mathrm{Sym}^2\left(\mathrm{Sym}^{\leq d}\left(\mathrm{Sym}^k\mathbb{R}^n\right)\right) \to \mathbb{U}_n$ is equivariant by definition of the action of G_n and the invariance of $\mathcal{I}_n$. If $\mathscr{V}_0 = \{\mathbb{R}^n\}$ as in Example 2.1.3, then $\mathscr{V} = \mathrm{Sym}^{\leq d}\left(\mathrm{Sym}^k\mathscr{V}_0\right)$ is a $\mathscr{V}_0$-module generated in degree kd, and presented in degree kd if $\mathsf{G}_n = \mathsf{S}_n, \mathsf{B}_n$ and in degree $kd+1$ if $\mathsf{G}_n = \mathsf{D}_n$ by Theorem 2.2.11. Thus, the result follows from Theorem 2.5.1 and Proposition 2.5.3. □

Note that Theorem 2.1.10 applies to *any* sequence of invariant ideals, not necessarily related to each other across dimensions, so that $\{\mathbb{U}_n\}$ in this result is not necessarily a consistent sequence. Nevertheless, it is a sequence of equivariant images of a consistent sequence, a fact crucial to the proof.

The Relative Entropy Cone and Variants

For a finite set $\mathcal{A}$, define the associated relative entropy cone

$$\mathrm{RE}_{\mathcal{A}} = \{(\nu, c, t) \in \mathbb{R}^{\mathcal{A}} \oplus \mathbb{R}^{\mathcal{A}} \oplus \mathbb{R} : \nu, c \geq 0,\ D(\nu, c) \leq t\}, \tag{2.26}$$

where $D(\nu, c) = \sum_{\alpha\in\mathcal{A}} \nu_\alpha \log\left(\frac{\nu_\alpha}{c_\alpha}\right)$ is the relative entropy.

Proposition 2.5.4. *Let $\mathscr{V}_0$ be a consistent sequence of $\{\mathsf{G}_n\}$ representations and $\mathscr{V} = \{\mathbb{R}^{\mathcal{A}_n}\}$ be a permutation $\mathscr{V}_0$-module for finite $\mathcal{A}_n$ satisfying $\mathcal{A}_n \subseteq \mathcal{A}_{n+1}$ (Definition 2.2.14). If $\dim \mathbb{V}_n^{\mathsf{G}_n}$ is constant for all $n \geq d$, then the invariant section $\{\mathrm{RE}_{\mathcal{A}_n}^{\mathsf{G}_n}\}$ of the cones (2.26) admit constant-sized descriptions for $n \geq d$.*

Proof. Let $k = \dim(\mathbb{R}^{\mathcal{A}_d})^{\mathsf{G}_d}$ and fix $n \geq d$. Let $\{\alpha_j\}_{j\in[k]} \subseteq \mathcal{A}_n$ be a set of G_n-orbit representatives, and let $\mathbb{1}_{j,n} = \sum_{g\in\mathsf{G}_n/\mathsf{Stab}_{\mathsf{G}_n}(\alpha_j)} e_{g\alpha_j}$ for each $j \in [k]$, so that $\{\mathbb{1}_{j,n}\}_{j\in[k]}$ is a basis for $(\mathbb{R}^{\mathcal{A}_n})^{\mathsf{G}_n} \cong \mathbb{R}^k$. Let $\mathscr{U} = \mathscr{V}^{\oplus 2} \oplus \mathbb{R} = \{\mathbb{U}_n\}$, which contains the relevant relative entropy cones $\{\mathrm{RE}_{\mathcal{A}_n}\}$. Then a basis for $\mathbb{U}_n^{\mathsf{G}_n}$ consists of $\{(\mathbb{1}_{j,n}, 0, 0), (0, \mathbb{1}_{j,n}, 0)\}_{j\in[k]} \cup \{(0, 0, 1)\}$.

If $(\nu, c, t) \in \mathbb{U}_n^{\mathsf{G}_n}$ for $n \geq d$ is expanded as $\nu = \sum_{j=1}^k \hat{\nu}_j \mathbb{1}_{j,n}$ and similarly for c, then

$$\mathrm{RE}_{\mathcal{A}_n}^{\mathsf{G}_n} = \left\{(\nu, c, t) \in \mathbb{U}_n^{\mathsf{G}_n} : \hat{\nu}, \hat{c} \geq 0,\ \sum_{j=1}^k |\mathsf{G}_n/\mathsf{Stab}_{\mathsf{G}_n}(\alpha_j)| \hat{\nu}_j \log\left(\frac{\hat{\nu}_j}{\hat{c}_j}\right) \leq t\right\}.$$

Let $\mathbb{U} = \mathbb{R}^L \oplus \mathbb{R}^L \oplus \mathbb{R}$, define

$$\mathcal{K} = \left\{(\hat{\nu}, \hat{c}, t) \in \mathbb{R}^L \oplus \mathbb{R}^L \oplus \mathbb{R} : \hat{\nu}, \hat{c} \geq 0,\ \sum_{j=1}^k \hat{\nu}_j \log\left(\frac{\hat{\nu}_j}{\hat{c}_j}\right) \leq t\right\},$$

and define $T_n : \mathbb{U} \to \mathbb{U}_n^{\mathsf{G}_n}$ sending $(e_j, 0, 0) \mapsto |\mathsf{G}_n/\mathsf{Stab}_{\mathsf{G}_n}(\alpha_j)|^{-1}(\mathbb{1}_{j,n}, 0, 0)$, sending $(0, e_j, 0) \mapsto |\mathsf{G}_n/\mathsf{Stab}_{\mathsf{G}_n}(\alpha_j)|^{-1}(0, \mathbb{1}_{j,n}, 0)$, and sending $(0, 0, 1) \mapsto (0, 0, 1)$. Then $\mathrm{RE}_{\mathcal{A}_n}^{\mathsf{G}_n} = T_n(\mathcal{K})$ for all $n \geq d$, giving the desired constant-sized descriptions. □

Now suppose $\mathscr{W} = \{\mathbb{R}^{\mathcal{B}_n}\}$ is another permutation $\mathscr{V}_0$-module. Let $\widetilde{\mathscr{U}} = \mathscr{W} \otimes \mathscr{U} = \{\widetilde{\mathbb{U}_n} = \mathcal{L}(\mathbb{R}^{\mathcal{B}_n}, \mathbb{U}_n)\}$ and consider the cones of maps

$$\mathrm{REM}_{\mathcal{A}_n,\mathcal{B}_n} = \left\{M \in \widetilde{\mathbb{U}_n} : M(\mathbb{R}_+^{\mathcal{B}_n}) \subseteq \mathrm{RE}_{\mathcal{A}_n}\right\} = \left\{M \in \widetilde{\mathbb{U}_n} : Me_\beta \in \mathrm{RE}_{\mathcal{A}_n} \text{ for all } \beta \in \mathcal{B}_n\right\}. \tag{2.27}$$

We obtain constant-sized descriptions for these cones for a specific $\mathscr{V}_0$.

Proposition 2.5.5. *Suppose $\mathscr{V}_0 = \{\mathbb{R}^n\}$ with $\mathsf{G}_n = \mathsf{B}_n, \mathsf{D}_n$, or S_n as in Example 2.1.3 and that $\{\mathbb{R}^{\mathcal{A}_n}\}$, $\{\mathbb{R}^{\mathcal{B}_n}\}$ are permutation $\mathscr{V}_0$-modules generated in degrees d_V, d_W, respectively. Then the sequence of cones $\{\mathrm{REM}^{\mathsf{G}_n}_{\mathcal{A}_n,\mathcal{B}_n}\}$ in (2.27) admits constant-sized descriptions for $n \geq d_V + d_W$ if $\mathsf{G}_n = \mathsf{S}_n, \mathsf{B}_n$ and $n \geq d_V + d_W + 1$ if $\mathsf{G}_n = \mathsf{D}_n$.*

Proof. Set $d_0 = 0$ if $\mathsf{G}_n = \mathsf{S}_n, \mathsf{B}_n$ and $d_0 = 1$ if $\mathsf{G}_n = \mathsf{D}_n$. Let $\hat{\mathcal{B}} \subseteq \mathcal{B}_{d_W}$ be a set of minimal-degree $\mathsf{G}_{d_W+d_0}$-orbit representatives in $\mathcal{B}_{d_W+d_0}$, which are also orbit representatives for $\mathcal{B}_n$ for all $n \geq d_W + d_0$ by Proposition 2.2.15(c). Any $M \in \widetilde{\mathbb{U}_n}^{\mathsf{G}_n} = \mathcal{L}(\mathbb{R}^{\mathcal{B}_n}, \mathbb{U}_n)^{\mathsf{G}_n}$ for $n \geq d_W + d_0$ is fully determined by the images $Me_\beta \in \mathbb{U}_n^{\mathrm{Stab}_{\mathsf{G}_n}(\beta)}$ of the basis elements e_β for $\beta \in \hat{\mathcal{B}}$ and conversely, for any collection $\left\{u_\beta \in \mathbb{U}_n^{\mathrm{Stab}_{\mathsf{G}_n}(\beta)}\right\}_{\beta\in\hat{\mathcal{B}}}$ there is a unique $M \in \widetilde{\mathbb{U}_n}^{\mathsf{G}_n}$ satisfying $Me_\beta = u_\beta$. Moreover, $M \in \widetilde{\mathcal{K}_n}$ if and only if $Me_\beta \in \mathcal{K}_n^{\mathrm{Stab}_{\mathsf{G}_n}(\beta)}$ for all $\beta \in \hat{\mathcal{B}}$. Thus, we have

$$\widetilde{\mathbb{U}_n}^{\mathsf{G}_n} = \bigoplus_{\beta\in\hat{\mathcal{B}}} \mathbb{U}_n^{\mathrm{Stab}_{\mathsf{G}_n}(\beta)} \supseteq \bigoplus_{\beta\in\hat{\mathcal{B}}} \mathrm{RE}_{\mathcal{A}_n}^{\mathrm{Stab}_{\mathsf{G}_n}(\beta)} = \mathrm{REM}^{\mathsf{G}_n}_{\mathcal{A}_n,\mathcal{B}_n}.$$

Applying Corollary 2.2.19 to the free module with which $\{\mathbb{R}^{\mathcal{A}_n}\}$ agrees for $n \geq d_V + d_0$ by Proposition 2.2.15(c), and which is presented in the same degree, we conclude that the projections $\mathcal{P}_{\mathbb{V}_n} \colon (\mathbb{R}^{\mathcal{A}_{n+1}})^{\mathrm{Stab}_{\mathsf{G}_{n+1}}(\beta)} \to (\mathbb{R}^{\mathcal{A}_n})^{\mathrm{Stab}_{\mathsf{G}_n}(\beta)}$ are isomorphisms for all $n \geq d_V + d_W + d_0$. Proposition 2.5.4 then gives constant-sized descriptions for $\left\{\mathrm{RE}_{\mathcal{A}_n}^{\mathrm{Stab}_{\mathsf{G}_n}(\beta)}\right\}_n$ for each $\beta \in \hat{\mathcal{B}}$. $\square$

As an application of Proposition 2.5.5, we obtain constant-sized descriptions for SAGE cones of signomials. Indeed, if $\mathcal{A}_n, \mathcal{B}_n \subseteq \mathbb{R}^n$ as in the above proposition, define the sequence $\mathscr{F} = \{\mathbb{F}_n\}$ of functions on $\{\mathbb{R}^n\}$

$$\mathbb{F}_n = \left\{ f(x) = \sum_{\alpha\in\mathcal{A}_n} c_\alpha e^{\langle\alpha,x\rangle} + \sum_{\beta\in\mathcal{B}_n} t_\beta e^{\langle\beta,x\rangle} : c_\alpha, t_\beta \in \mathbb{R} \right\} \cong \mathbb{R}^{\mathcal{A}_n} \oplus \mathbb{R}^{\mathcal{B}_n},$$

with G_n acting by $g \cdot f = f \circ g^{-1}$. Note that $\mathscr{F} = \mathscr{V} \oplus \mathscr{W}$ as consistent sequences, where $\mathscr{V} = \{\mathbb{R}^{\mathcal{A}_n}\}$ and $\mathscr{W} = \{\mathbb{R}^{\mathcal{B}_n}\}$ are as above. Sums of exponentials as in $\mathbb{F}_n$ are called signomials, and optimization problems involving such functions arise in a number of applications [162]. As usual, minimizing a signomial f over $\mathbb{R}^n$ can be recast as maximizing $\gamma \in \mathbb{R}$ such that $f - \gamma \geq 0$ on $\mathbb{R}^n$, so that optimizing signomials can be reduced to certifying their nonnegativity. This is NP-hard in general, but it can be done efficiently if only a single coefficient of f is nonnegative or if f is a sum of such signomials [40]. Formally, define the cones of (Sums of) nonnegative

AM/GM Exponential functions, called AGE (resp., SAGE) functions, by

$$\mathrm{AGE}_{\mathcal{A}_n,\beta} = \left\{ f(x) = \sum_{\alpha\in\mathcal{A}_n} c_\alpha e^{\langle\alpha,x\rangle} + te^{\langle\beta,x\rangle} : f \geq 0 \text{ on } \mathbb{R}^n \text{ and } c_\alpha \geq 0 \text{ for all } \alpha \in \mathcal{A}_n \right\},$$

$$\mathrm{SAGE}_{\mathcal{A}_n,\mathcal{B}_n} = \sum_{\beta\in\mathcal{B}_n} \mathrm{AGE}_{\mathcal{A}_n,\beta}.$$

Theorem 2.5.6. *Suppose $\mathcal{A}_n, \mathcal{B}_n \subseteq \mathbb{R}^n$ where $\mathbb{R}^n$ is embedded in $\mathbb{R}^{n+1}$ by zero-padding and with the standard action of $\mathsf{G}_n = \mathsf{S}_n, \mathsf{D}_n$ or B_n. If $\mathcal{A}_n = \bigcup_{g\in\mathsf{G}_n} g\mathcal{A}_{d_A}$ for all $n \geq d_A$ and $\mathcal{B}_n = \bigcup_{g\in\mathsf{G}_n} g\mathcal{A}_{d_B}$ for all $n \geq d_B$, then the invariant SAGE cones $\{\mathrm{SAGE}^{\mathsf{G}_n}_{\mathcal{A}_n,\mathcal{B}_n}\}_n$ admit constant-sized descriptions for $n \geq d_A + d_B$ if $\mathsf{G}_n = \mathsf{S}_n, \mathsf{B}_n$ and $n \geq d_A + d_B + 1$ if $\mathsf{G}_n = \mathsf{D}_n$.*

Proof. Identify $M \in \widetilde{\mathbb{U}_n}$ with tuples $(Me_\beta)_{\beta\in\mathcal{B}_n} = (\nu^{(\beta)}, c^{(\beta)}, t_\beta)_{\beta\in\mathcal{B}_n} \in \mathbb{R}^{\mathcal{A}_n} \oplus \mathbb{R}^{\mathcal{A}_n} \oplus \mathbb{R}$ for each $\beta \in \mathcal{B}_n$. The authors of [40] show that, in our notation,

$$\begin{aligned}\mathrm{SAGE}_{\mathcal{A}_n,\mathcal{B}_n} = \Big\{&(c,t) \in \mathbb{R}^{\mathcal{A}_n} \oplus \mathbb{R}^{\mathcal{B}_n} : \exists M = (\nu^{(\beta)}, c^{(\beta)}, t_\beta)_{\beta\in\mathcal{B}_n} \in \mathrm{REM}_{\mathcal{A}_n,\mathcal{B}_n} \\ &\text{s.t. } \textstyle\sum_{\beta\in\mathcal{B}_n} c^{(\beta)} = c, \quad \sum_{\alpha\in\mathcal{A}_n} \nu^{(\beta)}_\alpha(\alpha-\beta) = 0 \text{ for all } \beta \in \mathcal{B}_n\Big\} \\ = \ &\pi_n(\mathrm{REM}_{\mathcal{A}_n,\mathcal{B}_n} \cap \mathbb{L}_n),\end{aligned}$$

where

$$\mathbb{L}_n = \left\{ M = (\nu^{(\beta)}, c^{(\beta)}, t_\beta)_{\beta\in\mathcal{B}_n} \in \widetilde{\mathbb{U}_n} : \textstyle\sum_{\alpha\in\mathcal{A}_n} \nu^{(\beta)}_\alpha(\alpha-\beta) = 0 \text{ for all } \beta \in \mathcal{B}_n \right\},$$

$$\pi_n(M) = \left(\textstyle\sum_{\beta\in\mathcal{B}_n} c^{(\beta)}, (t_\beta)_{\beta\in\mathcal{B}_n} \right) \in \mathbb{R}^{\mathcal{A}_n} \oplus \mathbb{R}^{\mathcal{B}_n}.$$

Note that π_n is equivariant, since

$$(g\cdot M)e_\beta = gMg^{-1}e_\beta = gMe_{g^{-1}\beta} = g\cdot(\nu^{(g^{-1}\beta)}, c^{(g^{-1}\beta)}, t_{g^{-1}\beta}) = (g\cdot\nu^{(g^{-1}\beta)}, g\cdot c^{(g^{-1}\beta)}, t_{g^{-1}\beta}),$$

hence

$$\pi_n(g\cdot M) = \left(\sum_{\beta\in\mathcal{B}_n} g\cdot c^{(g^{-1}\beta)}, (t_{g^{-1}\beta})_{\beta\in\mathcal{B}_n} \right) = \left(g \cdot \sum_{\beta\in\mathcal{B}_n} c^{(\beta)}, g\cdot(t_\beta)_{\beta\in\mathcal{B}_n} \right) = g\cdot\pi_n(M).$$

Similarly, if $\sum_{\alpha\in\mathcal{A}_n} \nu^{(\beta)}_\alpha(\alpha-\beta) = 0$ for all $\beta \in \mathcal{B}_n$ then

$$\sum_{\alpha\in\mathcal{A}_n} (g\cdot\nu^{(g^{-1}\beta)})_\alpha(\alpha-\beta) = g\sum_{\alpha\in\mathcal{A}_n} \nu^{(g^{-1}\beta)}_{g^{-1}\alpha}(g^{-1}\alpha-g^{-1}\beta) = g\sum_{\alpha\in\mathcal{A}_n} \nu^{(g^{-1}\beta)}_\alpha(\alpha-g^{-1}\beta) = 0,$$

hence $\mathbb{L}_n$ is G_n-invariant. Thus, the result follows from Proposition 2.5.5 and Proposition 2.5.3. ◻

Theorem 2.5.6 generalizes [160, Thm. 5.3] beyond S_n to the other classical Weyl groups D_n and B_n. It would be interesting to further generalize to signomials defined on more general consistent sequences than $\{\mathbb{R}^n\}$, which would require generalizing the description of the AGE cone from [40].

2.6 Dimension-Free Convex Regression

In this section, we use our framework to describe a solution to the dimension-free convex regression problem. Recall that in this problem we are given evaluation data in different dimensions and we seek a sequence of convex functions that best fit the data and can be instantiated in any desired dimension (including those not represented in the data). To that end, we use the framework we developed in Section 2.3 to obtain parametric families of freely-described convex sets in a fully algorithmic manner. We then present a numerical procedure to fit elements of these families to the given data. Our implementation of the resulting algorithms can be found at `https://github.com/eitangl/anyDimCvxSets`.

Computationally Parametrizing Descriptions

Suppose we seek a parametric family of convex subsets $\{C_n \subseteq \mathbb{V}_n\}$ of a consistent sequence $\mathscr{V} = \{\mathbb{V}_n\}$ of $\mathscr{G} = \{\mathsf{G}_n\}$-representations. Both $\mathscr{V}$ and $\mathscr{G}$ are usually dictated by the application at hand and the symmetries it exhibits, as in Example 2.1.5. We then select description spaces $\mathscr{U} = \{\mathbb{U}_n\}, \mathscr{W} = \{\mathbb{W}_n\}$, usually constructed from $\mathscr{V}$ as described in Section 2.2. Their choice is dictated by the desired richness of the family of the freely-described sets and, as we shall see in Remark 2.6.3 below, by the dimensions of the available data. Once all of these are chosen, parametrizing the associated family of freely-described convex sets amounts to identifying bases for the spaces of freely-described vectors and linear maps appearing in (ConicSeq). Up to this point we have obtained such bases analytically, as the relevant spaces of invariants have been fairly low-dimensional, but this approach becomes impractical for richer description spaces. To address this challenge, we present a computational method for obtaining the relevant bases, which yields a fully algorithmic approach for deriving parametrized families of freely-described convex sets.

Computing a basis for the space of freely-described elements in a finitely-generated consistent sequence proceeds in two steps. First, we compute a basis for the space of invariants in a fixed, sufficiently large dimension. Second, we extend the elements of this basis to any other dimension, which is done by solving a linear system. Our procedure is summarized in Algorithm 1, which we proceed to describe in more detail.

Algorithm 1 Computationally parametrize a freely-described (and possibly compatible) sequence of convex sets.

1: **Input:** Finitely-presented consistent sequences $\mathscr{V}, \mathscr{W}, \mathscr{U}$.
2: **Output:** Bases $\{\mathscr{A}^{(i)} = \{A_n^{(i)}\}_{n\in\mathbb{N}}\}_{i=1}^{d_A}, \{\mathscr{B}^{(j)} = \{B_n^{(j)}\}_{n\in\mathbb{N}}\}_{j=1}^{d_B}$ for freely-described equivariant maps or morphisms.
3: **if** compatibility not required **then**
4:     Fix $n_0 \geq$ presentation degrees of $\mathscr{V} \otimes \mathscr{U}$ and $\mathscr{W} \otimes \mathscr{U}$.
5:     Find bases $\{A_{n_0}^{(i)}\}_{i=1}^{d_A}, \{B_{n_0}^{(j)}\}_{j=1}^{d_B}$ for $\mathcal{L}(\mathbb{V}_{n_0}, \mathbb{U}_{n_0})^{\mathsf{G}_{n_0}}$ and $\mathcal{L}(\mathbb{W}_{n_0}, \mathbb{U}_{n_0})^{\mathsf{G}_{n_0}}$.
6: **else**
7:     Fix $n_0 \geq$ presentation degrees of $\mathscr{V}, \mathscr{W}, \mathscr{U}$.
8:     Find bases $\{A_{n_0}^{(i)}\}_{i=1}^{d_A}, \{B_{n_0}^{(j)}\}_{j=1}^{d_B}$ for subspaces of $\mathcal{L}(\mathbb{V}_{n_0}, \mathbb{U}_{n_0})^{\mathsf{G}_{n_0}}$ and $\mathcal{L}(\mathbb{W}_{n_0}, \mathbb{U}_{n_0})^{\mathsf{G}_{n_0}}$ satisfying the hypotheses of Theorem 2.3.5.
9: **end if**
10: For any $n > n_0$ and each i, j, find unique equivariant $A_n^{(i)}, B_n^{(j)}$ projecting onto $A_{n_0}^{(i)}, B_{n_0}^{(j)}$. For each $n < n_0$, project $A_{n_0}^{(i)}, B_{n_0}^{(j)}$ to dimension n to obtain $A_n^{(i)}, B_n^{(j)}$.

Fix $n_0 \in \mathbb{N}$ as in Algorithm 1, computed for instance using Theorem 2.2.11. We now explain how to find bases for the desired spaces of equivariant linear maps A_{n_0} and B_{n_0} in a fixed dimension, and how to extend them to bases of freely-described elements of $\mathscr{V} \otimes \mathscr{U}$ and $\mathscr{W} \otimes \mathscr{U}$. These elaborate on steps 5, 8 and 10 in Algorithm 1.

Step 5: Computing basis for equivariant maps. We explain how to compute a basis for invariants in a fixed vector space, which we then instantiate in the context of Algorithm 1 to perform step 5. If $\mathbb{V}$ is a representation of a group G, a vector $v \in \mathbb{V}$ is G-invariant iff $g \cdot v = v$ for all $g \in \mathsf{G}$, which can be rewritten as $v \in \ker(g - I)$ for all $g \in \mathsf{G}$. Thus, finding a basis for invariants in a fixed vector space reduces to finding a basis for the kernel of a matrix, though this matrix may be very large or even infinite. We can dramatically reduce the size of this matrix by only considering discrete and continuous generators of G, as proposed in [79].

Theorem 2.6.1 ([79, Thm. 1]). *Let G be a real Lie group with finitely-many connected components acting on a vector space $\mathbb{V}$ via the homomorphism $\rho\colon \mathsf{G} \to \mathrm{GL}(\mathbb{V})$. Let $\{H_i\}$ be a basis for the Lie algebra $\mathfrak{g}$ of G and $\{h_j\} \subseteq \mathsf{G}$ be a finite collection such that G is generated by $\exp(\mathfrak{g})$ and $\{h_j\}$.*[6] *Then*

$$v \in \mathbb{V}^{\mathsf{G}} \iff \mathrm{D}\rho(H_i)v = 0 \text{ and } (\rho(h_j) - \mathrm{id}_{\mathbb{V}}) \cdot v = 0 \quad \text{for all } i, j.$$

[6] Such $\{h_j\}$ exist by [79, Appdx. H].

Here $\mathrm{D}\rho\colon \mathfrak{g} \to \mathcal{L}(\mathbb{V})$ is the differential of ρ. Sets of Lie algebra bases and discrete generators for various standard groups are given in [79]. For example, $\mathsf{G} = \mathsf{S}_n$ is generated by two elements, namely, the transposition $(1,2)$ and the n-cycle $(1,\dots,n)$, reducing the number of group elements that must be considered from the naïve $n!$ to two. For $\mathsf{G} = \mathsf{O}_n$, a basis for the Lie algebra $\mathfrak{g} = \mathrm{Skew}(n)$ is given by $E_{i,j} = e_ie_j^\top - e_je_i^\top$ for $i < j$, and only one discrete generator, e.g., $h_1 = \mathrm{diag}(-1,1,\dots,1)$, is needed, for a total of $\binom{n}{2} + 1$ elements.

As equivariant linear maps are precisely the invariants in the space of linear maps, Theorem 2.6.1 allows us to obtain a basis for equivariant maps between fixed vector spaces. Explicitly, if $\rho_V\colon \mathsf{G}_{n_0} \to \mathrm{GL}(\mathbb{V}_{n_0})$ and $\rho_U\colon \mathsf{G}_{n_0} \to \mathrm{GL}(\mathbb{U}_{n_0})$ are the group homomorphisms defining the actions of G_{n_0} on $\mathbb{V}_{n_0}, \mathbb{U}_{n_0}$, then $A_{n_0} \in \mathcal{L}(\mathbb{V}_{n_0}, \mathbb{U}_{n_0})^{\mathsf{G}_{n_0}}$ if and only if

$$\mathrm{D}\rho_U(H_i)A_{n_0} - A_{n_0}\mathrm{D}\rho_V(H_i) = 0, \quad \rho_U(h_j)A_{n_0} - A_{n_0}\rho_V(h_j) = 0, \quad \text{for all } i,j. \tag{2.28}$$

The equations (2.28) express the space $\mathcal{L}(\mathbb{V}_{n_0}, \mathbb{U}_{n_0})^{\mathsf{G}_{n_0}}$ as the kernel of a matrix, which is often very large and sparse. Forming this matrix only requires specifying the matrices representing $\{H_i\}$ and $\{h_j\}$ for G_{n_0}. Computing a basis for the kernel of this matrix can then by done using its LU decomposition as in [90, 129], or using the algorithm of [79, §5]. A basis for the space $\mathcal{L}(\mathbb{W}_{n_0}, \mathbb{U}_{n_0})^{\mathsf{G}_{n_0}}$ is obtained analogously.

Step 8: Computing a basis for extendable equivariant linear maps. We know from Example 2.3.1 that extending arbitrary invariants $A_{n_0}, B_{n_0}, u_{n_0}$ to freely-described elements does not yield a compatible sequence of convex sets. However, Theorem 2.3.5 identifies subspaces of invariant linear maps A_{n_0}, B_{n_0} whose extensions do yield a compatible sequence of sets, provided we fix a freely-described element $\{u_n\}$ satisfying $u_{n+1} - u_n \in \mathcal{K}_n$. Specifically, we need to find a basis for equivariant linear maps satisfying $A_{n_0}(\mathbb{V}_j) \subseteq \mathbb{U}_j$ for $j \leq d_V$ where d_V is the generation degree of $\mathscr{V}$. This is again a linear condition on A_{n_0}; defining $\varphi_{n_0,j} = \varphi_{n_0-1}\cdots\varphi_j\colon \mathbb{V}_j \hookrightarrow \mathbb{V}_{n_0}$ if $j < n_0$ and $\varphi_{n_0,n_0} = \mathrm{id}_{\mathbb{V}_{n_0}}$, and similarly for $\psi_{n_0,j}$, we have

$$A_{n_0}(\mathbb{V}_j) \subseteq \mathbb{U}_j \iff (I - \mathcal{P}_{\mathbb{U}_j})A_{n_0}|_{\mathbb{V}_j} = 0 \iff (I - \psi_{n_0,j}\psi_{n_0,j}^*)A_{n_0}\varphi_{n_0,j} = 0. \tag{2.29}$$

The subspace of $\mathcal{L}(\mathbb{V}_{n_0}, \mathbb{U}_{n_0})^{\mathsf{G}_{n_0}}$ satisfying the hypotheses of Theorem 2.3.5 is thus again the kernel of a matrix obtained by combining (2.28) and (2.29). To also impose $A_{n_0}(\mathbb{V}_j^\perp) \subseteq \mathbb{U}_j^\perp$ for $j \leq d_U$ where d_U is the generation degree of $\mathscr{U}$, so that $A_{n_0}^*$ extends to a morphism, note that

$$A_{n_0}(\mathbb{V}_j^\perp) \subseteq \mathbb{U}_j^\perp \iff \mathcal{P}_{\mathbb{U}_j} A_{n_0}|_{V_i^\perp} = 0 \iff \psi_{n_0,i}^* A_{n_0}(I - \varphi_{n_0,i}\varphi_{n_0,i}^*) = 0, \quad (2.30)$$

hence the corresponding subspace of $\mathcal{L}(\mathbb{V}_{n_0}, \mathbb{U}_{n_0})^{\mathsf{G}_{n_0}}$ is the kernel of the matrix obtained by combining (2.28)-(2.30). Forming this matrix requires specifying the matrices representing $\{H_i\}$ and $\{h_j\}$ for G_{n_0}, as well as the matrices representing the embeddings $\varphi_{n_0,i}$ and $\psi_{n_0,i}$. The subspace of $\mathcal{L}(\mathbb{W}_{n_0}, \mathbb{U}_{n_0})^{\mathsf{G}_{n_0}}$ satisfying the hypotheses of Theorem 2.3.5 is again the kernel of a matrix and its basis is computed similarly.

Step 10: Extending bases to higher dimensions. Given bases of equivariant $A_{n_0}^{(i)}, B_{n_0}^{(j)}$, we wish to extend them to freely-described elements. We do so computationally by applying a linear map for $n < n_0$ and solving a linear system for each $n > n_0$ to which we wish to extend. For each $n < n_0$, we set $A_n^{(i)} = \psi_{n_0,n}^* A_{n_0}^{(i)} \varphi_{n_0,n}$ and similarly for $B_n^{(j)}$. For each $n > n_0$, we set $A_n^{(i)}$ to be the unique solution to the linear system (2.28) (with n_0 replaced by n) and $\psi_{n,n_0}^* A_n^{(i)} \varphi_{n,n_0} = A_{n_0}^{(i)}$. Forming this linear system requires specifying the matrices representing $\{H_i\}$ and $\{h_j\}$ for G_n, as well as matrices representing φ_{n,n_0} and ψ_{n,n_0}. The above linear system is typically large and sparse, and we solve it using LSQR [170]. The extension of $B_{n_0}^{(j)}$ is handled similarly, except that n_0 needs to exceed the presentation degrees of both $\mathscr{W}$ and $\mathscr{U}$ to guarantee that both $B_{n_0}^{(j)}$ and $(B_{n_0}^{(j)})^*$ extend to morphisms.

Example 2.6.2 (Dimension counts)**.** *We use the above algorithm to obtain dimension counts for parametric families of dimension-free descriptions. See the functions `compute_dims_a`, `compute_dims_b`, and `compute_dims_c` on GitHub for the code computing these dimensions using Algorithm 1.*

(a) Let $\mathscr{V} = \{\mathbb{R}^n\}$ with $\mathsf{G}_n = \mathsf{S}_n$ as in Example 2.1.3, and let $\mathscr{W} = \mathscr{U} = \mathrm{Sym}^2(\mathrm{Sym}^{\leq 2}\mathscr{V})$. Then $\mathscr{V}, \mathscr{U}, \mathscr{V} \otimes \mathscr{U}, \mathscr{W} \otimes \mathscr{U}$ are all $\mathscr{V}$-modules and are presented in degrees $1, 4, 5, 8$, respectively, by Theorem 2.2.11.

The dimensions of invariants parametrizing dimension-free descriptions are $\dim \mathcal{L}(\mathbb{V}_n, \mathbb{U}_n)^{\mathsf{G}_n} = 39$, $\dim \mathcal{L}(\mathbb{W}_n, \mathbb{U}_n)^{\mathsf{G}_n} = 1068$, and $\dim \mathbb{U}_n^{\mathsf{G}_n} = 17$ for $n \geq 8$. The dimensions of linear maps $\{A_n\}$ and $\{B_n\}$ satisfying Proposition 2.3.2(a), yielding intersection-compatible convex sets, are

$$\dim \left\{\{A_n \colon \mathbb{V}_n \to \mathbb{U}_n\} \text{ morphism}\right\} = 6,$$
$$\dim \left\{\{B_n \colon \mathbb{W}_n \to \mathbb{U}_n\} : \text{both } \{B_n\} \text{ and } \{B_n^*\} \text{ morphisms}\right\} = 104.$$

If we further require $\{A_n^*\}$ to be a morphism to obtain projection compatibility, the dimension of $\{A_n\}$ decreases to 5.

(b) Let $\mathscr{V} = \{\mathbb{S}^n\}$ with $\mathsf{G}_n = \mathsf{S}_n$ used to obtain graph parameters in Example 2.1.5, and let $\mathscr{W}, \mathscr{U}$ be as in (a). Then $\mathscr{V}, \mathscr{U}, \mathscr{V} \otimes \mathscr{U}, \mathscr{W} \otimes \mathscr{U}$ are all $\mathscr{V}$-modules and are presented in degrees $2, 4, 6, 8$, respectively.

The dimension of invariant $\{A_n\}$ in this case is $\dim \mathcal{L}(\mathbb{V}_n, \mathbb{U}_n)^{\mathsf{G}_n} = 93$, and the dimensions of $\dim \mathcal{L}(\mathbb{W}_n, \mathbb{U}_n)^{\mathsf{G}_n}$ and $\dim \mathbb{W}_n^{\mathsf{G}_n}$ are as in (a), giving the number of parameters describing freely-described convex sets with these description spaces. The dimensions of linear maps $\{A_n\}$ and $\{B_n\}$ satisfying Proposition 2.3.2(a), and hence describing intersection-compatible sets, are 19 and 104, respectively, as given in (2.18). If we further require $\{A_n^*\}$ to be a morphism to get projection compatibility, the dimension of $\{A_n\}$ decreases to 12.

(c) Let $\mathscr{V} = \{\mathbb{R}^n\}$ with $\mathsf{G}_n = \mathsf{B}_n$ as in Example 2.1.3, used below to learn regularizers defined for vectors of any length, let $\mathscr{V}' = \{\mathbb{R}^{2n+1}\} = \mathscr{W}^{(2)} \oplus \mathbb{R}$ as in Example 2.3.1, and let $\mathscr{W} = \mathscr{U} = \mathrm{Sym}^2(\mathrm{Sym}^{\leq 1}\mathscr{V}')$. Then $\mathscr{V}, \mathscr{U}, \mathscr{V} \otimes \mathscr{U}, \mathscr{W} \otimes \mathscr{U}$ are all $\mathscr{V}$-modules and are presented in degrees $1, 2, 3, 4$, respectively.

The dimensions of invariants parametrizing freely-described convex sets in this case are

$$\dim \mathcal{L}(\mathbb{V}_n, \mathbb{U}_n)^{\mathsf{G}_n} = 4, \quad \dim \mathcal{L}(\mathbb{W}_n, \mathbb{U}_n)^{\mathsf{G}_n} = 108, \quad \dim \mathbb{U}_n^{\mathsf{G}_n} = 8, \quad \textit{for all } n \geq 4.$$

The dimensions of linear maps $\{A_n\}$ and $\{B_n\}$ satisfying Proposition 2.3.2(a) and parametrizing intersection-compatible sets are

$$\begin{aligned} &\dim \Big\{\{A_n \colon \mathbb{V}_n \to \mathbb{U}_n\} \textit{ morphism}\Big\} = 3, \\ &\dim \Big\{\{B_n \colon \mathbb{W}_n \to \mathbb{U}_n\} : \textit{both } \{B_n\} \textit{ and } \{B_n^*\} \textit{ morphisms}\Big\} = 37. \end{aligned} \tag{2.31}$$

If we further require $\{A_n^*\}$ to be a morphism, the dimension does not decrease in this case, so all of these intersection-compatible sets are also projection-compatible.

Fitting Freely-Described Convex Functions to Data

We now present an algorithm for the dimension-free convex regression problem from Section 2.1. Recall that in this problem we are given data $\{(x_i, \phi_i) \in \mathbb{V}_{n_i} \oplus \mathbb{R}\}$ in finitely-many dimensions n_i corresponding to a sequence of vector spaces $\{\mathbb{V}_n\}$, and our objective is to identify a sequence of convex functions $\{f_n : \mathbb{V}_n \to \mathbb{R}\}$ such that $f_{n_i}(x_i) \approx \phi_i$ in the dimensions in which data is available. We tackle this problem by endowing $\{\mathbb{V}_n\}$ with the structure of a consistent sequence of $\{\mathsf{G}_n\}$-representations, and choosing description spaces $\{\mathbb{W}_n\}, \{\mathbb{U}_n\}$ that yield a finitely-parametrized family of freely-described and compatible convex functions which we fit to the given data. We explain how to perform this fitting below in two key steps.

The first step is to define a finitely-parametrized family of freely-described convex functions. Let $\{\{A_n^{(i)}\}\}_{i=1}^{d_A}$ and $\{\{B_n^{(j)}\}\}_{j=1}^{d_B}$ be the bases for freely-described maps computed by Algorithm 1 where we choose $n_0 \geq \max_i n_i$ so that we do not have to extend the basis elements computed there to access the data dimensions. Further, we select intersection and projection compatible cones $\mathscr{K} = \{\mathcal{K}_n \subseteq \mathbb{U}_n\}$, and freely-described $\{u_n \in \mathbb{U}_n^{\mathsf{G}_n}\}$ satisfying $u_{n+1} - u_n \in \mathcal{K}_n$ and $u_n \in \mathrm{int}(\mathcal{K}_n)$ for all n. We consider families of nonnegative and positively-homogeneous convex functions f_n parametrized by $\alpha \in \mathbb{R}^{d_A}$, $\beta \in \mathbb{R}^{d_B}$, and $\lambda \in \mathbb{R}$ of the form

$$f_n(x;\alpha,\beta,\lambda) = \inf_{\substack{t\geq 0\\ y\in\mathbb{W}_n}} t + \lambda\|y\| \quad \text{s.t.} \sum_{i=1}^{d_A}\alpha_i A_n^{(i)}x + \sum_{j=1}^{d_B}\beta_j B_n^{(j)}y + tu_n \in \mathcal{K}_n \tag{P}$$

$$= \sup_{z\in\mathcal{K}_n^*} -\sum_{i=1}^{d_A}\alpha_i\langle z, A_n^{(i)}x\rangle \quad \text{s.t.} \left\|\sum_{j=1}^{d_B}\beta_j (B_n^{(j)})^* z\right\| \leq \lambda,\ \langle z, u_n\rangle \leq 1, \tag{D}$$

where the primal and dual programs above are equal since Slater's condition is satisfied by our choice of u_n. Here $\|\cdot\|$ is the norm on $\mathbb{W}_n$ induced by its inner-product and the parameter λ is chosen to be positive; the purpose of this term in (P) is to prevent numerical issues arising in our subsequent regression procedure (see (Regress)). Note that this is indeed a finitely-parametrized (by $d_A + d_B + 1$ parameters) infinite sequence of functions. Also note that if we require compatibility in Algorithm 1, then the sequence $\{f_n\}$ is intersection (and if desired, projection) compatible for any value of the parameters, since it is the sequence of gauge functions of a correspondingly compatible sequence of sets (see Section 2.1).

The second step concerns optimizing over the parameters α, β, λ to fit $\{f_n\}$ to the available data. To this end, we consider the following optimization problem, with a user-specified $\lambda_{\min} > 0$:

$$\min_{\substack{\varepsilon \in \mathbb{R}^D_+ \\ \alpha \in \mathbb{R}^{d_A}, \beta \in \mathbb{R}^{d_B}, \lambda \geq \lambda_{\min} \\ \{(t_i, y_i)\}, \{z_i\}}} \quad \|\varepsilon\|_{\ell_2} \quad \text{s.t.} \tag{Regress}$$

$$(y_i, t_i) \text{ feasible for (P) with } n = n_i \text{ and cost } \leq \phi_i + \varepsilon_i, \tag{PC}$$

$$z_i \text{ feasible for (D) with } n = n_i \text{ and cost } \geq \phi_i - \varepsilon_i, . \tag{DC}$$

The constraints (PC) and (DC) are required to hold for all $i \in [D]$ and they ensure that $\phi_i - \varepsilon_i \leq f_{n_i}(x_i; \alpha, \beta, \lambda) \leq \phi_i + \varepsilon_i$, so that minimizing $\|\varepsilon\|_{\ell_2}$ yields a good fit to the data. We emphasize that this problem is finite-dimensional even though the it yields an infinite sequence of convex functions $\{f_n\}$.

As (Regress) involves bilinear constraints, we solve the problem via *alternating minimization*, where we alternate between fixing α, β, λ and $\{(t_i, y_i)\}, \{z_i\}$ while optimizing over the rest of the variables. Note that Slater's condition holds in (Regress) for both steps of alternating minimization when $u_n \in \mathrm{int}(\mathcal{K}_n)$.

Remark 2.6.3. *The quantity n_0 in Algorithm 1 is governed by the choice of description spaces and leads to a tradeoff between the richness of the parametric family and the dimensions of the available data. The data we are given only contains information about $\{f_n\}_{n \leq \max_i n_i}$ which, in turn, only depends on $\{\{A_n^{(i)}\}_{n \leq \max_i n_i}\}_{i=1}^{d_A}$ and $\{\{B_n^{(j)}\}_{n \leq \max_i n_i}\}_{j=1}^{d_B}$. If n_0 in Algorithm 1 is strictly larger than the maximum dimension $\max_i n_i$ in which data is available, then the number of distinct basis elements for $n \leq \max_i n_i$ might be strictly smaller than d_A and d_B, respectively, in which case our parameters α, β are not identifiable from such low-dimensional data. Imposing compatibility decreases the bound on n_0, thus facilitating dimension-free convex regression from lower-dimensional data. More broadly, even when $\max_i n_i \geq n_0$, it is of interest to investigate the landscape of* (Regress).

Numerical Results

In all experiments below, we use $\mathscr{U} = \mathrm{Sym}^2(\mathrm{Sym}^{\leq k}\mathscr{V}')$ for some $\mathscr{V}$-module $\mathscr{V}'$ and some k, with the corresponding PSD cones $\mathscr{K} = \mathrm{Sym}^2_+(\mathrm{Sym}^{\leq k}\mathscr{V}')$. We choose $\{u_n\}$ to be the sequence of identity matrices, which satisfy $u_n \in \mathrm{int}(\mathcal{K}_n)$ and $u_{n+1} - u_n \in \mathcal{K}_n$ as required. We apply our algorithm to learn semidefinite approximations of two non-SDP-representale functions, comparing the results obtained with and without imposing compatibility in Algorithm 1.

The first function we approximate is the ℓ_π norm $\|x\|_\pi = (\sum_i |x_i|^\pi)^{1/\pi}$, which is not SDP-representable because π is irrational. We view the ℓ_π norm as defined on the sequence $\mathscr{V} = \{\mathbb{R}^n\}$ with $\mathsf{G}_n = \mathsf{B}_n$ from Example 2.1.3. It satisfies both intersection and projection compatibility. We choose description spaces $\mathscr{W} = \mathscr{U} = \{\mathrm{Sym}^2(\mathrm{Sym}^{\leq 1}\mathbb{R}^{2n+1}) = \mathbb{S}^{2n+2}\}$ as in Example 2.6.2(c), with the corresponding PSD cones $\{\mathcal{K}_n = \mathbb{S}^{2n+2}_+\}$. We used 50 data points in $\mathbb{R}^n$ for $n_i \in \{1, 2\}$. When we impose compatibility on our family of functions, we use $n_0 = 2 = \max_i n_i$ in Algorithm 1. If we do not impose compatibility and search over all freely-described (possibly incompatible) sequences, we take $n_0 = 4$ which is the presentation degree of $\mathscr{W} \otimes \mathscr{U}$, and which strictly exceeds $\max_i n_i$. The constraints in the fitting program (Regress) do not depend on some of the entries of α, β in this case, and we set such entries to zero. This highlights the advantage of imposing compatibility mentioned in Remark 2.6.3–it allows us to uniquely identify a dimension-free description from lower-dimensional data.

The second sequence of functions we approximate is the nonnegative and positively-homogeneous variant of the quantum entropy given by (2.2) given in Section 2.1, defined on the sequence $\{\mathbb{S}^n\}$ with embeddings by zero-padding and the action of $\mathsf{G}_n = \mathsf{O}_n$ by conjugation. Once again, the function (2.2) cannot be evaluated using semidefinite programming, though a family of semidefinite approximations is analytically derived in [74]. Here we aim to learn a semidefinite approximation entirely from evaluation data. To that end, we choose description spaces $\mathscr{W} = \{\mathbb{W}_n = \mathrm{Sym}^2(\mathrm{Sym}^{\leq 1}\mathbb{R}^n) = \mathbb{S}^{n+1}\}$ and $\mathscr{U} = \{\mathbb{U}_n = \mathrm{Sym}^2(\mathrm{Sym}^{\leq 2}\mathbb{R}^n) = \mathbb{S}^{\binom{n+2}{2}}\}$, with corresponding PSD cones $\{\mathcal{K}_n = \mathbb{S}_+^{\binom{n+2}{2}}\}$. Our data consists of 200 PSD matrices in $\mathbb{S}^n$ for $n \leq n_0 = 4$. Without a calculus for presentation degrees for $\mathsf{G}_n = \mathsf{O}_n$, our theory does not guarantee the existence of an O_n-invariant extension of our learned description. Our theory does however guarantee a unique B_n-invariant extension, and in practice we observe that the extension is, in fact, O_n-invariant.

To approximate the above functions, which we generically denote $\{f_n^{\mathrm{true}}\}$, we used (Regress) with 100 random initializations to fit the data in degree n_0. For the above two examples, not only is $f_{n_0}^{\mathrm{true}}$ positively-homogeneous and nonnegative, but also $f_{n_0}^{\mathrm{true}}(x) \neq 0$ for $x \neq 0$ in the domain. We therefore normalize the data x_i by $x_i \mapsto x_i / f_{n_0}^{\mathrm{true}}(x_i)$, so that $f_{n_0}^{\mathrm{true}}(x_i) = 1$ for all i and all points contribute equally to the objective of (Regress). We impose $\lambda \geq \lambda_{\min} = 10^{-3}$ in (Regress). To evaluate the results, we extended our learned descriptions to $n = 20$, sampled 10^3 unit-norm points (also PSD for the quantum entropy example) and computed the average normalized errors $\frac{|f_n(x) - f_n^{\mathrm{true}}(x)|}{f_n^{\mathrm{true}}(x)}$ in each n up to 20.

The resulting errors are plotted in Figure 2.1, shown and discussed in Section 2.1.

Since imposing compatibility conditions decreases the search space in (Regress), we expect the optimal solution of (Regress) with compatibility conditions to exhibit larger errors in dimensions in which data is available compared to the optimal freely-described (but possibly incompatible) solution. That is not the case in Figure 2.1(b), illustrating the nonconvexity of the fitting problem (Regress) and demonstrating another advantage of imposing compatibility—the resulting smaller parametric family allows our algorithm to better fit the data.

2.7 Discussion

We developed a systematic framework to study convex sets that can be instantiated in different dimensions using representation stability, as well as a computational method to parametrize such sets and fit them to data. We did so by formally defining dimension-free descriptions of convex sets and compatibility conditions relating sets in different dimensions. We then proved a number of structural results pertaining to dimension-free descriptions, namely, characterizing such descriptions certifying compatibility; giving conditions on fixed-dimensional descriptions ensuring their extendability to dimension-free ones; and studying infinite-dimensional limits of freely-described sequences of sets. We further used representation stability to systematically derive constant-sized descriptions for sequences of invariant sections of PSD and relative entropy cones. Finally, we developed an algorithm to computationally parametrize and search over dimension-free descriptions to fit them to data. Our work can be viewed as identifying and exploiting a new point of contact between representation stability and convex geometry through conic descriptions of convex sets.

Our work suggests questions and directions for future research in several areas.

(Computational algebra) Is there an algorithm to compute the generation and presentation degrees of a given consistent sequence?

(Lie groups) Can we extend our calculus for presentation degrees in Theorem 2.2.11 to Lie groups such as $\mathsf{G}_n = \mathsf{O}_n$?

(Constructing descriptions) Given a sequence of convex sets instantiable in any relevant dimension, can we systematically construct freely-described, and possibly compatible, approximations for the sequence? When are approximations derived from sums-of-squares hierarchies such as [75] dimension-free and certify compatibility?

(Complexity) Is there a systematic framework to study the smallest possible size of a dimension-free description for a given sequence of sets, extending the slack operator-based approach for fixed convex sets [91]?

(Dimension-free separation) Under what conditions can a point outside a compatible sequence of convex sets be separated by a freely-described sequence, generalizing the Effros–Winkler theorem [66]?

(Statistical inference) How much data do we need to learn a given sequence of sets or functions, and in what dimensions should this data lie?

Chapter 3

ANY-DIMENSIONAL EQUIVARIANT NEURAL NETWORKS

This chapter is based on [141].

3.1 Introduction

In order to train neural networks that accept inputs in any dimension, we must address three key challenges. First, how do we encode infinite sequences of neural networks, one in each dimension, using only finitely-many parameters? Doing so is necessary if we want to learn from a finite amount of data. Second, how do we ensure that the networks we learn from data in a fixed dimension generalize well to higher dimensions? Third, is there a user-friendly procedure for learning these networks? We proceed to tackle these challenges.

Free equivariant neural networks. Equivariance is the key ingredient in tackling the first challenge. The authors of [152] showed that the dimension of permutation-equivariant linear layers between tensors is independent of the size of the tensors, and gave free bases for such linear layers. For example, the space of permutation-equivariant linear layers for graph adjacency matrices is 15-dimensional for any n, and its basis contains the identity, taking a transpose and extracting the diagonal, among others. The authors of that paper use this fact to derive finite parameterizations of linear layers

$$W = \alpha_1 A_1^{(n)} + \cdots + \alpha_{15} A_{15}^{(n)} \tag{3.1}$$

where $\left\{A_1^{(n)}, \ldots, A_{15}^{(n)}\right\}$ is a basis of free equivariant maps for graphs on n nodes. This parameterization allows them to instantiate linear layers in any dimension using only the fifteen parameters α_i, which can be learned from data in a fixed dimension. Our **first contribution** is showing that this is not a coincidence but rather stems from a general, recently-identified phenomenon known as *representation stability*. We utilize this phenomenon to show that the dimension of equivariant linear layers stabilizes for a number of group actions, including those induced by (signed) permutations, the orthogonal groups $\mathrm{O}(n)$, rotations $\mathrm{SO}(n)$, and the Lorentz groups $\mathrm{O}(1, n)$. We leverage this observation to derive finite parametrizations of infinite sequences of equivariant neural networks.

Generalization across dimensions. The authors of [152] observed that neural networks obtained using (3.1) do not always generalize well to dimensions different from the one used for training. Intuitively, this is due to overparametrization, as there are many ways to represent the same function in a fixed dimension, but not all of them extend correctly to other dimensions. Our **second contribution** is to introduce a compatibility condition relating the mappings in different dimensions, which has a regularization effect that often leads to better generalization. Formally, it entails to the commutativity of the diagrams in Figure 3.1. We further explain how to impose this condition on the network architecture.

Computational recipe. The authors of [152] *manually* found the free basis elements in (3.1). However, for more complicated groups and spaces of linear layers, manually finding a free basis becomes prohibitive. In a fixed dimension, one can circumvent this issue using an algorithm proposed in [79] to compute a basis. For our **third contribution**, we extend this basis to any other dimension by solving a sparse linear system, enabling us to obtain free bases computationally.

We provide proof-of-concept experiments demonstrating the ability of our approach to learn functions defined for any dimension. We emphasize, however, that the primary contribution of this chapter is its conceptual novelty. To our knowledge, our method is the first *black-box approach* for training neural networks that can handle inputs in any dimension. Instead of customizing architectures for specific domains and data types, as has been done in the prior work discussed below, our strategy adapts itself to the application domain by selecting appropriate group actions and nested sequences of vector spaces. The downside of the more general approach is that we do not expect it to outperform application-specific methods. Thus, we envision our framework serving as an initial step when approaching novel and poorly understad problems, or serving as a baseline to evaluate the performance of application-specific architectures.

Notation. Given two finite-dimensional vector spaces U and V we use $\mathcal{L}(V,U)$ to denote the set of linear mappings between them. For a group G acting on both U and V, we denote invariant elements as $U^G = \{u \in U \mid g \cdot u = u\}$, and equivariant linear maps as $\mathcal{L}(V,U)^G = \left\{A \in \mathcal{L}(V,U) \mid gAg^{-1} = A\right\}$. The map $\mathcal{P}_W \colon U \to U$ denotes the projection onto a subspace $W \subseteq U$. The symbols $\mathbb{1}_n$ and I_n represent the all-ones vector and the identity in dimension n, respectively. The symbol $\mathbb{R}[G]$ denotes the set of finite linear combinations of elements in G. To avoid cluttering notation, we use bold font symbols whenever possible to denote the nth element in a sequence, e.g., the symbol $\boldsymbol{G}$ denotes $G^{(n)}$. We add a "+" superscript to denote the $(n+1)$th element, e.g., $G^{(n+1)}$ is denoted by $\boldsymbol{G}^+$.

Outline. The remainder of this section focuses on related work. Section 3.2 formally defines free neural networks. Section 3.3 introduces a compatibility condition that ensures good generalization across different dimensions. In Section 3.4, we describe our computational recipe to learn free neural networks from data in a fixed dimension and extend to other dimensions. Section 3.5 provides numerical experiments. We close the chapter in Section 2.7 with conclusions, limitations, and future work. We defer several technical details to the appendix.

Related work

Equivariant learning. The benefits of equivariant architectures first became apparent with the success of convolutional neural networks (CNNs) in computer vision [132]. Since then, equivariance has been applied to a range of applications. Examples include DeepSets [233] and graph neural networks [152, 200] using permutation equivariance to process sets and graphs; AlphaFold 2 [111] and ARES [212] using SE(3)-equivariance for protein and RNA structure prediction and assessment; steerable [49] and spherical [127] CNNs using rotation-equivariance to classify images; and physics-informed neural networks are equivariant under the symmetries of the corresponding physical systems [117]. Many of these architectures have been shown to implement a generalized notion of convolution over the groups at hand [48, 128]. Under additional assumptions, equivariant architectures have been derived using invariant theory to explicitly parametrize polynomial equivariant maps [222]. We refer the reader to [27, 144] for an introduction to equivariant deep learning.

Dimension-free learning. Certain architectures processing data of a particular type are defined for inputs of different sizes. Many architectures processing graph-based data update the features at each vertex by applying the same function of the features of the vertex's neighbors, hence can be applied to graphs of any size [87, 200, 219]. Convolutional neural networks (CNNs) processing signals and images convolve their inputs with filters of constant size, hence can also be applied to inputs of different sizes. Nevertheless, it has been observed that naïvely applying CNNs to inputs of size different from that used during training leads to artifacts, and several downsampling techniques have been proposed to resize inputs to a CNN [96]. Networks processing natural language embed words as vectors of the same length but are defined for arbitrarily long sequences of such vectors. Recurrent neural networks process such sequences one-by-one, using cycles in their architectures to sequentially combine

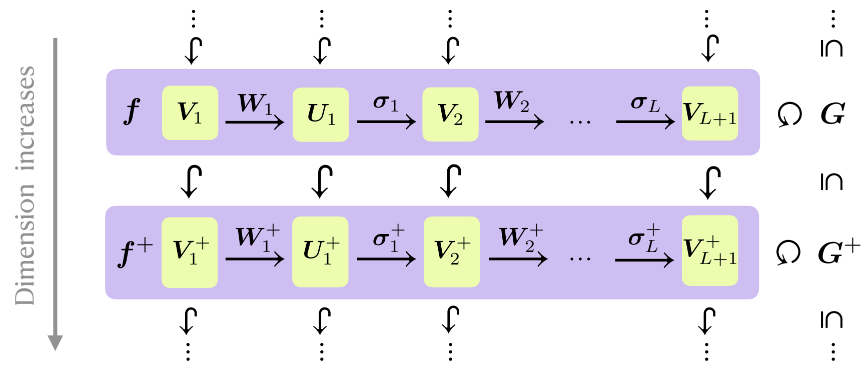


Figure 3.1: Equivariant free neural networks. We use bold font to denote the nth element in a sequence; see Notation in Section 3.1 for details.

a new input vector with a function of the previous inputs [156]. Recursive neural networks apply the same weights recursively to pairs of input vectors to combine them into one, until the entire sequence is reduced to one vector [206]. Notably, all these architectures must process the input in particular ways motivated by the specific application at hand, whereas we only need to assume that the network processes its input equivariantly.

Representation stability. Representation stability considers nested sequences of groups and their representation, and implies that such sequences of representations often stabilize. Specifically, there is a labelling of the irreducibles of these groups such that the decomposition of the representations in the sequence contain the same irreducibles with the same multiplicities. This phenomenon has been formalized in [47] and further studied in [46, 84, 194, 195, 197, 227]. It has been applied to study free convex sets and algebraic varieties [8, 44, 64, 139, 215], though to our knowledge, this chapter is the first to apply it to equivariant deep learning. We refer the reader to [70, 196, 228] for introductions to this area.

3.2 Free Neural Networks

In this section, we introduce the concept of free neural networks, i.e., networks that can be instantiated in every dimension. To set the stage, let us recall the classical notion of a neural network (NN). A NN is a mapping $f = f_L \circ \ldots \circ f_1$ where $f_i \colon V_i \to V_{i+1}$ is a composition $f_i(x) = \sigma_i(W_i x + b_i)$ of an affine map $x \mapsto W_i x + b_i$ and an activation map $\sigma_i \colon U_i \to V_{i+1}$. This yields a family of mappings parametrized by the weights $W_i \in \mathcal{L}(V_i, U_i)$ and biases $b_i \in U_i$. In turn, these parameters are chosen to minimize a prescribed loss function.

In several settings, we want f to respect symmetries in its inputs. Formally, we require f to be equivariant with respect to the action of a group G, i.e., $f \circ g = g \circ f$ for all $g \in G$. A natural way to obtain equivariance is by ensuring that each building block of f is equivariant, i.e., $W_i \in \mathcal{L}(V_i, U_i)^G$, $b_i \in U_i^G$, and σ_i is G-equivariant. NNs satisfying these properties are called *equivariant*. Equivariant NNs can be trained by computing a basis for the space of weights and biases and optimizing over the coefficients in this basis [79].

In this work, we seek equivariant NNs that extend to inputs in *any dimension*. To this end, we consider sequences $\{\boldsymbol{U}_i\}$ and $\{\boldsymbol{V}_i\}$ of nested[1] vector spaces $\boldsymbol{U}_i \subseteq \boldsymbol{U}_i^+$ and $\boldsymbol{V}_i \subseteq \boldsymbol{V}_i^+$, with actions of a nested sequence of groups $\{\boldsymbol{G}\}$, see Figure 3.1. For example, we have inclusions $\mathbb{R}^n \subseteq \mathbb{R}^{n+1}$ by zero-padding on which the group of permutations on n letters $\boldsymbol{G} = \mathrm{S}_n$ acts by permuting coordinates. These define a sequence of NNs whose weights $\boldsymbol{W}_i \colon \boldsymbol{V}_i \to \boldsymbol{U}_i$ and biases $\boldsymbol{b}_i \in \boldsymbol{U}_i$ increase in size. The activation functions $\boldsymbol{\sigma}$ are often defined for every dimension, as the next section shows. Thus, we focus here on extending the linear layers.

Free equivariant networks. Because the dimensions of the space of linear layers $\mathcal{L}(\boldsymbol{U}_i, \boldsymbol{V}_i)$ and the vector spaces $\boldsymbol{U}_i$ usually grow, sequences of *general* NNs are not finitely-parameterizable and cannot be learned from a finite amount of data. The situation radically simplifies when considering equivariant networks since the dimensions of $\mathcal{L}(\boldsymbol{U}_i, \boldsymbol{V}_i)^{\boldsymbol{G}}$ and $\boldsymbol{U}_i^{\boldsymbol{G}}$ often stabilize. This was previously observed in the context of simple graph NNs by [152], and follows from a general phenomenon known as representation stability [47]. To illustrate it with a concrete example, consider the nested sequence $\boldsymbol{U} = \boldsymbol{V} = \mathbb{R}^n$ with the action of $\boldsymbol{G} = \mathrm{S}_n$ as above. Then $\boldsymbol{V}^{\boldsymbol{G}}$ is one-dimensional and spanned by $\mathbb{1}_n$ while $\mathcal{L}(\boldsymbol{U}, \boldsymbol{V})^{\boldsymbol{G}}$ is two-dimensional and spanned by $I_n, \mathbb{1}_n\mathbb{1}_n^\top$, which are basis elements defined for every n. Similarly, [152] obtained explicit free bases for spaces of invariant tensors.

Interestingly, all the above basis elements project onto each other, e.g., the orthgonal projection of $\mathbb{1}_n$ onto $\mathbb{R}^{n-1}$ is $\mathbb{1}_{n-1}$, and similarly $I_n, \mathbb{1}_n\mathbb{1}_n^\top$ project onto $I_{n-1}, \mathbb{1}_{n-1}\mathbb{1}_{n-1}^\top$. Motivated by this observation, we say that a sequence of equivariant NNs is *free* if the weights $\boldsymbol{W}_i \colon \boldsymbol{V}_i \to \boldsymbol{U}_i$ and biases $\boldsymbol{b}_i \in \boldsymbol{U}_i$ satisfy

$$\boldsymbol{W}_i = \mathcal{P}_{\boldsymbol{U}_i}\, \boldsymbol{W}_i^+\big|_{\boldsymbol{V}_i}, \qquad \text{and} \qquad \boldsymbol{b}_i = \mathcal{P}_{\boldsymbol{U}_i}\boldsymbol{b}_i^+. \tag{FreeNN}$$

For many sequences $\{\boldsymbol{V}_i\}, \{\boldsymbol{U}_i\}$, equation (FreeNN) uniquely determines $\boldsymbol{W}_i^+, \boldsymbol{b}_i^+$ from $\boldsymbol{W}_i, \boldsymbol{b}_i$, allowing us to uniquely extend a network to accept larger inputs.

[1]Formally, we have embeddings $\boldsymbol{V} \hookrightarrow \boldsymbol{V}^+$ and $\boldsymbol{U} \hookrightarrow \boldsymbol{U}^+$.

Representation stability. To understand for which sequences (FreeNN) allows us to extend, we recall some key definitions and results from the representation stability literature [46, 47], presented in Section 2.2. Specifically, we fix a sequence of compact groups $\mathscr{G} = \{\boldsymbol{G}\}_{n\in\mathbb{N}}$ such that $\boldsymbol{G} \subseteq \boldsymbol{G}^+$ and assume that $\mathscr{V} = \{(\boldsymbol{V}, \boldsymbol{\varphi})\}_{n\in\mathbb{N}}$ is a consistent sequence of $\mathscr{G}$-representations. As before, we will identify $\boldsymbol{V}$ with its image $\boldsymbol{\varphi}(\boldsymbol{V}) \subseteq \boldsymbol{V}^+$, and omit the inclusions $\boldsymbol{\varphi}$ unless needed. We recall from Proposition 2.2.3 that if $\mathscr{V}$ is a finitely-generated consistent sequence, then projections become isomorphisms of invariants for all sufficiently-large dimensions. More precisely, Proposition 2.2.9 shows that the projections between invariants become isomorphisms for all n exceeding the *presentation degree* of the sequence. Note that (FreeNN) precisely requires $\boldsymbol{W}^+$ and $\boldsymbol{b}^+$ to project to $\boldsymbol{W}, \boldsymbol{b}$ inside the consistent sequences $\mathcal{L}(\mathscr{V}_i, \mathscr{U}_i), \mathscr{U}_i$, respectively. Thus, for all large n we can parameterize infinite sequences of equivariant linear layers with a finite set of parameters $\alpha_1, \ldots, \alpha_\ell$, just as in (3.1). This result enables us to train free NNs as we describe in Section 3.4.

Examples. We collect a few examples.

Scalar sequence. The sequence $\mathscr{S} = \{\mathbb{R}\}$ together with $\boldsymbol{\varphi}(x) = x$ yields a consistent sequence for any sequence of groups acting trivially, i.e., $g \cdot x = x$. It is generated in degree one. By setting the last layer $\boldsymbol{V}_{L+1} = \mathbb{R}$ in the free NN architecture of Figure 3.1, we obtain a sequence of invariant functions $\boldsymbol{f}$.

Permutation sequences. Let $\mathscr{R} = \{\mathbb{R}^n\}$ be the consistent sequence with zero-padding embeddings $\boldsymbol{\varphi}(x) = (x^\top, 0)^\top$ and the action of the symmetric group $\boldsymbol{G} = \mathrm{S}_n$ by permuting coordinates. Consider $\mathscr{V}_k = \mathscr{U}_{k-1} = \mathscr{R}^{\otimes k}$, which is generated in degree k. The dimensions of the space of weights and biases for low-order tensors are eventually

$$\begin{aligned} \dim \mathcal{L}(\boldsymbol{V}_1, \boldsymbol{U}_1)^{\boldsymbol{G}} &= 4, \qquad & \dim \boldsymbol{U}_1^{\boldsymbol{G}} &= 2,\\ \dim \mathcal{L}(\boldsymbol{V}_2, \boldsymbol{U}_2)^{\boldsymbol{G}} &= 52, \qquad & \dim \boldsymbol{U}_2^{\boldsymbol{G}} &= 5. \end{aligned}$$

Signed permutation sequences. Let B_n (resp. D_n) be the signed permutation group (resp. even-signed permutation group) in n elements. Consider the sequence of representations $\boldsymbol{V} = \mathbb{R}^n$ with the standard action of permuting and flipping the signs of the coordinates. Then, $\mathscr{V} = \{(\boldsymbol{V}, \varphi)\}$ is a consistent sequence with generation and presentation degrees equal to one.

Orthogonal rotation sequences. Let $O(n)$ (resp., $SO(n)$) be the orthogonal group (resp., special orthogonal group) in dimension n. Consider the sequence of representations $\boldsymbol{V} = \mathbb{R}^n$ with the standard rotation action. Then, $\mathscr{V} = \{(\boldsymbol{V}, \varphi)\}$ is again a consistent sequence with generation degree equal to one. Characterizing the presentation degree is an open question. Yet, numerically, the presentation degree for S_n seems to work well.

Lorentz sequences. Let $O(1, n)$ be the Lorentz group (resp., special orthogonal group) in dimension $n + 1$. Again, consider the sequence of representations $\boldsymbol{V} = \mathbb{R}^{n+1}$ with the natural action of the Lorentz group. Then, $\mathscr{V} = \{(\boldsymbol{V}, \varphi)\}$ is once more a consistent sequence with generation degree equal to one.

Graph(on) sequences. There are two ways to identify small graphs with larger ones, namely, appending isolated vertices and identifying graphs with their associated graphons. These two embeddings yield two consistent sequences. For the first embedding, let $\boldsymbol{V} = \mathbb{S}^n$ with the Frobenius inner product and embeddings $\boldsymbol{\varphi}(X) = \mathrm{blkdiag}(X, 0)$ padding X by a zero row and zero column. We endow $\boldsymbol{V}$ with the action of $\boldsymbol{G} = \mathrm{S}_n$ by simultaneously permuting rows and columns (so $g \cdot X = gXg^\top$ for all $g \in \boldsymbol{G}$ and $X \in \boldsymbol{V}$). This sequence is generated in degree 2. For the second embedding, let $\boldsymbol{V} = \mathbb{S}^{2^n}$ with the normalized inner product $\langle X, Y\rangle = 2^{-2n}\mathrm{Tr}(XY)$, embeddings $\boldsymbol{\varphi}(X) = X \otimes \mathbb{1}_{2\times 2}$, and the same action of the symmetric group $\boldsymbol{G} = \mathrm{S}_{2^n}$ as above. This sequence arises in the theory of graphons [146], where adjanceny matrices X are associated with a step function (the graphon) $W_X\colon [0,1]^2 \to \mathbb{R}$ in such a way that X and $X \otimes \mathbb{1}_{2\times 2}$ represent the same graphon and $\langle X, Y\rangle = \int_{[0,1]^2} W_X(x,y)W_Y(x,y)\,dx\,dy$. The graphon consistent sequence is generated in degree 2 by [139, Prop. 3.1].

3.3 Generalization Across Dimensions and Compatibility

Free NNs often do not generalize correctly to other dimensions. In Section 3.5, we provide examples where a free NN yields excellent test error in its trained dimension while exhibiting errors 10^{10} times worst in other dimensions. This discrepancy arises due to overparametrization, as there are numerous ways to encode the same function in a fixed dimension, and yet not every encoding extends seamlessly. In this section, we propose a regularization strategy that often leads to better generalization across dimensions.

Specifically, inspired by a similar condition for free convex sets [139], we introduce the following compatibility condition.

Definition 3.3.1 (Compatible networks)**.** *A sequence of maps* $\{\boldsymbol{f}: \boldsymbol{V} \to \boldsymbol{U}\}$ *is (intersection-)* compatible *if* $\boldsymbol{f}^+|_{\boldsymbol{V}} = \boldsymbol{f}$*. A sequence of equivariant networks (Figure 3.1) is compatible if*

$$\boldsymbol{W}_i^+|_{\boldsymbol{V}_i} = \boldsymbol{W}_i, \quad \boldsymbol{b}_i^+ = \boldsymbol{b}_i, \quad \textit{and} \quad \boldsymbol{\sigma}_i^+|_{\boldsymbol{U}_i} = \boldsymbol{\sigma}_i. \tag{CompNN}$$

Intuitively, this condition ensures that applying a function in a fixed dimension and then extending to a higher dimension is equivalent to first extending and then applying the function. Layer-wise compatibility (CompNN) guarantees that the sequence of NN $\{\boldsymbol{f}: \boldsymbol{V}_1 \to \boldsymbol{V}_{L+1}\}$, given by Figure 3.1, is compatible, or equivalently, that the diagram in that figure commutes. The condition (CompNN) is strictly stronger than (FreeNN), as compatible NNs are free, but free NNs are not compatible in general.

Compatibility in the wild. Compatibility is a natural condition satisfied by sequences of equivariant maps $\{\boldsymbol{f}\}$ arising in many problems of interest. Here we list some examples.

Graph(on) parameters. Graph parameters are functions of graphs (typically of any size) that do not depend on the labeling of their vertices. In other words, these are sequences of S_n-invariant functions $\{\boldsymbol{f}: \mathbb{S}^n \to \mathbb{R}\}$. Several graph invariants are also compatible with zero-padding, which is equivalent to adding an isolated vertex. For example, the max-cut value, the number of triangles, and cycles do not change if we add an isolated vertex, hence they are compatible. Other graph invariants only depend on a graph via its associated graphon, and hence are compatible with the embeddings in the graphon sequence (see example at the end of Section 3.2). These include graph homomorphism densities, which play a central role in extremal combinatorics [146].

Matrix mappings. Linear algebra operations are often compatible. For instance, the experiments in [152] aimed to learn the following matrix mappings: $X \mapsto \frac{1}{2}(X + X^\top)$, $X \mapsto \mathrm{diag}(\mathrm{diag}(X))$, $X \mapsto \mathrm{Tr}(X)$, and $X \mapsto \mathrm{argmax}_{\|v\|=1} \|Xv\|$. All of these are compatible sequences with the zero-padding embedding described in Section 3.2, and S_n-equivariant.

Orthogonal invariance. The papers [79, 222] consider the task of learning the function

$$\boldsymbol{f}(x_1, x_2) = \sin(\|x_1\|) - \frac{\|x_2\|^3}{2} + \frac{x_1^\top x_2}{\|x_1\|\|x_2\|}, \tag{3.2}$$

defined on $(\mathbb{R}^n)^{\oplus 2}$ with $n = 5$ from evaluation data. This function is $\mathrm{O}(n)$-invariant if we let rotations g act by $(g \cdot x_1, g \cdot x_2)$. Embedding $(\mathbb{R}^n)^{\oplus 2}$ into $(\mathbb{R}^{n+1})^{\oplus 2}$ by zero-padding each vector, we see that $\{\boldsymbol{f}\}$ is a compatible sequence of invariant functions.

Orthogonal equivariance. The $\mathrm{O}(3)$-equivariant task in [79, 222] consists of taking as input $n = 5$ particles in space, given by their masses and position vectors $(m_i, x_i)_{i=1}^n \in (\mathbb{R} \oplus \mathbb{R}^3)^{\oplus n}$, and outputting their moment of inertia matrix $\boldsymbol{f}(m_i, x_i) = \sum_{i=1}^n m_i (x_i^\top x_i I_3 - x_i x_i^\top)$. Embedding $\boldsymbol{V} = (\mathbb{R} \oplus \mathbb{R}^3)^{\oplus n}$ into $\boldsymbol{V}^+$ by zero-padding and letting $\boldsymbol{U} = \mathbb{S}^3$, we see that $\{\boldsymbol{f}\}$ is a compatible sequence of maps. Furthermore, let $\boldsymbol{G} = \mathrm{S}_n \times \mathrm{O}(3)$ act on $\boldsymbol{V}$ by $(\pi, g) \cdot (m_i, x_i)_{i=1}^n = (m_{\pi^{-1}(i)}, g x_{\pi^{-1}(i)})_{i=1}^n$, and on $\boldsymbol{U}$ by $(\pi, g) \cdot X = g X g^{-1}$ for $(\pi, g) \in \boldsymbol{G}$ (so permutations act trivially). Then each $\boldsymbol{f}$ is $\boldsymbol{G}$-equivariant.

There are many more examples of compatible mappings in the literature [139]. Therefore, we proceed to study compatible NNs. We derive conditions ensuring the compatibility of the linear layers and show that several standard activation functions are compatible.

Compatible linear layers. Since we only have data in some finite level n_0, we ask when do fixed-dimensional weights $W_i^{(n_0)}$ extend to a sequence satisfying (CompNN). The set of such $W_i^{(n_0)}$ forms a linear space, and we characterized this subspace in Theorem 2.3.4. The following assumption precisely guarantees that the above extension exists.

Assumption 3.3.2. *The sequences $\mathscr{V} = \{\boldsymbol{V}\}$, $\mathscr{U} = \{\boldsymbol{U}\}$ are obtained from direct sums and tensor products of the same sequence $\mathscr{V}_0$. The sequence $\mathscr{V}$ is generated in degree d_{g} and presented in degree d_{p}. The mapping $W^{(n_0)} \in \mathcal{L}(V^{(n_0)}, U^{(n_0)})^{G^{(n_0)}}$ at level $n_0 \geq d_{\mathrm{p}}$ satisfies*

$$W^{(n_0)}(V^{(m)}) \subseteq U^{(m)} \text{ for } m \leq d_{\mathrm{g}}.$$

Compatible activation functions. The majority of equivariant nonlinearities proposed in the literature are compatible, such as the following few examples.

Entrywise activations and permutations. Let $\sigma \colon \mathbb{R} \to \mathbb{R}$ be any nonlinear function. Define $\boldsymbol{\sigma} \colon (\mathbb{R}^n)^{\otimes k} \to (\mathbb{R}^n)^{\otimes k}$ by applying σ to each entry of a tensor. Then each $\boldsymbol{\sigma}$ is $\boldsymbol{G} = \mathrm{S}_n$-equivariant, and $\{\boldsymbol{\sigma}\}$ is a compatible sequence with respect to the zero-padding embeddings if and only if $\sigma(0) = 0$. On the other hand, if $\boldsymbol{\sigma}\colon (\mathbb{S}^{2^n})^{\otimes k} \to (\mathbb{S}^{2^n})^{\otimes k}$ is again applied entrywise and we use the graphon embedding (see example at the end of Section 3.2), then $\{\boldsymbol{\sigma}\}$ is a compatible sequence for *any* σ.

Bilinear layers. The authors of [79] map $(\boldsymbol{V}\otimes\boldsymbol{U})\oplus\boldsymbol{V}\to\boldsymbol{U}$ by sending $(v\otimes u, v')\mapsto \langle v, v'\rangle u$, which is equivariant for *any* group action on $\boldsymbol{V},\boldsymbol{U}$. This sequence of maps is also compatible.

Gated nonlinearities. The authors of [225] map $\boldsymbol{V}\oplus\mathbb{R}\to\boldsymbol{V}$ by $(v,\alpha)\mapsto v\sigma(\alpha)$ where $\sigma\colon\mathbb{R}\to\mathbb{R}$ is a nonlinearity. These are equivariant maps for any group acting trivially on the $\mathbb{R}$ component, and they form a compatible sequence.

3.4 A Computational Receipe for Learning Free Neural Networks

In this section, we describe an algorithm to train free and compatible NNs. We do so in two stages. First, we train our NN at a large enough level n_0 by finding bases for the weights and biases at that level and optimizing over the coefficients in this basis. Second, for any higher level $n > n_0$, we extend the trained NN at level n_0 by solving linear systems for the higher-dimensional weights and biases; for any lower level $n < n_0$ we project the NN using (`FreeNN`). Throughout, we fix consistent sequences $\mathscr{V}_i, \mathscr{U}_i$ and sequences of compatible equivariant nonlinearities $\{\boldsymbol{\sigma}_i\}$. We summarize our procedure in Algorithm 2. We proceed to elaborate on the steps of this algorithm, and then prove its correctness.

Finding bases for free NN. To train free networks (`FreeNN`), we fix n_0 *exceeding the presentation degrees* of $\mathscr{V}_i\otimes\mathscr{U}_i$ and $\mathscr{U}_i$ for all $i = 1,\dots,L$, and find a basis for equivariant weights and invariant biases at level n_0 using the algorithm of [79]. Specifically, Theorem 1 in [79] states that $b\in (U_i^{(n_0)})^{G^{(n_0)}}$ if and only if b satisfies that

$$\begin{aligned} B\,b &= 0 \qquad \left(\text{for all } B\in\mathcal{B}_i^{(n_0)}\right), \\ (D-I)\,b &= 0 \qquad \left(\text{for all } D\in\mathcal{D}_i^{(n_0)}\right), \end{aligned} \tag{3.3}$$

where $\mathcal{B}_i^{(n_0)}$ is a basis for the Lie algebra of $G^{(n_0)}$ and $\mathcal{D}_i^{(n_0)}$ are discrete generators for $G^{(n_0)}$, which are finite sets. Here B and D are represented as matrices acting on $U_i^{(n_0)}$. Since equivariant linear maps $W\colon V_i^{(n_0)}\to U_i^{(n_0)}$ are just the invariants in $\mathcal{L}(V_i^{(n_0)}, U_i^{(n_0)})$, they constitute the kernel of the analogously-defined set of equations. Thus, finding a basis for equivariant weights and invariant biases reduces to finding a basis for the kernel of a matrix (3.3), which is typically large and sparse.

Algorithm 2 Train a freely-described (resp., compatible) neural network.

Input: Architecture $\mathscr{V}_i, \mathscr{U}_i$, group sequence $\{\boldsymbol{G}\}$, and training data at level n_0.

Stage 1 Find free (resp., compatible) bases of the linear layers at level n_0.
Optimize over the coefficients in the bases to train the network at level n_0.

Stage 2 For any $n > n_0$, solve (ExtSyst) to extend the trained network to level n.
For any $n < n_0$, project via (FreeNN) to restrict the trained network to level n.

This can be tackled either using a Krylov-subspace method [79] or a sparse LU decomposition [90, 129].

Finding bases for Compatible NN. If we rather want to find basis for layers satisfying (CompNN), we fix n_0 *exceeding the presentation degrees* of $\mathscr{V}_i$ and let d_i be its generation degree. For most representations, (CompNN) implies zero bias, so we focus on the weights. We find a basis for weights $W_i^{(n_0)}$ satisfying the hypotheses of Theorem 2.3.4 by noting that Assumption 3.3.2 holds if and only if $W_i^{(n_0)}$ is $G^{(n_0)}$-equivariant and satisfies

$$\left[\mathcal{P}_{V_i^{(m)}} \otimes \left(I - \mathcal{P}_{U_i^{(m)}}\right)\right] \operatorname{vec}\left(W_i^{(n_0)}\right) = 0, \tag{3.4}$$

for all $m \leq d_i$, where vec denotes the vectorization of a matrix by stacking its columns. This characterization allows us to find a basis for weights by finding a basis for the kernel of a sparse matrix.

Extending to arbitrary dimensions. For any $n > n_0$, we extend our trained network at level n_0 by finding the unique equivariant weights and biases at level n projecting onto the trained ones as in (FreeNN). Formally, we find the unique $W_i^{(n)}$ and $b_i^{(n)}$ satisfying (3.3) with n_0 replaced by n, and

$$\begin{aligned} &\left[\mathcal{P}_{V_i^{(n_0)}} \otimes \mathcal{P}_{U_i^{(n_0)}}\right] \operatorname{vec}\left(W_i^{(n)}\right) = \operatorname{vec}\left(W_i^{(n_0)}\right), \\ &\mathcal{P}_{U_i^{(n_0)}} b_i^{(n)} = b_i^{(n_0)}. \end{aligned} \tag{ExtSyst}$$

This amounts to solving a linear system, which is again typically sparse. It can be solved via iterative methods, e.g., stochastic gradient descent or LSQR [170]. For any $n < n_0$, we set $W_i^{(n)}$ and $b_i(n)$ via (FreeNN), which is equivalent to (ExtSyst) with n and n_0 swapped.

Having described Algorithm 2 in detail, we prove its correctness.

Theorem 3.4.1. *Consider a neural network architecture with consistent sequences $\mathscr{V}_1, \mathscr{U}_1, \ldots, \mathscr{V}_L \mathscr{U}_L$, and $\mathscr{V}_{L+1}$ and consistent activation functions $\boldsymbol{\sigma}_1, \ldots, \boldsymbol{\sigma}_L$ as described Figure 3.1. Assume we run Algorithm 2 to learn a neural network at level n_0. Then, the following two hold.*

(Free) *If $n_0 \geq \max_i d_{\mathrm{p}}(\mathscr{V}_i \otimes \mathscr{U}_i)$ and $n_0 \geq \max_i d_{\mathrm{p}}(\mathscr{U}_i)$, then, Algorithm 2 extends the trained network at level n_0 to a unique free neural network.*

(Compatible) *If $n_0 \geq \max_i d_{\mathrm{p}}(\mathscr{V}_i)$ and $n_0 \geq \max_i d_{\mathrm{p}}(\mathscr{U}_i)$, then, Algorithm 2 extends the trained network at level n_0 to a unique compatible neural network.*

Proof. We start by proving the second statement. This is a simple corollary of Theorem 2.3.4. Note that once n_0 is bigger than the presentation degrees of $\mathscr{V}_i$ and $\mathscr{U}_i$, then, by Theorem 2.3.4, there is a unique extension of the weights and biases satisfying `CompNN`, which corresponds to the unique solution of (`ExtSyst`). The first statement follows by an analogous argument by substituting Theorem 2.3.4 with Proposition 2.2.9. This completes the proof. □

3.5 Numerical Experiments

In this section, we use Algorithm 2 to learn some of the compatible mappings from Section 3.3. Our implementation is based on that of [79], and is available at

https://github.com/mateodd25/free-nets.

For each experiment, we aim to learn a sequence of mappings $\{\boldsymbol{f} : \boldsymbol{V}_1 \to \boldsymbol{V}_{L+1}\}$. To do so, we fix a level n_0 and randomly generate data $\{(X_i, f^{(n_0)}(X_i))\} \subseteq V_1^{(n_0)} \times V_{L+1}^{(n_0)}$. To fit the data, we consider the architectures described at the end of this section and compute bases for weights and biases as we describe in Section 3.4. We then optimize over the coefficients in those bases using ADAM [121] starting from random initialization. Finally, we extend the trained network at level n_0 to several levels n. We evaluate our extended network on random test data generated at each level from the ground-truth sequence of maps.

We consider five examples from Section 3.3: the trace $f(X) = \mathrm{Tr}(X)$, diagonal extraction $f(x) = \mathrm{diag}(\mathrm{diag}(X))$, symmetric projection $f(X) = (X + X^\top)/2$, the top right-singular vector $f(X) = \mathrm{argmax}_{\|v\|=1} \|Xv\|$, and the function defined in (3.2). The first and last examples are invariant functions; the rest are equivariant mappings. For the first four examples, we set $\boldsymbol{G} = \mathrm{S}_n$, and for the last one, we set

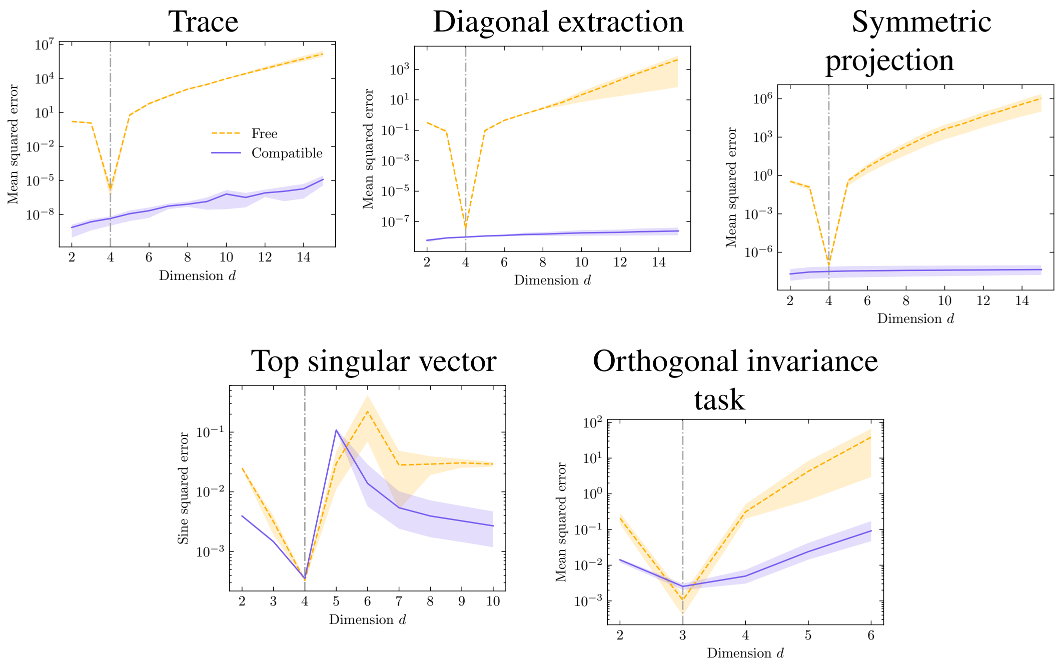


Figure 3.2: Test errors across dimensions for free and compatible networks. Each experiment is run three times; the lighter bands show the max and min runs, while the bold line shows the average. For diagonal extraction and symmetric projection, we measure the average MSE per entry. Vertical gray lines mark the dimension used for learning.

$\boldsymbol{G} = \mathrm{O}(n)$. We train with 3000 data points generated randomly and evaluate the test error at each dimension on 1000 fresh samples.

We use the mean squared error (MSE) as our loss function for all the experiments except for learning the top singular vector. For that experiment, we use the squared sine loss proposed by [152], i.e., $\ell(\widehat{y}, y) = 1 - \langle \widehat{y}, y\rangle^2/\|\widehat{y}\|^2\|y\|^2$. The resulting errors in each dimension for both free and compatible NNs are shown in Figure 3.2. The test errors at the trained dimension are competitive with those obtained in [79, 152, 222] for learning the same mappings, see Table 3.1 for comparison. Moreover, the errors we obtain for higher-dimensional extensions are competitive with errors obtained in the literature for NNs trained at those higher dimensions. For the orthogonally invariant task example, we trained our compatible NN in 3 dimensions using 3000 data points and got an error of 2e-2 when extending to 5 dimensions. In contrast, the NN of [222], which is tailored to orhogonal equivariance, achieves an error of 9e-3 when trained in 5 dimensions on the same number of points, while the NN of [79] trained in the same dimension achieves 4e-2.

Due to memory limitations, we set the training level to $n_0 = 3$ for the orthogonal

Example	Ours	Previous
trace	3e-8	1e-3 [152]
diag	1e-8	7e-6 [152]
sym	3e-8	7e-6 [152]
svd	4e-4	1.6e-3 [152]
orth.	2e-3	1e-2 [79]; 5e-4 [222]

Table 3.1: Comparison of errors from the literature

Experiment	Architecture		
	Input	Hidden	Output
Trace	$\mathscr{V}^{\otimes 2}$	$2\mathscr{V} + 2\mathscr{V}^{\otimes 2}$	$\mathscr{S}$
Diagonal extraction	$\mathscr{V}^{\otimes 2}$	$4\mathscr{V} + 4\mathscr{V}^{\otimes 2}$	$\mathscr{V}^{\otimes 2}$
Symmetric projection	$\mathscr{V}^{\otimes 2}$	$4\mathscr{V} + 4\mathscr{V}^{\otimes 2}$	$\mathscr{V}^{\otimes 2}$
Top singular vector	$\mathscr{V}^{\otimes 2}$	$25\mathscr{S} + 10\mathscr{V} + 2\mathscr{V}^{\otimes 2} + \mathscr{V}^{\otimes 3}$	$\mathscr{V}$
Orthogonal invariance	$2\mathscr{V}$	$25\mathscr{S} + 10\mathscr{V} + 2\mathscr{V}^{\otimes 2} + \mathscr{V}^{\otimes 3}$	$\mathscr{S}$

Table 3.2: Architectures used for the numerical examples. Recall that $\mathscr{S} = \{\mathbb{R}\}$ is the scalar sequence endowed with the trivial action $g \cdot v = v$.

invariance task, whereas Algorithm 2 would require $n_0 = 6$ to uniquely extend free networks[2] since the presentation degree of the linear maps between the hidden layers equal to six — see Theorem 3.4.1 in the appendix. This highlights an advantage of imposing compatibility — it allows us to uniquely extend a trained network from lower-dimensional data. Moreover, we see that imposing our compatibility condition yields substantially lower errors across dimensions. Remarkably, the test error for free NNs increases by many orders of magnitude for simple functions, a phenomenon that was previously noted by [152]. This further underscores the importance of compatibility when generalizing to other dimensions.

Detailed description of numerical experiments. In this section, we elaborate on the implementation details of the numerical experiments. All experiments were run on a 2021 Macbook Air M1 with 16GB of RAM.

Initilization. We initialize the weights and biases at random with small i.i.d Gaussian entries $N(0, 1/100)$.

[2]We find the minimum norm solution for the weights and biases using LSQR [170].

Data generation. The datasets $\{(X_i, f^{(n)}(X_i))\}$ for both training and testing (in every dimension) are generated as i.i.d. standard Gaussian samples with the appropriate size.

Architecture. For all the experiments, we use three layers, i.e., two hidden layers. All the sequences we take are sums and tensors products of the base sequence $\mathscr{V} = \{\mathbb{R}^d\}$. The activation functions are a composition of the form

$$\boldsymbol{h}\left(\mathbf{bilinear}(x) + x\right)$$

where **bilinear** [79] is described in Section 3.3 and $\boldsymbol{h}$ changes depending on the group; for the permutation group, $\boldsymbol{h}$ applies entry-wise ReLU for the permutation group, while for the orthogonal group, it applies a gated nonlinearity [225] — also described in Section 3.3.

We chose the smallest hidden layer size that yielded competitive results in order to control the size of the linear systems we employ for extending the networks. The top singular vector and orthogonal invariant examples required a larger architecture, so we only extended them to smaller dimensions. The architecture for each example are summarized in Table 3.2.

3.6 Discussion

We leveraged representation stability to prove that a broad family of equivariant NNs extends to higher dimensions, which enables us to train an infinite sequence of NNs using a finite amount of data. Extending networks to higher dimensions often results in substantial test errors. To improve generalization, we introduced a compatibility condition relating networks across dimensions. We characterized compatibility and developed an algorithm to train free and compatible NNs. Finally, we applied our method to several numerical examples from the literature. In these examples, free NNs trained without imposing compatibility generalize poorly to higher dimensions, even when learning simple linear functions. In contrast, compatible NNs generalize significantly better.

The following are some of the limitations of our approach that we hope to address in future work.

Extension to non-compact groups. We assumed that the group is compact and acts orthogonally so that the orthogonal projection $\mathcal{P}_V = \boldsymbol{\varphi}^*$ is $\boldsymbol{G}$-equivariant, used for instance in the proof of Proposition 2.2.3. This property holds more generally when $\boldsymbol{G}$ is linearly reductive. Thus, our framework extends to non-compact groups such as the Lorentz group $O(3,1)$ and so-called rescaling groups $\mathbb{R}^n_{>0}$ consisting of diagonal matrices with strictly positive entries, both of which are important to applications of equivariant deep learning to physics [19, 221].

Some groups arising in applications are not reductive however, including any group containing translations such as the group of rigid motions $SE(n)$. This difficulty can be circumvented in some cases, such as centering the data when learning an $SE(n)$-invariant function and imposing only $SO(n)$-invariance on the architecture. In general, however, more work is needed to address the non-reductive case.

Compatible activation functions and biases. Our compatibility conditions (CompNN) require the activation functions we use to be compatible with our embeddings relating the different dimensions. For example, applying any activation function $\sigma\colon \mathbb{R} \to \mathbb{R}$ entrywise always satisfies (CompNN) for the graphon embeddings. However, only activations with $\sigma(0) = 0$ satisfy (CompNN) for the zero-padding embeddings. On the one hand, some activation functions used in practice, such as the sigmoid $\sigma(x) = (1 + e^{-x})^{-1}$, do not satisfy this condition. Moreover, compatibility requires the biases to be the same across dimensions, which implies that the biases must be multiples of the first canonical basis vector when using the zero-padding embedding, and a multiple of the all-1's matrix when using the graphon embedding. Therefore, compatibility limits the architectures that we can use. On the other hand, if the sequence of mappings we are trying to learn is known in advance to satisfy compatibility, then it is desirable to use architectures that enforce compatibility as well. In that case the mapping we learn also generalizes correctly to other dimensions, as the numerical experiments in Section 3.5 demonstrate.

Scalability. We hope to address several limitations of our approach in future work. First, while our compatibility condition improves generalization, it can be restrictive in certain settings. For example, it often yields zero bias. It would be interesting to learn the relation between maps in different dimensions directly from the data rather than assume it is known in advance. Second, for examples like the leading eigenvector, the error increases substantially when extending to higher dimensions, even if we impose compatibility. It would be interesting to improve generalization to higher dimensions in these examples by imposing additional compatibility conditions. Third, extending a trained network to higher dimensions in our examples involved solving large and sparse systems, which we currently do not fully exploit. Incorporating sparsity in the toolbox of [79] will enable training larger networks.

Our current implementation faces scalability challenges, primarily due to the inadequate support for GPU-accelerated sparse linear algebra operations in Python. Notably, there are already some efforts to close this gap [179]. Moving forward, we are optimistic about enhancing our framework to leverage these advancements effectively. Furthermore, the compatibility condition constraints the family of functions one could hope to learn. This presents a compelling avenue for future research: exploring other regularization strategies that lead to good generalization across dimensions. Moreover, the concepts introduced in this chapter paved the way to studying any-dimensional statistical learning, raising several natural questions: Can we bound the generalization error across dimensions? How do we characterize the notation of statistical complexity for classes of any-dimensional problems? And how do we approach the formulation of lower bounds within this context?

Chapter 4

ANY-DIMENSIONAL POLYNOMIAL OPTIMIZATION

This chapter is based on [138].

4.1 Introduction

We begin by illustrating our construction of lower bounds and the role de Finetti's theorem plays in it in the context of Example 1.3.1 from Chapter 1, highlighting the main steps that we shall generalize in the rest of the chapter.

A First Illustration

In the sequence of POPs (1.1), for any $n|N$ the map $(x_1,\dots,x_n)\mapsto(\underbrace{x_1,\dots,x_1}_{N/n\text{ times}},\dots,\underbrace{x_n,\dots,x_n}_{N/n\text{ times}})$ takes a feasible point of the n'th problem to a feasible point of the N'th problem with the same objective value. This implies that the optimal values of (1.1) satisfy $u_{nk}\leq u_n$ for each n,k. The goal is to obtain lower bounds on the limiting optimal value $u_\infty=\inf_n u_n$, which may be expressed as:

$$u_\infty=\inf_{\mu\in\mathcal{P}(\Theta)} f\left(\mathbb{E}_\mu[X^{\alpha_1}],\dots,\mathbb{E}_\mu[X^{\alpha_m}]\right). \tag{4.1}$$

where $X\in\mathbb{R}^d$ is a random variable and $\mathcal{P}(\Theta)$ denotes the set of probability measures supported on Θ. Our approach proceeds in four steps.

First, we express u_∞ as a linear optimization problem over product measures. Suppose f is equal to a single monomial with exponent $\beta\in\mathbb{N}^m$. In this case, we have that $f\left(\mathbb{E}_\mu[X^{\alpha_1}],\dots,\mathbb{E}_\mu[X^{\alpha_m}]\right)=\prod_{i=1}^m(\mathbb{E}_\mu[X^{\alpha_i}])^{\beta_i}$, which can be rewritten as:

$$\prod_{i=1}^m(\mathbb{E}_\mu[X^{\alpha_i}])^{\beta_i}=\prod_{i=1}^m\mathbb{E}_{\mu^{\otimes\beta_i}}[Y_1^{\alpha_i}\cdots Y_{\beta_i}^{\alpha_i}]=\mathbb{E}_{\mu^{\otimes 1'\beta}}[(Z_{1,1}^{\alpha_1}\cdots Z_{1,\beta_1}^{\alpha_1})\cdots(Z_{m,1}^{\alpha_m}\cdots Z_{m,\beta_m}^{\alpha_m})]$$

where the $Y_i\in\mathbb{R}^d$ and the $Z_{i,j}\in\mathbb{R}^d$ are random variables. In both equalities, we appeal to the observation that the expectation of a product of independent random variables is the product of their expectations. Thus, when f is a monomial of degree $\deg(f)=\mathbb{1}_m^\top\beta$, we have identified a monomial q in $d\cdot\deg(f)$ variables such that:

$$f\left(\mathbb{E}_\mu[X^{\alpha_1}],\dots,\mathbb{E}_\mu[X^{\alpha_m}]\right)=\mathbb{E}_{\mu^{\otimes\deg(f)}}[q]$$

By linearity, such a polynomial q exists for any polynomial f, enabling us to rewrite the limiting optimal value u_∞ of (1.1) as:

$$u_\infty = \inf_{\nu} \; \mathbb{E}_\nu[q] \quad \text{s.t.} \quad \nu = \mu^{\otimes \deg(f)} \text{ for some } \mu \in \mathcal{P}(\Theta). \tag{4.2}$$

This is the promised reformulation, as the objective here is linear in the decision variable ν.

Our second step reformulates (4.2) by appealing to de Finetti's theorem [10]. Specifically, due to the linearity of the objective in (4.2) in the decision variable ν, we can replace the constraint on ν by its convex hull, and optimize over measures ν that are mixtures of product measures $\mu^{\otimes \deg(f)}$ for $\mu \in \mathcal{P}(\Theta)$. By de Finetti's theorem, ν is a mixture of product measures precisely when it extends to an exchangeable measure on vectors of any length (and hence to the law of an infinite exchangeable sequence of random variables). Formally, this means that there is an infinite sequence $(\nu_i \in \mathcal{P}(\Theta^i))_{i\in\mathbb{N}}$ of exchangeable measures such that $\nu_{\deg(f)} = \nu$ and $(X_1, \dots, X_{i+1}) \sim \nu_{i+1}$ implies $(X_1, \dots, X_i) \sim \nu_i$ for all $i \in \mathbb{N}$. We therefore obtain

$$u_\infty = \inf_{(\nu_i)_{i\in\mathbb{N}}} \mathbb{E}_{\nu_{\deg(f)}}[q] \;\; \text{s.t.} \;\; \nu_i \in \mathcal{P}(\Theta^i) \text{ is exchangeable, } \nu_j \text{ projects to } \nu_i \text{ for } j \geq i, \tag{4.3}$$

where the projection condition in the constraint means that ν_i equals the pushforward of ν_j under the projection $(x_1, \dots, x_j) \mapsto (x_1, \dots, x_i)$ extracting the first i entries.

The third step is simply the observation that the reformulation (4.3) yields a sequence of lower bounds on u_∞ by truncating the sequence of measures. Specifically, for each $n \geq \deg(f)$ consider

$$\ell_n = \inf_{(\nu_i)_{i=1}^n} \mathbb{E}_{\nu_{\deg(f)}}[q] \quad \text{s.t.} \quad \nu_i \in \mathcal{P}(\Theta^i) \text{ is exchangeable, } \nu_j \text{ projects to } \nu_i \text{ for } j \geq i. \tag{4.4}$$

Equivalently, we are optimizing over measures $\nu \in \mathcal{P}(\Theta^{\deg(f)})$ that extend to a measure on Θ^n, but not necessarily to Θ^N for $N > n$. We therefore have $\ell_n \leq u_\infty$. Moreover, we can conclude that for all $n \geq \deg(f)$ we have $\ell_n \leq \ell_{n+1}$ for all n and that the limit of these lower bounds $\ell_\infty = \lim_{n\to\infty} \ell_n$ equals u_∞.

Our final step is to show that the optimization problem (4.4) is in fact a finite-dimensional POP for each n. While the polynomial q that we derived in (4.3) is a function of $d \cdot \deg(f)$ variables, it can be viewed as a polynomial in $d \cdot n$ variables as $n \geq \deg(f)$, i.e., $q(x_1, \dots, x_{\deg(f)}) = q(x_1, \dots, x_n)$ with each $x_i \in \mathbb{R}^d$. The objective of (4.4) can then be rewritten as $\mathbb{E}_{\nu_n}[q]$ thanks to the constraints of (4.4), and this problem (4.4) becomes

$$\ell_n = \inf_{\nu} \mathbb{E}_\nu[q] \quad \text{s.t.} \quad \nu \in \mathcal{P}(\Theta^n) \text{ is exchangeable.}$$

Now we make the observation that the objective is bilinear in ν and q. Consequently, we may remove the exchangeability constraint on ν by symmetrizing q in the objective. Specifically, let $\mathrm{sym}_n q$ represent the following symmetrization of q (as a function of $d \cdot n$ variables):

$$\mathrm{sym}_n q = \frac{1}{|\mathsf{S}_n|} \sum_{\sigma \in \mathsf{S}_n} q(\sigma \cdot (x_1, \ldots, x_n)),$$

where S_n denotes the permutation group which acts by permuting the elements of the tuple $(x_1, \ldots, x_n)$. For any measure $\nu \in \mathcal{P}(\Theta^n)$, not necessarily exchangeable, we then have $\mathbb{E}_\nu[\mathrm{sym}_n q] = \mathbb{E}_{\mathrm{sym}_n \nu} q$ where $\mathrm{sym}_n \nu \in \mathcal{P}(\Theta^n)$ is the exchangeable law of ΠX with $X \sim \nu$ and Π a uniformly random permutation independent of X. Using this observation, we can rewrite (4.4) as:

$$\ell_n = \inf_{\nu} \mathbb{E}_\nu[\mathrm{sym}_n q] \quad \text{s.t.} \quad \nu \in \mathcal{P}(\Theta^n).$$

The objective is linear in the decision variable ν, and the constraint set is the convex hull of the collection of Dirac measures supported at each point in Θ^n. By only optimizing over the Dirac measures, we have that:

$$\ell_n = \inf_{y \in \Theta^n} \mathrm{sym}_n q(y). \tag{4.5}$$

If Θ is semialgebraic, this problem is a fixed-dimensional POP in dimension $d \cdot n$.

In summary, we have presented a constructive method for obtaining a sequence of increasing lower bounds on the limiting value $u_\infty = \inf_n u_n$ of the any-dimensional POP (1.1) via fixed-dimensional POPs. Comparing (4.5) with the original sequence of POPs (1.1), we see that our lower bounds consist of minimizing a modified objective over the same sequence of constraint sets. The resulting optimal values are related to each other via

$$\cdots \leq \ell_n \leq \ell_{n+1} \leq \cdots \ell_\infty = u_\infty \cdots \leq u_{nk} \leq u_n \leq \cdots, \tag{4.6}$$

for all $n, \ell, k \in \mathbb{N}$. By applying a quantitative variant of de Finetti's theorem due to Diaconis and Freeman [59], we further obtain the following rate of convergence for these optimal values

$$u_n - \ell_n \lesssim \frac{1}{n}. \tag{4.7}$$

In this chapter, we generalize the above construction to other any-dimensional POPs, such as the ones in Example 1.3.2 and 1.3.3. The limiting optimal value of these problems cannot be reformulated in terms of product measures like in (4.2), and hence the classical de Finetti theorem does not apply. Nevertheless, we prove new generalizations of de Finetti's theorem that allow us to reformulate these limiting optimal values as in (4.3) in terms of sequences of measures projecting onto each other in a suitable sense. We then obtain lower bounds in terms of finite-dimensional POPs analogous to (4.5), and derive rates of convergence for the resulting bounds using our new de Finetti-type theorems.

Our Contributions

Our framework consists of three components—formally defining any-dimensional POPs, proving new de Finetti-type theorems, and applying these theorems to obtain lower bounds on the limiting optimal values of any-dimensional POPs in terms of finite-dimensional POPs. We outline each of these in some more detail:

Formalizing any-dimensional POPs (Section 4.3). For concreteness, denote a sequence of POPs by $\{u_n = \inf_{x\in\Omega_n}\ p_n(x)\}$ with semialgebraic constraints $\Omega_n \subseteq \mathbb{V}_n$ and polynomial objectives $p_n : \mathbb{V}_n \to \mathbb{R}$, where $\mathbb{V}_n$ is the vector space underlying the n'th problem. To obtain lower bounds on $u_\infty = \inf_n u_n$ via finite-time procedures, we require that the problems in the sequence are related to each other in some fashion, as without such relations we would be seeking lower bounds on arbitrary sequences of numbers. In the illustration in Section 4.1, these relations are given by the map $\delta_{N,n}(x_1,\ldots,x_n) = (\underbrace{x_1,\ldots,x_1}_{N/n\text{ times}},\ldots,\underbrace{x_n,\ldots,x_n}_{N/n\text{ times}})$ from $\mathbb{R}^{d\times n}$ to $\mathbb{R}^{d\times N}$ for $n|N$. This map takes a feasible point of the n'th problem map to a feasible point for the N'th problem, and the sequence of costs (p_n) satisfies $p_n = p_N \circ \delta_{N,n}$. The consequence is that the optimal values satisfy $u_N \leq u_n$ for $n|N$. The sequence of problems in Example 1.3.3 can also be understood as being related across dimension by an appropriate type of duplication map (see Example 4.3.2), and the corresponding sequence of optimal values also satisfy $u_N \leq u_n$ for $n|N$. In other cases such as Example 1.3.2, one can check that the problems are similarly related across dimension by the zero-padding map $\zeta_{N,n}(x) = (x, 0_{N-n})$ for $n \leq N$, so that the associated optimal values satisfy $u_N \leq u_n$ for $n \leq N$. We refer to such sequences of POPs related via embeddings across problem dimensions as *freely-described* (see Definition 4.3.1). In Section 4.1, the lower bounds are given by a sequence of POPs $\{\ell_n = \inf_{x\in\Omega_n}\ q_n(x)\}$ in which the costs are higher-dimensional symmetrizations of a fixed polynomial $q_n = \mathrm{sym}_n q$; we refer to such sequences of POPs as *freely-symmetrized* (see Definition 4.3.3). Generalizing the case study in Section 4.1, our objective is to obtain lower bounds via freely-symmetrized POPs on the limiting optimal values of freely-described POPs. The definitions of these two types of sequence of POPs are based on ideas from *representation stability*, which concerns sequences of group representations related across dimension [47].

Generalizing de Finetti (Section 4.4). De Finetti's theorem characterizes an infinite sequence of measures that project onto each other as in (4.3). Diaconis and Freedman [59] show that one can obtain a dense subset of such infinite sequences of measures by sampling with replacement the coordinates of finite random vectors. We broaden this perspective in two ways. First, we generalize the notion of infinite sequences of measures that project onto each other by again leveraging representation stability. In particular, for sequences of representations of permutation groups related across dimension, we consider projections induced by these relations and associated sequences of exchangeable measures which project onto each other (see Definition 4.4.1). Second, we generalize the idea of sampling with replacement using actions of maps between finite sets on the underlying vector spaces (see Definitions 4.4.12 and 4.4.18). These notions enable us to generalize the result of Diaconis and Freedman in Theorems 4.4.15 and 4.4.20. As an illustration, we obtain the following new 'dual' de Finetti theorem:

Theorem 4.1.1 (informal, special case; see Theorem 4.4.20)**.** *Consider any sequence of random vectors $(X^{(n)} \in \mathbb{R}^n)_n$ projecting onto each other in the following sense*

$$X^{(nk)} = \Big(\underbrace{X_1^{(nk)}, \dots, X_k^{(nk)}}, \quad \cdots \quad , \underbrace{X_{(n-1)k+1}^{(nk)}, \dots, X_{nk}^{(nk)}}\Big)$$

$$\Big\downarrow$$

$$X^{(n)} \overset{d}{=} \Big(X_1^{(nk)} + \dots + X_k^{(nk)}, \quad \cdots \quad , X_{(n-1)k+1}^{(nk)} + \dots + X_{nk}^{(nk)}\Big),$$

Any such sequence can be approximated arbitrarily well by sequences constructed as follows: For a finite random vector $Y \in \mathbb{R}^d$, and a uniformly random map $F_{n,d}\colon [d] \to [n]$ that is independent of Y, construct the sequence $(\tau(F_{n,d})Y \in \mathbb{R}^n)_{n\in\mathbb{N}}$ where the ith entry of $\tau(F_{n,d})Y$ is given by $\sum_{j\in F_{n,d}^{-1}(i)} Y_j$.

Applying $\tau(F_{n,d})$ to Y amounts to randomly binning the coordinates of Y into n bins and summing all the coordinates in each bin. Its adjoint $\tau(F_{n,d})^\star$ with respect to the standard inner products corresponds to sampling d entries with replacement from a length-n vector, hence the duality with classical de Finetti. Our results also apply to higher-dimensional arrays. In particular, we derive new dual analogs of results from Aldous–Hoover theory and the theory of graphons [115, 146].

Lower Bounds on Any-Dimensional POPs (Section 4.5). We apply our generalized de Finetti theorems to construct lower bounds for the limiting optimal values of freely-described POPs in Section 4.5. As we remarked earlier in Section 4.1, step 2 corresponding to (4.3) represents the right abstraction to reformulate many families of freely-described POPs, for appropriate notions of 'project'. We formalize this point in Theorem 4.5.2. The following is an informal version of Theorem 4.5.2 applied to the problem of minimizing symmetric functions as in Example 1.3.2.

Theorem 4.1.2 (informal, special case; see Theorem 4.5.2). *Consider* $u_n = \inf_{x\in\Omega_n} f(s_1(x),\dots,s_k(x))$ *where* $f(s_1,\dots,s_k)$ *is a symmetric function as in Example 1.3.2 and* Ω_n *is either* $\mathbb{R}^n$*, the simplex* Δ^{n-1}*, or the* ℓ_1 *unit ball.*

1. *There exists a fixed polynomial* q *in* $d = \deg f(s_1,\dots,s_k)$ *variables satisfying*

$$f(s_1(x),\dots,s_k(x)) = \mathbb{E}q\left(\sum_{i\in F_{d,n}^{-1}(1)} x_i,\dots,\sum_{i\in F_{d,n}^{-1}(d)} x_i\right), \tag{4.8}$$

   for all $x\in\mathbb{R}^n$ *and all* $n\in\mathbb{N}$*, where* $F_{d,n}\colon [n]\to[d]$ *is a uniformly random map.*

2. *The optimal values* $\ell_n = \inf_{x\in\Omega_n} q_n$ *yield increasing lower bounds* $\ell_n \le u_\infty$*, where* q_n *is a certain symmetrization of* q*.*

3. *If* Ω_n *is compact, we get* $u_n - \ell_n = O(1/\sqrt{n})$*.*

Note that $f(s_1,\dots,s_k)$ and q are related precisely by the sampling procedure arising in our generalization of de Finetti's theorem in Theorem 4.1.1. Moreover, the construction of q given $f(s_1,\dots,s_k)$ is fully automatic, much like in Section 4.1. For the symmetric function in Example 1.3.5, the fixed polynomial q is the Motzkin polynomial.

Our results suggest a duality between the duplication and zero-padding embeddings in the following sense. For freely-described POPs related across dimension by duplication (resp. zero-padding) maps, the lower bounds generalizing (4.3) optimize over measures projecting onto each other by the adjoint of zero-padding (resp.

duplication) maps. Our final bounds are given in terms of freely-symmetrized POPs consisting of the same constraint set as the original sequence of freely-descirbed POPs but with a different sequence of costs. We also interpret and analyze our bounds from the perspective of any-dimensional certificates of nonnegativity in Section 4.5.

Applications (Section 4.6). We illustrate our methods with several examples in Section 4.6, with the code to reproduce all our numerical results available at `https://github.com/eitangl/anyDimPolyOpt`. In some cases, we clarify the relationship between any-dimensional POPs appearing in the literature, such as between the monomial means of Example 1.3.1 and power means leveraged in [2] (Section 4.6), or between the homomorphism densities of Example 1.3.3 and their injective analogs (Section 4.6). These are revealed to be consequences of our generalized de Finetti theorems. In other cases, we compute new bounds for symmetric functions (Section 4.6) and homomorphism numbers (Section 4.6), which are derived via bases for symmetric functions obtained using our de Finetti theorems; to the best of our knowledge, these bases and their applications have not been explored previously in the literature.

Chapter Organization. Section 4.2 provides background on representation stability, which serves as a foundation for the later sections. For readers primarily interested in the optimization applications, it is possible to get to the illustrations of our framework in Section 4.6 after reading Sections 4.2, 4.3, 4.4, and 4.5. This shorter route through the chapter presents the statements of our generalizations of de Finetti's theorem in Section 4.4 as well as statements and proofs of our results on freely-symmetrized bounds for freely-described POPs in Section 4.5, and these suffice for the developments in Section 4.6. On the other hand, for readers primarily interested in a full treatment of our generalizations of de Finetti's theorem, it is possible to follow all of Section 4.4 directly after Section 4.2.

Related Work

Our work is related to several areas. We now survey past work in these areas and explain the connection with our own results in more detail.

Polynomial Optimization. Leveraging the literature on sums-of-squares and moment (SOS) methods for fixed-dimensional POPs [16, 133, 172], a recent body of work has developed convex relaxations for any-dimensional problems in special cases. SDP relaxations for unconstrained problems of the form of Example 1.3.1 for the particular case of $d = 1$ were studied in [2], while specific unconstrained cases of Example 1.3.2 were studied in [3, 4] (see Section 4.6 for a 'higher-dimensional' generalization of Example 1.3.2). There is also a long line of work on certifying inequalities between graph homomorphism densities using SOS techniques, also known as flag algebras in the literature [28, 184–186]. Setting up these SDP relaxations often involves serious and case-specific analytic work related to taking certain limits of symmetry-reduced SDPs. These, in turn, are based on understanding decompositions into irreducibles of spaces of polynomials with increasing degrees; see [2, 3, 28, 177] for example. In contrast, our methods are more straightforward to apply, as our finite-dimensional POPs for obtaining lower bounds can be constructed automatically for a given any-dimensional POP, including in settings where the relevant decompositions into irreducibles become quite involved [177]. SDP relaxations of these finite-dimensional POPs can subsequently be computed using standard software packages. Another hierarchy of SDP relaxations were developed in [124] for a class of optimization problems over measures encompassing Example 1.3.1. In that work, the authors apply techniques inspired by noncommutative SOS directly to the reformulation of the limiting optimal value as an optimization problem over measures (4.1). It is not, however, clear how to extend their approach to other any-dimensional POPs, such as Examples 1.3.2-1.3.3, where the limiting optimal value does not admit such a formulation.

Representation Stability. The study of sequences of representations related across dimensions was initiated in [47], where it was observed that the decompositions into irreducibles of such sequences often stabilize. The seminal paper [46] studied sequences of representations of symmetric groups from the perspective of representations of categories, and proved that decompositions of such representations into irreducibles stabilize starting from a prescribed dimension. Subsequently, a large literature has emerged studying representations of various categories, in particular relating the combinatorics of the morphisms in the category with properties of its representations [45, 84, 195, 197, 227]. We use the language and results of the representation stability literature to formalize sequences of symmetric POPs related across dimensions. Further, we extend this body of work by showing that representations of the (opposite of the) category FinSet of finite sets with all maps between them underlie de Finetti's theorem and our generalizations of this theorem. Specifically, our extension of the notion of sampling with replacement appearing in de Finetti's theorem consists of applying a uniformly random morphism from the $\mathrm{FinSet}^{\mathrm{op}}$ and FinSet categories (see Section 4.4). We then prove de Finetti-type theorems for representations of $\mathrm{FinSet}^{\mathrm{op}}$ and FinSet in Theorems 4.4.15 and 4.4.20, respectively. Thus, our work represents a new point of contact between representation stability and probability via random actions of a category on its representations.

Exchangeable Arrays. Since de Finetti's original result [56] on binary infinite exchangeable arrays, a number of extensions have been obtained pertaining both to the types of random variables involved and the sense in which they are exchangeable. The former includes extensions to random variables taking real values [57, 65] and values in a compact Hausdorff space [102]. The latter includes extensions to finite exchangeable arrays [59], higher-dimensional exchangeable arrays and partially-exchangeable such arrays [7, 103, 115]. In a separate line of work [20], the convergence of sequences of dense graphs to so-called graphons (graph functions) is studied. These were then shown to characterize consistent graph models [148], which are sequences of measures on simple graphs projecting onto each other similarly to (4.4), and their characterization in terms of graphons was shown to be a special case of Aldous–Hoover theory in [60]. In this chapter, we generalize de Finetti's theorem in a different—and to our knowledge as yet unexplored—direction that is motivated by our optimization applications. Specifically, we view the preceding results as pertaining to sequences of exchangeable measures that project onto each other in specific ways, similarly to (4.4). We then study sequences of measures projecting onto each other in more general ways, and we generalize the results in [59] to give dense subsets of such sequences as well as associated nonasymptotic error bounds.

Graphons and Flag Algebras. The limiting optimal value u_∞ for minimizing homomorphism densities in Example 1.3.3 can be reformulated as optimization over graphons, which are symmetric measurable functions $W\colon [0,1]^2 \to [0,1]$, see [146, §16]. From this perspective, one can obtain $u_n - u_\infty = O(1/n)$ using [147, Lemma 2.4(b)]. Finally, taking SOS relaxations of u_n and taking their limits as $n \to \infty$ yields precisely the lower bounds on u_∞ that can be proved using flag algebras [28, 184, 185]. Relative to this previous work, our framework gives several new results and suggests avenues for future work. To the best of our knowledge, proving inequalities in graph densities by minimizing the corresponding expression in injective densities has not explicitly appeared in the literature before. These lower bounds are generally better than the ones obtained by computing u_n and subtracting $O(1/n)$, as explained in Section 4.5 and illustrated in Example 4.6.7. Further, they open the door to non-SOS-based proofs of graph density inequalities, since any relaxation applied to ℓ_n yields a lower bound on u_∞. Finally, our bounds apply not just for simple graphs but also for weighted ones.

Acknowledgments

We thank Emma Church, Jordan Ellenberg, and Graham White for suggesting the idea of using representations of FI and related categories to study de Finetti theorems. We also thank Jordan Ellenberg for helpful and productive discussions and for his feedback on our generalized de Finetti theorems. We thank Jorge Garza Vargas for the idea of using a Wasserstein distance in Lemma 4.4.19. The authors were partially supported by AFOSR grants FA9550-23-1-0070 and FA9550-23-1-0204.

Notation

Finite-dimensional real vector spaces are denoted by blackboard bold letters such as $\mathbb{V}, \mathbb{U}$, their product by $\mathbb{V} \times \mathbb{U}$, and their tensor product by $\mathbb{V} \otimes \mathbb{U}$. The space of symmetric order-d tensors over $\mathbb{V}$ is denoted by $\mathrm{Sym}^d\mathbb{V}$. The ring of polynomials over such a vector space $\mathbb{V}$ is denoted by $\mathbb{R}[\mathbb{V}]$, with $\mathbb{R}[\mathbb{V}]_{\leq d}$ and $\mathbb{R}[\mathbb{V}]_d$ denoting the spaces of polynomials of degree at most d and homogeneous polynomials of degree exactly d. Note that $\mathrm{Sym}^d\mathbb{V} \cong \mathbb{R}[\mathbb{V}]_d$. The space of symmetric $n \times n$ matrices is denoted by $\mathbb{S}^n$, and the set of symmetric matrices with entries in a set $\Omega \subseteq \mathbb{R}$ is denoted by $\mathbb{S}^n(\Omega)$. All vector spaces in this chapter are endowed with an inner product, and if $A \colon \mathbb{V} \to \mathbb{U}$ is a linear map between two vector spaces, we denote by $A^\star \colon \mathbb{U} \to \mathbb{V}$ the adjoint of A with respect to the inner products on $\mathbb{V}, \mathbb{U}$. The inner product on $\mathbb{R}^n$ is always the standard one $\langle x, y\rangle = x^\top y$, and inner products on tensors $[\mathbb{R}^n]^{\otimes k}$ (in particular, on symmetric tensors $\mathrm{Sym}^k\mathbb{R}^n$, and hence on polynomials via the above isomorphism) are the Frobenius inner products. We denote by S_n the group of permutations on n letters. We denote by $\preceq$ a partial order. We take the set of natural numbers to be $\mathbb{N} = \{0, 1, 2 \ldots\}$. We write $n|N$ for $n, N \in \mathbb{N}$ to denote divisibility of N by n, meaning that there is $k \in \mathbb{N}$ satisfying $N = nk$. Random variables are denoted by capital letters. The distribution of a random variable X is denoted $\mathrm{Law}(X)$. If μ is a probability measure, we write $X \sim \mu$ when $\mathrm{Law}(X) = \mu$. If X and Y are random variables, we write $X \overset{d}{=} Y$ to mean $\mathrm{Law}(X) = \mathrm{Law}(Y)$. Expectations are denoted by $\mathbb{E}$.

4.2 Background on Representation Stability

In this section, we discuss the relevant concepts from the representation stability literature [46, 47, 139]. We begin by briefly reviewing linear representations of groups; more comprehensive introductions are available in [81, 202]. A (linear) representation of a group G on a vector space $\mathbb{V}$ is a group homomorphism ρ mapping elements of G to linear maps from $\mathbb{V}$ to itself. A group element $g \in \mathsf{G}$ then acts on vectors $x \in \mathbb{V}$ via $g \cdot x := \rho(g)x$. It is customary to omit the homomorphism ρ since it is usually clear from context and to simply write gx for the action of $g \cdot x$. A vector $x \in \mathbb{V}$ is G-invariant if $gx = x$ for all $g \in \mathsf{S}$. The collection of invariants in $\mathbb{V}$ forms a subspace denoted by $\mathbb{V}^{\mathsf{G}}$. In this chapter, all groups are symmetric groups, that is, $\mathsf{G} = \mathsf{S}_n$ for some n. In this case, we can endow $\mathbb{V}$ with a G-invariant inner product, so we shall always assume that our inner products are group-invariant (or equivalently, $\rho(\mathsf{G})$ consists of orthogonal matrices). With respect to such an inner product, the orthogonal projection onto $\mathbb{V}^{\mathsf{G}}$ is given by the *Reynolds operator* $\mathrm{sym}_{\mathsf{G}} = \frac{1}{|\mathsf{G}|}\sum_{g\in\mathsf{G}} g$.

Consistent Sequences

Representation stability studies nested sequences of representations called consistent sequences. These sequences are indexed by a directed poset (one in which every two elements have a common upper bound) that specifies how these representations are nested. For this chapter, it suffices to consider the natural numbers $\mathbb{N}$ with one of two partial orders $\preceq$, the standard total order denoted by $(\mathbb{N}, \leq)$, and the divisibility partial order denoted by $(\mathbb{N}, \cdot|\cdot)$. In the first case $n \preceq N$ corresponds to $n \leq N$, while in the second case $n \preceq N$ corresponds to $N = nk$ for some $k \in \mathbb{N}$.

Definition 4.2.1 (Consistent sequence)**.** *A* consistent sequence *is a collection* $\mathscr{V} = \{\mathcal{N}, (\mathsf{S}_n)_{n\in\mathcal{N}}, (\mathbb{V}_n)_{n\in\mathcal{N}}, (\varphi_{N,n})_{n\preceq N}\}$ *composed of:*

(Poset) *a directed poset* $\mathcal{N}$*;*

(Groups) *nested symmetric groups* $\mathsf{S}_n \subseteq \mathsf{S}_N$ *whenever* $n \preceq N$*;* [1]

(Vector spaces) *a sequence of vector spaces* $(\mathbb{V}_n)_{n\in\mathcal{N}}$ *such that* $\mathbb{V}_n$ *is a* S_n*-representation;*

(Embeddings) *linear injective maps* $(\varphi_{N,n}\colon \mathbb{V}_n \to \mathbb{V}_N)_{n\preceq N}$ *such that* $\varphi_{N,n}$ *is a* S_n*-equivariant.*

Throughout this chapter, we focus on the following two consistent sequences arising from the sequence of permutation groups S_n acting on the sequence of vector spaces

[1] Consistent sequences have been defined for more general sequences of groups [47], but only symmetric groups are used in this chapter.

$\mathbb{R}^n$ by permutating coordinates, namely, by $(\sigma \cdot x)_i = x_{\sigma^{-1}(i)}$. They are both indexed by $\mathbb{N}$ but with different partial orders, and they involve different embeddings between dimensions.

(Zero-padding) Let

$$\mathscr{Z} = \{(\mathbb{N}, \leq), (\mathsf{S}_n)_{\mathscr{Z}}, (\mathbb{R}^n), (\zeta_{N,n})_{n\leq N}\}, \tag{4.9}$$

be the consistent sequence in which $\mathsf{S}_n \subseteq \mathsf{S}_N$ by identifying $g \in \mathsf{S}_n$ with $\mathrm{blkdiag}(g, I_{N-n}) \in \mathsf{S}_N$ and with embeddings by zero-padding $\zeta_{N,n}(x) = (x, 0_{N-n})$.

(Duplication) Let

$$\mathscr{D} = \{(\mathbb{N}, \cdot \mid \cdot), (\mathsf{S}_n)_{\mathscr{D}}, (\mathbb{R}^n), (\delta_{N,n})_{n|N}\}, \tag{4.10}$$

be the consistent sequence in which $\mathsf{S}_n \subseteq \mathsf{S}_N$ by identifying $g \in \mathsf{S}_n$ with $g \otimes I_{N/n} \in \mathsf{S}_N$ and with embeddings by duplicating entries $\delta_{N,n}(x) = x \otimes \mathbb{1}_{N/n} = (x_1, \dots, x_1, \dots, \underbrace{x_n, \dots, x_n}_{N/n \text{ times}})$.

We can construct new consistent sequences from existing ones such as $\mathscr{Z}$ and $\mathscr{D}$ using standard operations on vector spaces.[2] Moreover, it is possible to quantify the complexity of the new consistent sequence by keeping track of the constructions applied. Specifically, if $\mathscr{W} = \{\mathcal{N}, (\mathsf{S}_n)_{n\in\mathcal{N}}, (\mathbb{W}_n)_{n\in\mathcal{N}}, (\varphi_{N,n})_{n\leq N}\}$, $\mathscr{U} = \{\mathcal{N}, (\mathsf{S}_n)_{n\in\mathcal{N}}, (\mathbb{U}_n)_{n\in\mathcal{N}}, (\psi_{N,n})_{n\leq N}\}$ are two consistent sequences indexed by the same poset and with the same nested sequence of symmetric groups acting on them, we can construct the following new consistent sequences:

(Products) The *(Cartesian) product* of $\mathscr{W}$ and $\mathscr{U}$ is $\mathscr{W} \times \mathscr{U} = \{\mathcal{N}, (\mathsf{S}_n)_{n\in\mathcal{N}}, (\mathbb{W}_n \times \mathbb{U}_n)_{n\in\mathcal{N}}, (\varphi_{N,n} \times \psi_{N,n})_{n\leq N}\}$. We denote by $\mathscr{W}^d$ the product of d copies of $\mathscr{W}$.

(Tensors) The *tensor product* of $\mathscr{W}$ and $\mathscr{U}$ is $\mathscr{W} \otimes \mathscr{U} = \{\mathcal{N}, (\mathsf{S}_n)_{n\in\mathcal{N}}, (\mathbb{W}_n \otimes \mathbb{U}_n)_{n\in\mathcal{N}}, (\varphi_{N,n} \otimes \psi_{N,n})_{n\leq N}\}$. The k'th tensor power $\mathscr{W}^{\otimes k}$ of $\mathscr{W}$ is the tensor product of $\mathscr{W}$ with itself k times.

(Polynomials) The set of *degree-k polynomials over $\mathscr{W}$* is $\mathrm{Sym}^k\mathscr{W} = \{\mathcal{N}, (\mathsf{S}_n)_{n\in\mathcal{N}}, (\mathrm{Sym}^k\mathbb{W}_n)_{n\in\mathcal{N}}, (\varphi_{N,n}^{\otimes k})_{n\leq N}\}$, which is also the set of order-k symmetric tensors over $\mathscr{W}$. Here we view $\mathrm{Sym}^k\mathbb{W}_n \subseteq \mathbb{W}_n^{\otimes k}$ and restrict $\varphi_{N,n}^{\otimes k}$ to that subspace.

[2]Formally, these ‘constructions’ are (bi)functors from the category of vector spaces to itself.

The set of polynomials of degree at most k over $\mathscr{W}$ is denoted $\mathrm{Sym}^{\leq k}\mathscr{W} = \bigoplus_{j=1}^{k} \mathrm{Sym}^{j}\mathscr{W}$.

The actions of S_n on $\mathbb{W}_n$ and on $\mathbb{U}_n$ naturally extends to all of the above vector spaces. For instance, if $g \in \mathsf{S}_n$ we have $g \cdot (w, u) = (g \cdot w, g \cdot u)$ for $(w, u) \in \mathbb{W}_n \times \mathbb{U}_n$ and $g \cdot \sum_i w_i \otimes u_i = \sum_i (gw_i) \otimes (gu_i)$ for $\sum_i w_i \otimes u_i \in \mathbb{W}_n \otimes \mathbb{U}_n$.

If we start from a consistent sequence $\mathscr{V}$ and apply some sequence of the above operations to it to obtain another sequence $\mathscr{U}$, then we say that $\mathscr{U}$ is a $\mathscr{V}$-sequence. We denote a $\mathscr{V}$-sequence by $\mathscr{U} = \mathcal{F}(\mathscr{V})$ where $\mathcal{F}$ is the sequence of operations applied to $\mathscr{V}$ to produce $\mathscr{U}$. We furthermore quantify the complexity of the operations $\mathcal{F}$ applied to $\mathscr{V}$ using its *degree*, defined inductively as follows. We set the degree of the identity construction that doesn't change its input to $\deg \mathrm{id} = 1$, and inductively set

$$\deg \mathcal{F}_1 \times \mathcal{F}_2 = \max\{\deg \mathcal{F}_1, \deg \mathcal{F}_2\}; \quad \deg \mathcal{F}_1 \otimes \mathcal{F}_2 = \deg \mathcal{F}_1 + \deg \mathcal{F}_2;$$
$$\deg \mathrm{Sym}^k \mathcal{F}_1 = k \deg \mathcal{F}_1.$$

If $\mathscr{U}$ is a $\mathscr{V}$-sequence, we then say that $\mathscr{U}$ has degree d over $\mathscr{V}$ if there is a construction $\mathcal{F}$ of degree d satisfying $\mathscr{U} = \mathcal{F}(\mathscr{V})$, in which case we write $\deg_{\mathscr{V}} \mathscr{U} = \deg \mathcal{F}$. We omit the subscript $\mathscr{V}$ when it is clear from context. Note that $\mathscr{Z}$ and $\mathscr{D}$ are indexed by different posets, so a given consistent sequence cannot be both a $\mathscr{Z}$-sequence and a $\mathscr{D}$-sequence.

Example 4.2.2. *The consistent sequences* $\mathrm{Sym}^k\mathscr{Z}$ *and* $\mathrm{Sym}^k\mathscr{D}$ *both consist of the sequence of vector spaces* $\mathbb{V}_n = \mathrm{Sym}^k\mathbb{R}^n$*, and they are degree-$k$ $\mathscr{Z}$- and $\mathscr{D}$-sequences, respectively. Viewing an element $X \in \mathbb{V}_n$ as an order-k symmetric tensor in n dimensions, permutations act by* $(\sigma \cdot X)_{i_1,\ldots,i_k} = X_{\sigma^{-1}(i_1),\ldots,\sigma^{-1}(i_k)}$*, and this action is common to both consistent sequences. The embeddings between dimensions are different, however, and given by $\zeta_{N,n}^{\otimes k}$ and $\delta_{N,n}^{\otimes k}$. Explicitly, these correspond to zero-padding and duplicating the entries of X as follows:*

$$(\zeta_{N,n}^{\otimes k} X)_{i_1,\ldots,i_k} = \begin{cases} X_{i_1,\ldots,i_k} & \text{if } i_j \in [n] \text{ for all } j \in [k] \\ 0 & \text{if } i_j \in [N] \setminus [n] \text{ for some } j \in [k] \end{cases}$$
$$(\delta_{N,n}^{\otimes k} X)_{(i_1-1)N/n+j_1,\ldots,(i_k-1)N/n+j_k} = X_{i_1,\ldots,i_k}, \quad \text{for } i_\ell \in [n] \text{ and } j_\ell \in [N/n].$$

Here $n \leq N$ for the first line and $n|N$ for the second. Equivalently, we can view the elements of $\mathbb{V}_n = \mathrm{Sym}^k\mathbb{R}^n$ *as homogeneous polynomials $p(x)$ of degree k over* $\mathbb{R}^n$*. From this perspective, permutations act by* $(\sigma \cdot p)(x) = p(\sigma^{-1}x)$ *and the above* *embeddings become* $(\zeta_{N,n}^{\otimes k}p)(x) = p(\zeta_{N,n}^{\star}x)$ *where* $\zeta_{N,n}^{\star}$ *is the adjoint of* $\zeta_{N,n}$ *in the standard inner product, which extracts the first n entries of a length-N vector. The analogous identity hold for* $\delta_{N,n}^{\otimes k}$.

Free Invariants and Their Stabilization

The sequences of polynomials that arise in the any-dimensional POPs in Examples 1.3.1-1.3.3 and in the lower bounds (4.5) in Section 3.1 are invariant polynomials defined over a consistent sequence, related to each other across dimensions in particular ways. In this subsection, we show that sequences of such polynomials of bounded degree form a finite-dimensional vector space as a consequence of representation stability [47]. This fact allows us to describe any-dimensional POPs by finitely-many parameters in Section 4.3, and it plays an important role in our construction of lower bounds in Section 4.5

Invariants in a consistent sequence $\mathscr{V} = \{\mathcal{N}, (\mathsf{S}_n), (\mathbb{V}_n), (\varphi_{N,n})\}$ are related to each other across dimensions in two ways. First, projections of high-dimensional invariants yield low-dimensional ones, that is, for any $n \preceq N$, the orthogonal projection $\varphi_{N,n}^{\star}\colon \mathbb{V}_N \to \mathbb{V}_n$ applied to a high-dimensional invariant $v_N \in \mathbb{V}_N^{\mathsf{S}_N}$ yields a low-dimensional invariant $\varphi_{N,n}^{\star}v_N \in \mathbb{V}_n^{\mathsf{S}_n}$. Second, embedding low-dimensional invariants and symmetrizing them yields high-dimensional invariants, that is, if $v_n \in \mathbb{V}_n^{\mathsf{S}_n}$ then $\mathrm{sym}_{\mathsf{S}_N}\varphi_{N,n}v_n \in \mathbb{V}_N^{\mathsf{S}_N}$ whenever $n \preceq N$. In what follows, we abbreviate $\mathrm{sym}_{\mathsf{S}_N}\varphi_{N,n}$ for any $n \preceq N$ by $\mathrm{sym}_N^{\mathscr{V}}$.

Definition 4.2.3. *Let* $\mathscr{V}$ *be a consistent sequence.*

(Freely-described) *The space of* freely-described elements *is the space of sequences*

$$\varprojlim \mathbb{V}_n^{\mathsf{S}_n} = \left\{(v_n) \in \prod_{n\in\mathcal{N}} \mathbb{V}_n^{\mathsf{S}_n} : \varphi_{N,n}^{\star}v_N = v_n \textit{ for all } n \preceq N\right\}. \tag{4.11}$$

In other words, a freely-described element is a sequence of invariants projecting onto each other.

(Freely-symmetrized) *The space of* freely-symmetrized elements *is the space of equivalence classes*

$$\varinjlim \mathbb{V}_n^{\mathsf{S}_n} = \bigsqcup_{n\in\mathcal{N}} \mathbb{V}_n^{\mathsf{S}_n}/\sim \quad \textit{where } v_n \sim v_N \textit{ if } v_N = \mathrm{sym}_N^{\mathscr{V}} v_n, \tag{4.12}$$

where $\bigsqcup$ denotes the disjoint union. A freely-symmetrized element is one of these equivalence classes, which we write as a sequence[3] $(\mathrm{sym}_n^{\mathscr{V}} v_k)_{n\geq k}$ *when* $v_k \in \mathbb{V}_k^{\mathsf{S}_k}$.

The limits here correspond to inverse and direct limits from category theory. Informally, the directions of the arrows denote the 'direction' in which a sequence of invariants must satisfy a relation—the left and right arrows correspond to mapping higher-dimensional invariants to lower-dimensional ones and vice-versa, respectively.

Example 4.2.4. *The spaces of freely-described and freely-symmetrized elements in* $\mathscr{X}$ *are both one-dimensional and spanned by* $(\mathbb{1}_n)$ *and* $(\frac{1}{n}\mathbb{1}_n)$*, respectively.*

For $\mathrm{Sym}^{\leq 2}\mathscr{X}$*, which consists of polynomials* $\mathbb{R}[x_1,\dots,x_n]_{\leq 2}$ *in n variables of degree at most 2, we have:*

$$\varprojlim \mathbb{R}[x_1,\dots,x_n]_{\leq 2}^{\mathsf{S}_n} = \mathrm{span}\left\{(1), \left(\sum_{i=1}^n x_i\right), \left(\sum_{i=1}^n x_i^2\right), \left(\sum_{1\leq i<j\leq n} x_i x_j\right)\right\},$$

$$\varinjlim \mathbb{R}[x_1,\dots,x_n]_{\leq 2}^{\mathsf{S}_n} = \mathrm{span}\left\{(1), \left(\frac{1}{n}\sum_{i=1}^n x_i\right), \left(\frac{1}{n}\sum_{i=1}^n x_i^2\right), \left(\frac{1}{\binom{n}{2}}\sum_{1\leq i<j\leq n} x_i x_j\right)\right\}.$$

As we alluded to previously, in all of these examples the spaces of freely-described and freely-symmetrized elements have the same finite dimension, and moreover the difference between the basis elements for both is in how they are "averaged" across dimensions. Remarkably, this is not a coincidence but a consequence of a general phenomenon whereby the projections $\varphi_{N,n}^\star$ and embeddings $\mathrm{sym}_N^{\mathscr{V}}$ become isomorphisms between invariants. To explain this phenomenon, we begin by observing that the space of freely-described elements can be identified with the dual space of freely-symmetrized elements.

Proposition 4.2.5. *Let* $\mathscr{V}$ *be a consistent sequence. We have* $(\varinjlim \mathbb{V}_n^{\mathsf{S}_n})^\star = \varprojlim \mathbb{V}_n^{\mathsf{S}_n}$*, under the duality pairing* $\langle(\mathrm{sym}_n^{\mathscr{V}} v_k), (u_n)\rangle = \langle v_k, u_k\rangle$ *for any* $v \in \mathbb{V}_k$ *and* $(u_n) \in \varprojlim \mathbb{V}_n^{\mathsf{S}_n}$*. Here* $\mathbb{V}^\star$ *denotes the (algebraic) dual space of* $\mathbb{V}$ *consisting of all linear functionals on* $\mathbb{V}$.

Proof. The given pairing is well-defined since if $v_N = \mathrm{sym}_n^{\mathscr{V}} v_n$ then

$$\langle v_N, u_N\rangle = \langle \mathrm{sym}_N \varphi_{N,n} v_n, u_N\rangle = \langle v_n, \varphi_{N,n}^\star u_N\rangle = \langle v_n, u_n\rangle.$$

[3] Two such sequences are equivalent if they agree for all $n \geq N$ for some fixed N.

Furthermore, distinct freely-described elements define distinct linear functionals via this pairing because if $\langle(\mathrm{sym}_n^{\mathscr{V}} v), (u_n)\rangle = 0$ for all $v \in \mathbb{V}_k^{\mathsf{S}_k}$ and all k, then in particular $\langle(\mathrm{sym}_n^{\mathscr{V}} u_m), (u_n)\rangle = \|u_m\|^2 = 0$ so $u_m = 0$ for all m. Thus, it suffices to show that every linear functional on $\varinjlim \mathbb{V}_n^{\mathsf{S}_n}$ comes from a freely-described element. Suppose $\ell\colon \varinjlim \mathbb{R}[\mathbb{V}_n]^{\mathsf{S}_n} \to \mathbb{R}$ is a linear functional. Define the sequence of linear functionals $\ell_n\colon \mathbb{V}_n^{\mathsf{S}_n} \to \mathbb{R}$ by $\ell_n = \ell|_{\mathbb{V}_n^{\mathsf{S}_n}}$, and let $u_n \in \mathbb{V}_n^{\mathsf{S}_n}$ satisfy $\ell_n(\cdot) = \langle \cdot, u_n\rangle$. Note that for any $n \leq N$,

$$\langle \varphi_{N,n}^{\star} u_N, v_n\rangle = \langle u_N, \mathrm{sym}_N \varphi_{N,n} v_n\rangle = \ell(\mathrm{sym}_N^{\mathscr{V}} v_n) = \ell(v_n) = \langle u_n, v_n\rangle,$$

for all $v_n \in \mathbb{V}_n^{\mathsf{S}_n}$, so we conclude that $\varphi_{N,n}^{\star} u_N = u_n$ for all $n \leq N$. Thus, (u_n) is a freely-described element representing the linear functional ℓ under the above duality pairing. □

Based on this result, the space of freely-described elements is finite-dimensional if and only if the space of freely-symmetrized elements is, and in this case the two spaces have the same dimension. The next proposition shows this finite dimensionality for $\mathscr{L}$- and $\mathscr{D}$-sequences.

Proposition 4.2.6 (Finite-dimensional invariants)**.** *If $\mathscr{V}$ is a $\mathscr{L}$- or a $\mathscr{D}$-consistent sequence of degree d, then* $\dim \varprojlim \mathbb{V}_n^{\mathsf{S}_n} = \dim \varinjlim \mathbb{V}_n^{\mathsf{S}_n} \leq \dim \mathbb{V}_d^{\mathsf{S}_d}$. *Equality holds when $\mathscr{V}$ is a $\mathscr{L}$-sequence.*

Proof. If $\mathscr{V}$ is a $\mathscr{L}$-sequence, all this follows from [139, Prop. 2.9, Thm. 2.11]. If $\mathscr{V} = \mathcal{F}(\mathscr{D})$ is a $\mathscr{D}$-sequence, applying this result for the corresponding $\mathscr{L}$-sequence $\mathcal{F}(\mathscr{L})$ we conclude that $\dim \mathbb{V}_n^{\mathsf{S}_n}$ is constant for all $n \geq d$. This implies that $\dim \varinjlim \mathbb{V}_n^{\mathsf{S}_n} \leq \dim \mathbb{V}_d^{\mathsf{S}_d}$ since if $(\mathrm{sym}_n^{\mathscr{V}} v_{k_i})$ are linearly independent and $N \geq k_i$ for all i and $N \geq d$, then $\mathrm{sym}_N^{\mathscr{V}} v_{k_i}$ must be linearly independent in $\mathbb{V}_N^{\mathsf{S}_N}$ whose dimension is $\dim \mathbb{V}_d^{\mathsf{S}_d}$. Proposition 4.2.5 now yields the result. □

In fact, we prove in Proposition 4.5.4 below that equality is always attained in Proposition 4.2.6 as we prove there that the projections and embeddings become isomorphisms of invariants for all sufficiently large dimensions. This was proved for $\mathscr{L}$-sequences in [139] using the fact that the high-dimensional spaces in such sequences are spanned by the orbits of the low-dimensional ones. Below, we use our de Finetti theorems to extend this result to $\mathscr{D}$-sequences.

4.3 Any-Dimensional Polynomial Problems

In this section, we define the any-dimemsional POPs that we study. In particular, we define two classes of sequences of POPs defined on a sequence of growing-dimensional vector spaces. The first captures the relations across dimensions satisfied by the problems whose limiting values we seek to bound, e.g., in Examples 1.3.1-1.3.3, and the second captures the relations satisfied by our lower bounds for them, e.g., in (4.5).

Freely-Described POPs

The first class of problems, from which we start and for which we seek lower bounds, is one in which freely-described polynomials are minimized over a growing sequence of semialgebraic sets.

Definition 4.3.1 (Freely-Described POPs)**.** *Let* $\mathscr{U} = \{\mathcal{N}, (\mathsf{S}_n), (\mathbb{V}_n), (\varphi_{N,n})_{n\leq N}\}$ *be a consistent sequence. A sequence of POPs* $(u_n = \inf_{x\in\Omega_n} p_n(x))$ *is called* freely-described *if the following two conditions hold.*

(Cost) *The sequence of costs* (p_n) *is a freely-described element of* $\mathrm{Sym}^{\leq d}\mathscr{V}$ *for some degree* d*. Equivalently,* (p_n) *is a sequence of polynomials of degree at most* d *that satisfy* $p_N \circ \varphi_{N,n} = p_n$ *for all* $n \preceq N$*.*

(Constraint) *The sequence* (Ω_n) *consists of constraint sets that are semialgebraic and increasing in the sense that* $\varphi_{N,n}\Omega_n \subseteq \Omega_N$ *if* $n \preceq N$*.*

As we illustrate next, Examples 1.3.1-1.3.3 in Section 3.1 are all instances of freely-described POPs.

Example 4.3.2. *The any-dimensional POP* (1.1) *in Example 1.3.1 is freely-described with respect to the consistent sequence* $\mathscr{D}^d$*; indeed, as we noted in Section 4.1, the embeddings* $\delta^d_{N,n}(x_1,\ldots,x_n) = (\underbrace{x_1,\ldots,x_1}_{N/n \text{ times}},\ldots,\underbrace{x_n,\ldots,x_n}_{N/n \text{ times}})$ *for* $n|N$ *map feasible points for the nth problem to feasible points for the* N*th one without changing the objective value.*

In Example 1.3.2 on symmetric functions, the sequence of POPs is freely-described with respect to the consistent sequence $\mathscr{Z}$*, since power sums are unchanged under zero-padding.*

Finally, in Example 1.3.3 on graph density inequalities, we view the homomorphism densities as polynomials over adjacency matrices of simple graphs, which are points *in the hypercubes* $\Omega_n = \mathbb{S}^n(\{0,1\})$ *consisting of symmetric matrices with binary entries. It can then be shown that these homomorphism densities are unchanged under duplication of the vertices sending an adjacency matrix* $X \in \Omega_n$ *to* $X \otimes \mathbb{1}_k \mathbb{1}_k^\top \in \Omega_{nk}$*, see Section 4.6. Thus, these any-dimensional POPs are freely-described with respect to* $\mathrm{Sym}^2\mathscr{D}$.

Although any-dimensional POPs are given as an infinite sequence of POPs of growing dimension, they can often be described by finitely-many parameters. In particular, if $\mathscr{V}$ is a $\mathscr{L}$- or $\mathscr{D}$-sequence, then Proposition 4.2.6 shows that the freely-described sequence of costs (p_n) is an element of a finite-dimensional vector space, so it can be specified by its coefficients in a basis.

A direct consequence of the definition of a freely-described POP is that the sequence of optimal values (u_n) satisfies $u_N \leq u_n$ whenever $n \leq N$:

$$u_n = \inf_{x\in\Omega_n} p_n(x) = \inf_{x\in\Omega_n} p_N \circ \varphi_{N,n}(x) = \inf_{y\in\varphi_{N,n}\Omega_n} p_N(y) \geq \inf_{y\in\Omega_N} p_N(y) = u_N. \tag{4.13}$$

Indeed, all of the examples above of freely-described POPs from Section 3.1 satisfy this relation between the optimal values. We seek lower bounds on the limiting optimal value $u_\infty = \inf_n u_n$ of a freely-described POP. As each u_n is, in general, larger than u_∞ by (4.13), merely obtaining a lower bound on one of the u_n's in the sequence, e.g., using sums-of-squares methods, does not yield a bound on the limiting value.

Freely-Symmetrized POPs

To derive lower bounds on u_∞, we generalize the example from Section 4.1 and construct a sequence of finite-dimensional POPs whose optimal values are monotonically *increasing* to u_∞. The sequences of POPs we construct belong to the following class of freely-*symmetrized* problems.

Definition 4.3.3 (Freely-Symmetrized POPs)**.** *Let* $\mathscr{L} = \{\mathcal{N}, (\mathsf{S}_n), (\mathbb{V}_n), (\psi_{N,n})_{n\leq N}\}$ *be a consistent sequence. A sequence of POPs* $(\ell_n = \inf_{x\in\Omega_n} q_n(x))$ *is called* freely-symmetrized *if:*

(Cost) *The sequence of costs* (q_n) *is a freely-symmetrized element of* $\mathrm{Sym}^{\leq d}\mathscr{L}$ *for some degree* d*. That is, there is a* k *and* $q \in \mathbb{R}[\mathbb{V}_k]_{\leq d}$ *such that* $q_n = \mathrm{sym}_n^{\mathscr{L}} q$ *for all* $n \geq k$*.*

(Constraint) *The sequence* (Ω_n) *consists of constraint sets that are semialgebraic and project onto each other in the sense that* $\psi^\star_{N,n}\Omega_N \subseteq \Omega_n$ *if* $n \le N$.

In a freely-symmetrized POP ($\ell_n = \inf_{x\in\Omega_n}\ q_n(x)$), the sequence of optimal values (ℓ_n) is in general increasing with n. Specifically, as $q_n(x) = \mathrm{sym}_n^{\mathscr{L}} q_k(x) = q_k(\psi^\star_{n,k}\mathrm{sym}_{\mathsf{S}_n} x)$ for some $q_k \in \mathbb{R}[\mathbb{V}_k]^{\mathsf{S}_k}$ with $n \ge k$, we have

$$\ell_n = \inf_{x\in\Omega_n} \mathrm{sym}_n^{\mathscr{L}} q_k(x) = \inf_{\mu\in\mathcal{P}(\Omega_n)} \mathbb{E}_\mu[q_k(\psi^\star_{n,k}\mathrm{sym}_{\mathsf{S}_n} X)] = \inf_{\mu\in\mathcal{P}(\Omega_n)^{\mathsf{S}_n}} \mathbb{E}_\mu[q_k(\psi^\star_{n,k} X)],$$

since if $X \sim \mu$ then $\mathrm{sym}_{\mathsf{S}_n} X \sim \mathrm{sym}_{\mathsf{S}_n}\mu \in \mathcal{P}(\Omega_n)^{\mathsf{S}_n}$. Next, we can rewrite ℓ_n as an optimization over the pushforwards under $\psi^\star_{n,k}$ of measures in $\mathcal{P}(\Omega_n)^{\mathsf{S}_n}$, to obtain

$$\ell_n = \inf_{\nu\in\psi^\star_{n,k}\mathcal{P}(\Omega_n)^{\mathsf{S}_n}} \mathbb{E}_\nu[q_k(X)]. \tag{4.14}$$

Since $\psi^\star_{N,n}\Omega_N \subseteq \Omega_n$ for any $n \le N$, we have

$$\psi^\star_{N,k}\mathcal{P}(\Omega_N)^{\mathsf{S}_N} = \psi^\star_{n,k}\psi^\star_{N,n}\mathcal{P}(\Omega_N)^{\mathsf{S}_N} \subseteq \psi^\star_{n,k}\mathcal{P}(\Omega_n)^{\mathsf{S}_n}, \tag{4.15}$$

for $k \le n \le N$. From (4.14) we conclude that ℓ_N is the minimum value of the same objective over a smaller constraint set than ℓ_n, hence the optimal values (ℓ_n) of freely-symmetrized POP are increasing

$$\ell_n \le \ell_N, \quad \text{whenever } k \le n \le N. \tag{4.16}$$

Example 4.3.4. *For graphs H and G, define $t_{\mathrm{inj}}(H;G)$ to be the fraction of injective maps $V(H) \to V(G)$ between their vertex sets that are graph homomorphisms. Note that the only difference with $t(H;G)$ from Example 1.3.3 is that we restrict ourselves to injective maps. Identifying a graph X on n vertices with its adjacencny matrix $X \in \mathbb{S}^n$, and defining the monomial $X^H = \prod_{1\le i\le j\le |V(H)|} X_{i,j}^{H_{i,j}}$ over $\mathbb{S}^{|V(H)|}$, we have $t_{\mathrm{inj}}(H;X) = \mathrm{sym}_n^{\mathrm{Sym}^2\mathscr{Z}} X^H$ as we show in (4.61) in Section 4.6 below. Therefore, for any $\alpha_1,\ldots,\alpha_m \in \mathbb{R}$ and graphs $H_1,\ldots,H_m$, certifying inequalities in* injective *graph densities over unweighted graphs can be written as*

$$\ell_n = \inf_{X\in\mathbb{S}^n\cap\{0,1\}^{n\times n}} \sum_{j=1}^m \alpha_j t_{\mathrm{inj}}(H_j;X) = \mathrm{sym}_n^{\mathrm{Sym}^2\mathscr{Z}} \sum_{j=1}^m \alpha_j X^{H_j},$$

for all $n \ge \max_j |V(H_j)|$, and therefore constitutes a freely-symmetrized problem over $\mathrm{Sym}^2\mathscr{Z}$. In particular, we have $\ell_n \le \ell_{n+1}$ for all such n, which shows that $\ell_n \ge 0$ for all such n precisely when $\ell_k \ge 0$ for $k = 1,\ldots,\max_j |V(H_j)|$. This *observation implies that checking whether an inequality in injective homomorphism densities is nonnegative for graphs of all sizes is decidable (it involves checking nonnegativity over a finite set of graphs of size up to* $\max_j |V(H_j)|$*), in contrast to the undecidability of the analogous question for non-injective densities [97].*

Suppose now that we are given a freely-described POP with limiting optimal value u_∞, and we wish to obtain lower bounds on u_∞. If we can identify a freely-symmetrized POP with optimal values (ℓ_n) such that their limiting optimal value $\ell_\infty = \sup_n \ell_n$ equals u_∞, then each ℓ_n yields a lower bound on u_∞ in terms of a finite-dimensional POP, and the sequence (ℓ_n) of these values gives monotonically increasing, convergent lower bounds for u_∞.

Pairings of Freely-Described and Freely-Symmetrized Problems

Given a freely-described POP with optimal values (u_n), our goal is to construct a freely-symmetrized POP with optimal values (ℓ_n) such that $u_\infty = \inf_n u_n$ is equal to $\ell_\infty = \sup_n \ell_n$, generalizing the example in Section 4.1. Similarly to that example, we shall modify the sequence of objectives while keeping the same sequence of constraints. Importantly, the freely-described problems (u_n) and their corresponding freely-symmetrized lower bounds (ℓ_n) are defined with respect to different consistent sequences. In the example of Section 4.1, the freely-described problem is defined over the $\mathscr{D}$-sequence $\mathscr{D}^d$, while the freely-symmetrized problem is defined over the $\mathscr{L}$-sequence $\mathscr{L}^d$. We generalize this example in Section 4.5 to any freely-described POP over a $\mathscr{L}$- or a $\mathscr{D}$-sequence. If the freely-described POP is defined over a $\mathscr{D}$-sequence, then our freely-symmetrized lower bounds are defined over the corresponding $\mathscr{L}$-sequence on the same underlying vector spaces, like in Section 4.1, and vice-versa.

Since the sequence of constraint sets (Ω_n) will be the same for the two sequences of POPs, we further require these sets to embed into each other in one consistent sequence and project into each other in the other consistent sequence. This ensures Definitions 4.3.1 and 4.3.3 are satisfied and the resulting optimal values (u_n) and (ℓ_n) are monotonic as in (4.13) and (4.16). We now formally define the pairs of consistent sequences with respect to which our freely-described POPs and associated freely-symmetrized POPs are defined, and the conditions on the constraint sets we impose. These correspond precisely to the class of problems to which our framework is applicable.

Definition 4.3.5 (FinSet-pair)**.** *An* FinSet-pair of degree d *is a pair of consistent sequences*

$$\mathscr{U} = \{(\mathbb{N}, \preceq_{\mathscr{U}}), (\mathsf{S}_n), (\mathbb{V}_n), (\varphi_{N,n})_{n\preceq_{\mathscr{U}}N}\} \quad \textit{and} \quad \mathscr{L} = \{(\mathbb{N}, \preceq_{\mathscr{L}}), (\mathsf{S}_n), (\mathbb{V}_n), (\psi_{N,n})_{n\preceq_{\mathscr{L}}N}\},$$

on the same sequence of vector spaces $(\mathbb{V}_n)$ *such that there is a degree-*d *sequence of constructions* $\mathcal{F}$ *satisfying* $(\mathscr{U}, \mathscr{L}) = (\mathcal{F}(\mathscr{X}), \mathcal{F}(\mathscr{D}))$ *or* $(\mathcal{F}(\mathscr{D}), \mathcal{F}(\mathscr{X}))$*. That is,* $\mathscr{U}$ *and* $\mathscr{L}$ *are obtained by applying the same constructions to* $\mathscr{X}$ *and* $\mathscr{D}$*, or vice-versa.*

A sequence of subsets $(\Omega_n \subseteq \mathbb{V}_n)$ *is said to be* compatible with respect to the FS-pair $(\mathscr{U}, \mathscr{L})$*, or* FinSet*-compatible for short, if* (i) *each* Ω_n *is* S_n*-invariant;* (ii) $\varphi_{N,n}\Omega_n \subseteq \Omega_N$ *whenever* $n \preceq_{\mathscr{U}} N$*; and* (iii) $\psi_{N,n}^{\star}\Omega_N \subseteq \Omega_n$ *whenever* $n \preceq_{\mathscr{L}} N$*.*

We show in Section 4.4 below that such pairs of consistent sequences can be derived from certain actions of maps between finite sets, which is the reason for the terminology ‘FS’. Such actions play a central role in our generalization of de Finetti’s theorem there. We leverage these generalized de Finetti theorems in Section 4.5 to accomplish our goal: for any FinSet-pair $(\mathscr{U}, \mathscr{L})$ we start from a freely-described POP over $\mathscr{U}$ with FinSet-compatible constraints (Ω_n), and derive lower bounds in terms of freely-symmetrized POPs over $\mathscr{L}$ with the same constraint sets.

4.4 Free Measures and Generalized de Finetti Theorems

As we have seen in the illustration of Section 4.1, de Finetti’s theorem plays a prominent role in our derivation of lower bounds. In this section we present generalizations of this theorem by using ideas from representation stability. We begin in Section 4.4 by reformulating de Finetti’s theorem as an assertion pertaining to sequences of measures, and state our generalizations in terms of such sequences of measures in Theorems 4.4.5 and 4.4.9. The former pertains to exchangeable multi-dimensional arrays and recovers some aspects of Aldous–Hoover–Kallenberg theory, while the latter is a certain ‘dual’ result that seems to not have been previously studied. These are the two results that will be used in Section 4.5 to construct lower bounds for more general freely-described sequences of problems. Then in Sections 4.4 and 4.4 we explain the key ingredients in proving our theorems, stated in the more abstact language of actions of maps between finite sets that yields cleaner statements and proofs. Finally, the proofs of all the key results are given in Section 4.4. This section is organized so that it is possible to proceed directly to Section 4.5 after Section 4.4.

Reformulating and Generalizing de Finetti's Theorem

De Finetti's theorem states that an infinite exchangeable array is a mixture of iid arrays. More usefully for us, Diaconis and Freedman [59] proved that infinite exchangeable arrays can be approximated arbitrarily well by particularly simple ones obtained by sampling the entries of finite random vectors with replacement. Formally, given a finite random vector $(X_1, \dots, X_n)$, one can form the infinite exchangeable array $(X_{i_1}, X_{i_2}, \dots)$ where the indices i_j are drawn iid from $[n]$. In [59] it is shown that infinite exchangeable arrays constructed in this fashion can approximate a target infinite exchangeable array in the sense that the marginal distributions of subsets of m variables from the two arrays are close in total variation for $m \le n$. (We state this result more formally in the sequel.)

Observe that different random vectors give rise to the same exchangeable array under the preceding sampling scheme, as sampling with replacement from $(X_1, \dots, X_n)$ yields the same distribution as sampling from $(X_{\pi(1)}, \dots, X_{\pi(n)})$ for any permutation $\pi \in \mathsf{S}_n$. Further, sampling from $(X_1, \dots, X_n)$ yields the same distribution as sampling from $(X_1, \dots, X_1, \dots, X_n, \dots, X_n)$ where each entry is repeated the same number of times. Thus, yet another equivalent formulation of the above result from [59] is that any infinite exchangeable array can be approximated by those obtained from *equivalence classes* of random vectors via iid sampling of their entries. We generalize this version of de Finetti's theorem by generalizing the three components of the result, namely, what an infinite exchangeable array is, how we can obtain such arrays by sampling the entries of finite random vectors, and which finite vectors are equivalent to each other in the sense that they yield the same infinite exchangeable arrays in this way.

Generalizing infinite exchangeable arrays: An infinite exchangeable array $(X_1, X_2, , \dots)$ corresponds to a sequence of measures $(\mathrm{Law}(X_1), \mathrm{Law}(X_1, X_2), \dots)$ that project onto each other, with each measure in the sequence being invariant under permutations. This is analogous to the freely-described elements of Definition 4.2.3 in which sequences of invariants project onto each other. We similarly define freely-described measures, which generalize infinite exchangeable arrays.

Definition 4.4.1 (Freely-described measures)**.** *Let $\mathscr{V}$ be a consistent sequence, and let $(\Omega_m \subseteq \mathbb{V}_m)$ be a sequence of subsets satisfying $\varphi^\star_{M,m}\Omega_M \subseteq \Omega_m$ for all $m \le M$. The collection of* freely-described probability measures *supported on (Ω_m) is*

$$\varprojlim_{\mathscr{V}} \mathcal{P}(\Omega_m)^{\mathsf{S}_m} = \left\{ (\mu_m) \in \prod_m \mathcal{P}(\Omega_m)^{\mathsf{S}_m} : \varphi^\star_{M,m}\mu_M = \mu_m \text{ for all } m \le M \right\}, \quad (4.17)$$

where $\varphi^{\star}_{M,m}\mu_M$ is the pushforward of μ_M under $\varphi^{\star}_{M,m}$.

Example 4.4.2 (Freely-described measures over $\mathscr{Z}$)**.** *The distributions of infinite exchangeable arrays correspond precisely to freely-described measures over $\mathscr{Z}$. Concretely, a freely-described measure over $\mathscr{Z}$ defines a sequence of S_m-invariant random vectors $X^{(m)} = (X_1, \dots, X_m)$ such that $\zeta^{\star}_{m+1,m}X^{(m+1)} \overset{d}{=} X^{(m)}$ for all m, where $\zeta^{\star}_{m+1,m}$ extracts the first m entries. Conversely, if $(X_1, X_2, \dots)$ is an infinite exchangeable array then the laws of the truncations $X^{(m)} = (X_1, \dots, X_m)$ form a freely-described measure over $\mathscr{Z}$.*

Generalizing sampling with replacement: Sampling the entries of a vector may be interpreted via maps between finite sets. For a map $f\colon [m] \to [n]$ between finite sets, consider the associated linear map $\rho(f)\colon \mathbb{R}^n \to \mathbb{R}^m$ acting by $(\rho(f)x)_i = x_{f(i)}$. If $F_{n,m}\colon [m] \to [n]$ is a uniformly random map between finite sets, then each $i \in [m]$ is mapped independently to a uniformly random element $F_{n,m}(i) \in [n]$, so that $\rho(F_{n,m})x \in \mathbb{R}^m$ is a vector whose entries are sampled with replacement from those of $x \in \mathbb{R}^n$. Sampling with replacement from the entries of a random vector $X^{(n)} = (X_1, \dots, X_n)$ then yields the freely-described measure $\left(\mathrm{Law}(\rho(F_{n,m})X^{(n)})\right)_{m\in\mathbb{N}}$ where $X^{(n)}$ is independent of $F_{n,m}$. The key observation to generalize de Finetti's theorem is that the above action $\rho(f)$ of maps between finite sets on $(\mathbb{R}^n)$ extends to an action on any sequence of vector spaces obtained from $(\mathbb{R}^n)$ via the standard constructions from Section 4.2 as follows.

(Products) For sequences of vector spaces $(\mathbb{W}_n)$, $(\mathbb{U}_n)$, setting $\rho(f)(v, u) = (\rho(f)w, \rho(f)u)$ defines an action on $(\mathbb{W}_n \times \mathbb{U}_n)$

(Tensors and Polynomials) For sequences $(\mathbb{W}_n)$, $(\mathbb{U}_n)$, setting $\rho(f)(w \otimes u) = (\rho(f)w) \otimes (\rho(f)u)$ defines an action on $(\mathbb{W}_n \otimes \mathbb{U}_n)$, which restricts to symmetric tensors to give an action on $(\mathrm{Sym}^k\mathbb{W}_n)$.

As an illustration, we get an action on the spaces $((\mathbb{R}^n)^{\otimes d})$ via $(\rho(f)X)_{i_1,\dots,i_d} = X_{f(i_1),\dots,f(i_d)}$. In this manner, maps between finite sets yield linear maps between vector spaces constituting any $\mathscr{Z}$- or $\mathscr{D}$-sequence.

Generalizing equivalence under sampling: Finally, fix a random vector $X^{(n)} = (X_1, \dots, X_n)$ and recall that sampling with replacement from any vector in the sequence $(g_N \cdot X^{(n)} \otimes \mathbb{1}_{N/n})_{n|N}$ for any $g_N \in \mathsf{S}_N$ yields the same infinite exchangeable array. Therefore, from the perspective of our sampling map, all measures in the sequence $(\mathrm{Law}(G_N \cdot X^{(n)} \otimes \mathbb{1}_{N/n}))_{n|N}$ should be viewed as equivalent; here $G_N \sim$

$\mathrm{Haar}(\mathsf{S}_N)$ is independent of $X^{(n)}$. These sequences of exchangeable measures that are related by embeddings are analogous to freely-symmetrized elements from Definition 4.2.3. We similarly define freely-symmetrized measures, which capture the ambiguity in our sampling maps more generally.

Definition 4.4.3 (Freely-symmetrized measures). *Let $\mathscr{V}$ be a consistent sequence, and let $(\Omega_n \subseteq \mathbb{V}_n)$ be a sequence of subsets satisfying $\varphi_{N,n}\Omega_n \subseteq \Omega_N$ for all $n \leq N$. The collection of* freely-symmetrized probability measures *over (Ω_n) is*

$$\varinjlim_{\mathscr{V}} \mathcal{P}(\Omega_n)^{\mathsf{S}_n} = \bigsqcup_n \mathcal{P}(\Omega_n)^{\mathsf{S}_n} \quad \text{where } \mu_n \sim \mu_N \text{ if } \mu_N = \mathrm{sym}_N^{\mathscr{V}} \mu_n. \tag{4.18}$$

A freely-symmetrized measure is one of these equivalence classes, denoted as a sequence $(\mathrm{sym}_N^{\mathscr{V}} \mu_n)_{N \geq n}$ for $\mu_n \in \mathcal{P}(\Omega_n)^{\mathsf{S}_n}$.

Example 4.4.4 (Freely-symmetrized measures over $\mathscr{D}$). *The exchangeable random vectors yielding the same infinite exchangeable array under sampling with replacement correspond to freely-symmetrized measures over $\mathscr{D}$. More precisely, a freely-symmetrized measure over $\mathscr{D}$ is the equivalence class of $\left(\mathrm{sym}_N^{\mathscr{D}} \mathrm{Law}(X^{(n)}) = \mathrm{Law}(G_N X^{(n)} \otimes \mathbb{1}_{N/n})\right)_{n|N}$ with $G_N \sim \mathrm{Haar}(\mathsf{S}_N)$ independent of $X^{(n)}$.*

With these notions in hand, we can restate de Finetti's theorem in terms of free measures. Specifically, consider the following map from freely-symmetrized measures $\varinjlim_{\mathscr{D}} \mathcal{P}(\mathbb{R}^n)^{\mathsf{S}_n}$ over $\mathscr{D}$ to freely-described measures $\varprojlim_{\mathscr{Z}} \mathcal{P}(\mathbb{R}^m)^{\mathsf{S}_m}$ over $\mathscr{Z}$:

$$\left(\mathrm{sym}_N^{\mathscr{D}} \mu_n\right)_{n|N} \mapsto \left(\mathrm{Law}(\rho(F_{n,m})X\right)_m, \tag{4.19}$$

where $X \sim \mu_n$ is independent of the uniformly random map $F_{n,m} \colon [m] \to [n]$. Diaconis and Freedman [59] prove the error bound $\|\mathrm{Law}(\zeta_{n,m}^{\star} X) - \mathrm{Law}(\rho(F_{n,m})X)\|_{\mathrm{TV}} \leq \frac{m(m-1)}{n}$ whenever $m \leq n$, which holds for any exchangeable random vector X in $\mathbb{R}^n$; here $\zeta_{n,m}^{\star} X$ is the vector of the first m coordinates of X, while $\rho(F_{n,m})X$ is a vector of m entries sampled with replacement from the coordinates of X. From this bound, one can conclude that the map (4.19) has a dense image in $\varprojlim_{\mathscr{Z}} \mathcal{P}(\mathbb{R}^m)^{\mathsf{S}_m}$ in total variation, meaning that if (μ_m) is a freely-described measure over $\mathscr{Z}$, then there exists a sequence of freely-described measures $(\mu_m^{(n)})$ in the image of the map (4.19) such that $\lim_{n\to\infty} \|\mu_m - \mu_m^{(n)}\|_{\mathrm{TV}} = 0$ for each fixed m. That is because for each $n \geq m$ and $X \sim \mu_n$ we have $\mu_m = \mathrm{Law}(\zeta_{n,m}^{\star} X)$ since (μ_m) is freely-described, while the image of $\left(\mathrm{sym}_N^{\mathscr{D}} \mu_n\right)_{n|N}$ under the map (4.19) is $(\mu_m^{(n)} = \mathrm{Law}(\rho(F_{n,m})X))_m$, so $\|\mu_m - \mu_m^{(n)}\|_{\mathrm{TV}} \to 0$ as $n \to \infty$ by the above error bound. To obtain de Finetti's characterization of *all* freely-described measures on $\mathscr{Z}$, not just the dense subset above, one can use a simple compactness argument [59, Thm. 14].

For more general consistent sequences, the characterization of all freely-described measures is more complicated [115, Chap. 7]. The map (4.19) does, however, readily generalize and yields simple dense subsets of freely-described measures, as we show in the following theorem. Fortunately, these dense subsets, together with error bounds generalizing those of [59], are all we require to construct finite-dimensional POP lower bounds and obtain their convergence rates (see Section 4.5).

Theorem 4.4.5. *Let $\mathcal{F}(\mathscr{Z}) = \{(\mathbb{N}, \leq), (\mathsf{S}_n), (\mathbb{V}_n), (\varphi_{N,n})\}$ be a $\mathscr{Z}$-sequence, and let $\mathcal{F}(\mathscr{D}) = \{(\mathbb{N}, \cdot|\cdot), (\mathsf{S}_n), (\mathbb{V}_n), (\psi_{N,n})\}$ be the corresponding $\mathscr{D}$-sequence. Let $(\Omega_n \subseteq \mathbb{V}_n)$ be a sequence of S_n-invariant sets satisfying $\varphi_{N,n}^{\star}\Omega_N \subseteq \Omega_n$ whenever $n \leq N$ and $\psi_{N,n}\Omega_n \subseteq \Omega_N$ whenever $n|N$. Consider the map from freely-symmetrized measures $\varinjlim_{\mathcal{F}(\mathscr{D})} \mathcal{P}(\Omega_n)^{\mathsf{S}_n}$ over $\mathcal{F}(\mathscr{D})$ to freely-described measures $\varprojlim_{\mathcal{F}(\mathscr{Z})} \mathcal{P}(\Omega_m)^{\mathsf{S}_m}$ over $\mathcal{F}(\mathscr{Z})$ sending*

$$\left(\mathrm{sym}_N^{\mathcal{F}(\mathscr{D})}\mu_n\right)_{n|N} \mapsto \left(\mathrm{Law}(\rho(F_{n,m})X)\right)_m, \tag{4.20}$$

where $X \sim \mu_n$ is independent of the uniformly random map $F_{n,m}\colon [m] \to [n]$. This map has a dense image in total variation, meaning that for any freely-described measure $(\mu_m) \in \varprojlim_{\mathscr{Z}} \mathcal{P}(\Omega_n)^{\mathsf{S}_n}$ there exist $(\mu_m^{(n)})$ in the image of (4.20) such that $\lim_{n\to\infty} \|\mu_m - \mu_m^{(n)}\|_{\mathrm{TV}} = 0$ for each m. More precisely, for any $\mu_n \in \mathcal{P}(\mathbb{V}_n)^{\mathsf{S}_n}$ we have $\|\mathrm{Law}(\varphi_{n,m}^{\star}X) - \mathrm{Law}(\rho(F_{n,m})X)\|_{\mathrm{TV}} \leq \frac{m(m-1)}{n}$ where $X \sim \mu_n$ is independent of $F_{n,m}$.

Note that the error bound above holds for any exchangeable measure on any fixed $\mathbb{V}_n$, not necessarily compactly supported. The proof of this result is provided at the end of Section 4.4 as a consequence of the more general Theorem 4.4.15. We instantiate Theorem 4.4.5 in the setting of the original de Finetti's theorem as well as for two-dimensional arrays.

Example 4.4.6 (Classic de Finetti)**.** *With $\mathcal{F} = \mathrm{id}$ and $\Omega_n = \{0,1\}^n$ in Theorem 4.4.5, the set $\varprojlim_{\mathscr{Z}} \mathcal{P}(\Omega_m)^{\mathsf{S}_m}$ corresponds to infinite binary exchangeable arrays, the original setting studied by de Finetti [56]. Theorem 4.4.5 states that a dense subset of such arrays is given by starting from a random exchangeable vector $X^{(n)} \in \{0,1\}^n$ and sampling its entries with replacement to get $(X_1, X_2, \ldots)$; equivalently, if $P^{(n)} \in \{0, 1/n, 2/n, \ldots, 1\}$ is the (random) fraction of the entries of $X^{(n)}$ which equal 1, we can first sample $P^{(n)}$ and then sample X_j iid from $\mathrm{Ber}(P^{(n)})$. In comparison, de*

Finetti proved that all infinite binary exchangeable arrays are obtained by sampling $P \in [0,1]$ *according to some distribution on* $[0,1]$*, and then sampling* X_j *iid from* $\mathrm{Ber}(P)$*. Thus, Theorem 4.4.5 yields the dense subset of these arrays for which the random* P *is supported on* $\{0, 1/n, 2/n, \dots, 1\}$ *for some* $n \in \mathbb{N}$*.*

Example 4.4.7 (Graphons and Aldous–Hoover)**.** *Let* $\mathcal{F} = \mathrm{Sym}^2$ *in Theorem 4.4.5, so* $\mathcal{F}(\mathscr{Z}) = \mathrm{Sym}^2\mathscr{Z}$ *and* $\mathcal{F}(\mathscr{D}) = \mathrm{Sym}^2\mathscr{D}$ *are the consistent sequences from Example 4.2.2 consisting of symmetric matrices on which permutations act by simultaneously permuting rows and columns, and with zero-padding and duplication embeddings. Next, let* $\Omega_n = \mathbb{S}^n(\{0,1\})$*, which we view as adjacency matrices of unweighted undirected graphs. An exchangeable measure on* Ω_n *is the law of a random symmetric binary matrix* $(X_{i,j})_{i,j\in[n]}$*, which can be viewed as a random graph on n vertices. Freely-described measures on* (Ω_n) *can be viewed as infinite exchangeable two-dimensional binary arrays, or equivalently, as a sequence of random graphs on increasingly-many vertices closed under taking induced subgraphs. Theorem 4.4.5 states that a dense subset of such sequences of measures is obtained by starting from a fixed-dimensional random* $X \in \Omega_n$ *and sampling its row and column indices* $i_1, i_2, \dots$ *iid from* $[n]$ *to obtain the infinite array* $(X_{i_j,i_{j'}})_{j,j'\in\mathbb{N}}$*. Equivalently, define the step-function* $W_X\colon [0,1]^2 \to \{0,1\}$ *by first partitioning* $[0,1] = \bigcup_{i=1}^n I_{i,n}$ *with* $I_{i,n} = [(i-1)/n, i/n)$ *if* $i < n$ *and* $I_{n,n} = [1-1/n, 1]$*, and setting* $W_X(x,y) = X_{i,j}$ *if* $(x,y) \in I_{i,n} \times I_{j,n}$*. Then sampling* $x_1, x_2, \dots$ *iid uniformly on* $[0,1]$ *and setting* $X_{j,j'} = W_X(x_j, x_{j'})$ *yields an infinite array with the same distribution as before.*

To obtain all *freely-described measures on* (Ω_n)*, it was shown in [148] that any infinite exchangeable binary array* $(X_{j,j'})$ *is obtained from a random symmetric measurable function* $W\colon [0,1]^2 \to [0,1]$ *called a* graphon *by first sampling* $x_1, x_2, \dots$ *iid uniformly from* $[0,1]$ *and then sampling* $X_{j,j'} \sim \mathrm{Ber}(W(x_j, x_{j'}))$ *iid for* $j, j' \in \mathbb{N}$*. Thus, Theorem 4.4.5 yields the dense subset of such arrays that can be obtained from* step graphons *of the form* W_X *for some* $X \in \Omega_k$*.*

For more general sequences of sets $(\Omega_n \subseteq \mathbb{S}^n)$*, corresponding to weighted graphs, the characterization of all freely-described measures on* (Ω_n) *is even more involved, and requires random functions defined on* $[0,1]^2 \times [0,1]^{\binom{\mathbb{N}}{2}}$*, as a consequence of Aldous–Hoover–Kallenberg theory [115, Chap. 7]. Nevertheless, freely-described measures obtained from step graphons as above continue to be dense by Theorem 4.4.5, and the error bound there still applies.*

Theorem 4.4.5 gives a dense subset of freely-described measures over $\mathscr{Z}$-sequences in terms of freely-symmetrized measures over $\mathscr{D}$-sequences. This result allows us to construct lower bounds for freely-described problems over $\mathscr{D}$-sequences in terms of freely-symmetrized problems over $\mathscr{L}$-sequences, generalizing the example in Section 4.1. To construct bounds for freely-described problems over $\mathscr{L}$-sequences, however, we need a dual de Finetti theorem. In more detail, we require a dense subset of freely-described measures over $\mathscr{D}$-sequences in terms of freely-symmetrized measures over the corresponding $\mathscr{L}$-sequences.

Example 4.4.8 (Freely-described measures over $\mathscr{D}$)**.** *A freely-described measure over $\mathscr{D}$ corresponds to the laws of a sequence of exchangeable random vectors $(X^{(m)} \in \mathbb{R}^m)_m$ satisfying*

$$\sum_{j=1}^{M/m} X^{(M)}_{(i-1)(M/m)+j} \overset{d}{=} X^{(m)}_i, \quad \text{for all } m|M \text{ and } i \in [m]. \tag{4.21}$$

Examples of such sequences of random variables include $X^{(m)} \sim \mathrm{Unif}\{e_1, \ldots, e_m\}$ being a uniformly random coordinate vector, the sequence of iid Gaussians $X^{(m)} \sim \mathcal{N}(0, \frac{1}{m} I_m)$, and the sequence of Dirichlet distributions $X^{(m)} \sim \mathrm{Dir}(\mathbb{1}_m/m)$.

We obtain the desired 'dual' version of Theorem 4.4.5 by considering the adjoint of the action ρ of maps between finite sets. Specifically, for any map $f\colon [n] \to [m]$ we get the linear map $\rho(f)^\star\colon \mathbb{R}^n \to \mathbb{R}^m$ defined by $(\rho(f)^\star x)_i = \sum_{j \in f^{-1}(i)} x_j$. This extends to an action between the vector spaces constituting any $\mathscr{L}$- or $\mathscr{D}$-sequence as before. Considering the adjoint action of a uniformly random map, we obtain the following theorem.

Theorem 4.4.9. *Let $\mathcal{F}(\mathscr{D}) = \{(\mathbb{N}, \cdot|\cdot), (\mathsf{S}_n), (\mathbb{V}_n), (\psi_{N,n})\}$ be a $\mathscr{D}$-sequence of degree $d_{\mathcal{F}}$, and let $\mathcal{F}(\mathscr{L}) = \{(\mathbb{N}, \leq), (\mathsf{S}_n), (\mathbb{V}_n), (\varphi_{N,n})\}$ be the corresponding $\mathscr{L}$-sequence. Let $(\Omega_n \subseteq \mathbb{V}_n)$ be a sequence of compact S_n-invariant sets satisfying $\psi^\star_{N,n}\Omega_N \subseteq \Omega_n$ whenever $n|N$ and $\varphi_{N,n}\Omega_n \subseteq \Omega_N$ whenever $n \leq N$. Consider the map from freely-symmetrized measures $\varinjlim_{\mathcal{F}(\mathscr{L})} \mathcal{P}(\Omega_n)^{\mathsf{S}_n}$ over $\mathcal{F}(\mathscr{L})$ to freely-described measures $\varprojlim_{\mathcal{F}(\mathscr{D})} \mathcal{P}(\Omega_m)^{\mathsf{S}_m}$ over $\mathcal{F}(\mathscr{D})$ sending*

$$\left(\mathrm{sym}_N^{\mathcal{F}(\mathscr{L})} \mu_n\right)_{N \geq n} \mapsto \left(\mathrm{Law}(\rho(F_{m,n})^\star X)\right)_m, \tag{4.22}$$

where $X \sim \mu_n$ is independent of the uniformly random map $F_{m,n}\colon [n] \to [m]$. This map has a dense image in the Wasserstein distance W_1 with respect to an induced $\|\cdot\|_1$ norm on $(\mathbb{V}_n)$. More precisely, for any $\mu_n \in \mathcal{P}(\mathbb{V}_n)^{\mathsf{S}_n}$ and $m|n$ we have

$$W_1\left(\mathrm{Law}(\psi^\star_{n,m} X), \mathrm{Law}(\rho(F_{m,n})^\star X)\right) \leq 2 d_{\mathcal{F}} \sqrt{\frac{m(m-1)}{n/m}} \mathbb{E}_{\mu_n}[\|X\|_1],$$

where $X \sim \mu_n$ is independent of $F_{m,n}$.

Note that the error bound in Theorem 4.4.9 holds for any exchangeable measure on $\mathbb{V}_n$, but is stated in Wasserstein-1 distance with respect to a particular choice of norms on $(\mathbb{V}_n)$. These norms are certain extensions of the 1-norms on $(\mathbb{R}^n)$ to spaces obtained via standard constructions from them (see (4.30) and Lemma 4.4.21 below). In contrast, the error bounds in Theorem 4.4.5 are given in the larger total variation distance, and decay to zero at the rate of $1/n$, faster than the present $1/\sqrt{n}$. We show in Sections 4.4 and 4.4 that the use of Wasserstein-1 here is necessary, and that both rates are tight. The proof of Theorem 4.4.9 is provided at the end of Section 4.4 as a consequence of the more general Theorem 4.4.20. We illustrate Theorem 4.4.9 on two examples, one corresponding to the freely-described measures on $\mathscr{D}$ from Example 4.4.8 and the other corresponding to a dual version of Aldous–Hoover.

Example 4.4.10 (Freely-described measures on $\mathscr{D}$, Example 4.4.8 continued)**.** *In the setting of Example 4.4.8, Theorem 4.4.9 states that any sequence of random vectors satisfying 4.21 and supported on simplices $\Omega_n = \Delta^{n-1}$ can be approximated arbitrarily well by those of the form $X^{(m)} = \rho(F_{m,n})^\star(X) = \sum_{i=1}^n X_i e_{j_i}$ where $X \in \Delta^{n-1}$ is a fixed random vector supported on the simplex and $j_1, \dots, j_n$ are sampled iid uniformly from $[m]$ (independently of X). That is, the right notion of sampling in this setting puts each coordinate of $x \in \mathbb{R}^n$ independently into one of m bins uniformly at random, and each coordinate of $\rho(F_{m,n})^\star x$ is obtained by summing the coordinates of x in the corresponding bin. While Theorem 4.4.9 yields a dense subset of all freely-described measures over $\mathscr{D}$ supported on (Δ^{n-1}), a full characterization of all such measures is not available to the best of our knowledge.*

Example 4.4.11 (Freely-described measures on $\mathrm{Sym}^2\mathscr{D}$)**.** *Setting $\mathscr{F} = \mathrm{Sym}^2$ and $\Omega_n = \{X \in \mathbb{S}^n : X_{i,j} \geq 0,\ \sum_{i,j} X_{i,j} = 1\}$ in Theorem 4.4.9, freely-described measures over $\mathrm{Sym}^2\mathscr{D}$ can be approximated arbitrarily well by the following construction. A finite random graph $X \in \Omega_n$ defines a freely-described measure over $\mathrm{Sym}^2\mathscr{D}$ by defining the sequence of random graphs $(X^{(m)})$ via $X^{(m)} = \sum_{1\leq i\leq j\leq n} X_{i,j} E_{k_i,k_j}$ where $E_{i,j} = e_i e_j^\top + e_j e_i^\top$ and $k_1, \dots, k_n$ are drawn iid uniformly from $[m]$. Equivalently, $X^{(m)}$ is the random graph obtained by labeling the n vertices of X by uniformly random labels from $[m]$ and identifying vertices with the same labels (which might then result in self-loops and multiple edges). As before, Theorem 4.4.9 yields a dense subset of freely-described measures over $\mathrm{Sym}^2\mathscr{D}$ supported on (Ω_n), and we are not aware of a full characterization of all such measures.*

The key insight underlying the proofs of Theorems 4.4.5 and 4.4.9 is a link between actions of maps between finite sets and consistent sequences. Specifically, the above actions of maps between finite sets recover all of the maps relating vector spaces in different dimensions that we have seen so far (e.g., zero-padding and duplication), and as a consequence, yield all $\mathscr{Z}$- and $\mathscr{D}$-sequences. Leveraging this link, we reduce the proofs of the above theorems to proving bounds on random maps between finite sets.

Freely-Described Measures Over $\mathscr{Z}$-Sequences

The aim of this section is to present a generalization of de Finetti's theorem that characterizes (a dense subset of) freely-described measures over $\mathscr{Z}$-sequences. We begin by abstracting away the various properties of the action ρ of maps between finite sets on $(\mathbb{R}^n)$ from Section 4.4. Note that if $f_1\colon [n] \to [m]$ and $f_2\colon [m] \to [k]$, then $\rho(f_2 \circ f_1) = \rho(f_1) \circ \rho(f_2)$ and $\rho(\mathrm{id}_{[n]}) = \mathrm{id}_{\mathbb{R}^n}$. We call such an assignment of linear maps to maps between finite sets an FinSet$^{\mathrm{op}}$-action, where FinSet stands for finite sets and 'op' stands for 'opposite', and pertains to the opposite order in which compositions act.[4]

Definition 4.4.12 (FinSet$^{\mathrm{op}}$-representation)**.** *An* FinSet$^{\mathrm{op}}$-action *on a sequence of vector spaces* $(\mathbb{V}_n)_{n\in\mathbb{N}}$ *is an assignment of a linear map* $\tau^{\mathrm{op}}(f)\colon \mathbb{V}_m \to \mathbb{V}_n$ *for each map* $f\colon [n] \to [m]$ *between finite sets satisfying* $\tau^{\mathrm{op}}(\mathrm{id}_{[n]}) = \mathrm{id}_{\mathbb{V}_n}$ *for all* $n \in \mathbb{N}$ *and* $\tau^{\mathrm{op}}(f_1 \circ f_2) = \tau^{\mathrm{op}}(f_2) \circ \tau^{\mathrm{op}}(f_1)$ *whenever the composition is well-defined. A sequence of vector spaces endowed with an* FinSet$^{\mathrm{op}}$*-action* $\mathtt{V} = \{(\mathbb{V}_n), (\tau^{\mathrm{op}})\}$ *is called an* FinSet$^{\mathrm{op}}$-representation.

We denote the standard FinSet$^{\mathrm{op}}$ representation defined earlier on $(\mathbb{R}^n)$ by $\mathtt{R} = \{(\mathbb{R}^n), (\rho)\}$. Next, we illustrate how FinSet$^{\mathrm{op}}$-representations yield consistent sequences. Starting with the representation $\mathtt{R}$, we observe that specific choices of maps f between finite sets yield many of the maps we have been using in our development.

(Permutations) If $\sigma\colon [n] \to [n]$ is a permutation, then $\rho(\sigma) = \sigma^{-1}$ gives the action of σ^{-1} on $\mathbb{R}^n$.

(Zero-padding) If $\iota_{N,n}\colon [n] \hookrightarrow [N]$ is inclusion, then $\rho(\iota_{N,n}) = \zeta^{\star}_{N,n}$ is the projection in $\mathscr{Z}$ extracting the first n entries of a length-N vector.

[4] In fact, an FinSet$^{\mathrm{op}}$-action precisely corresponds to a representation of the category FinSet$^{\mathrm{op}}$, the opposite of the category FinSet whose objects are finite sets and whose morphisms are all maps between them.

(Duplication) If $d_{n,nk}\colon [nk] \to [n]$ is the standard equipartition $d_{n,nk}(k(i-1)+j) = i$ for $i \in [n]$, $j \in [k]$, then $\rho(d_{n,nk}) = \delta_{nk,n}$ is the duplication embedding in $\mathscr{D}$.

Consequently, the consistent sequences $\mathscr{L}$ and $\mathscr{D}$ can be derived from $\mathtt{R}$ as follows:

$$\mathscr{L} = \{(\mathbb{N}, \leq), (\mathsf{S}_n)_{\mathscr{L}}, (\mathbb{R}_n), (\rho(\iota_{N,n})^\star)_{n\leq N}\} \quad \text{and} \quad \mathscr{D} = \{(\mathbb{N}, \cdot|\cdot), (\mathsf{S}_n)_{\mathscr{D}}, (\mathbb{R}_n), (\rho(d_{n,N}))_{n|N}\},$$

where $(\mathsf{S}_n)_{\mathscr{L}}$ and $(\mathsf{S}_n)_{\mathscr{D}}$ denote the chains of symmetric groups from (4.9) and (4.10), respectively. We can generalize this observation to obtain $\mathscr{L}$- and $\mathscr{D}$-sequences from any $\mathsf{FinSet}^{\mathrm{op}}$-representation. Specifically, given a $\mathsf{FinSet}^{\mathrm{op}}$-representation $\mathtt{V} = \{(\mathbb{V}_n), (\tau^{\mathrm{op}})\}$, for each $n \in \mathbb{N}$ we endow $\mathbb{V}_n$ with the S_n action $\sigma \cdot x = \tau^{\mathrm{op}}(\sigma^{-1})x$ for a permutation $\sigma\colon [n] \to [n]$ and $x \in \mathbb{V}_n$. We then define the two consistent sequences:

$$\begin{aligned} \mathscr{L}(\mathtt{V}) &= \{(\mathbb{N}, \leq), (\mathsf{S}_n)_{\mathscr{L}}, (\mathbb{V}_n), (\tau^{\mathrm{op}}(\iota_{N,n})^\star)_{n\leq N}\} \quad \text{and} \\ \mathscr{D}(\mathtt{V}) &= \{(\mathbb{N}, \cdot|\cdot), (\mathsf{S}_n)_{\mathscr{D}}, (\mathbb{V}_n), (\tau^{\mathrm{op}}(d_{n,N}))_{n|N}\}. \end{aligned} \tag{4.23}$$

Any $\mathscr{L}$- and $\mathscr{D}$-sequence can be obtained in this way for some $\mathsf{FinSet}^{\mathrm{op}}$-representation. To see this, recall that the standard constructions of Section 4.2, which can be applied to $\mathscr{L}$ and $\mathscr{D}$ to obtain more complicated consistent sequences, can also be applied to $\mathtt{R}$ as discussed in Section 4.4 to obtain more complicated $\mathsf{FinSet}^{\mathrm{op}}$-representations. In particular, for any standard construction $\mathcal{F}$, we observe that $\mathcal{F}(\mathtt{R})$ is also an $\mathsf{FinSet}^{\mathrm{op}}$-representation. Moreover, one can check that $\mathscr{L}(\mathcal{F}(\mathtt{R})) = \mathcal{F}(\mathscr{L})$ and $\mathscr{D}(\mathcal{F}(\mathtt{R})) = \mathcal{F}(\mathscr{D})$. That is, we get the same consistent sequences from $\mathcal{F}$ whether we apply it directly to $\mathscr{L}$ and $\mathscr{D}$, or apply it to the $\mathsf{FinSet}^{\mathrm{op}}$-representation $\mathtt{R}$ and invoke (4.23).

We proceed to generalize de Finetti's theorem to free measures over consistent sequences of the form (4.23) derived from any $\mathsf{FinSet}^{\mathrm{op}}$-representation. The proofs of our results in this section are presented in Section 4.4 to make the architecture of our argument more transparent. We begin with a lemma characterizing the invariances of uniformly random maps. It will be used to argue that our sampling maps send equivalent freely-symmetrized measures to the same image, and that these images are indeed freely-described measures.

Lemma 4.4.13. *Let $F_{m,n}\colon [n] \to [m]$ be uniformly random maps for each $m, n \in \mathbb{N}$. For any injective map $\phi_{n,k}\colon [k] \to [n]$ we have $F_{m,n} \circ \phi_{n,k} \overset{d}{=} F_{m,k}$, and for any equipartition $\Psi_{k,nk}\colon [nk] \to [k]$ we have $\Psi_{k,nk} \circ F_{nk,m} \overset{d}{=} F_{k,m}$ for all $k, m, n \in \mathbb{N}$.*

Next, we state the key technical lemma used in the proof of Theorem 4.4.15, which is the quantitative version of the asymptotic equivalence of sampling with and without replacement from [208], stated in our formalism. In what follows, if $\mu, \nu \in \mathcal{P}(\Omega)$ are two measures on a finite set Ω, then their total variation distance is $\|\mu - \nu\|_{\mathrm{TV}} = \sum_{x\in\Omega} |\mu(x) - \nu(x)| = \sup_{f\colon \Omega\to[-1,1]} |\mathbb{E}_\mu f - \mathbb{E}_\nu f|$.

Lemma 4.4.14. *For $m \le n$, let $F_{n,m}\colon [m] \to [n]$ be a uniformly random map, and let $\Phi_{n,m}\colon [m] \hookrightarrow [n]$ be a uniformly random* injective *map. Then*

$$\|\mathrm{Law}(F_{n,m}) - \mathrm{Law}(\Phi_{n,m})\|_{\mathrm{TV}} \le \frac{m(m-1)}{n}. \tag{4.24}$$

The above total variation bound applies to any pushforward of the random maps $F_{n,m}$ and $\Phi_{n,m}$, and in particular applies to their linear representatives acting on an $\mathrm{FinSet}^{\mathrm{op}}$-representation, so that $\|\mathrm{Law}(\tau^{\mathrm{op}}(F_{n,m})) - \mathrm{Law}(\tau^{\mathrm{op}}(\Phi_{n,m}))\|_{\mathrm{TV}} \le \frac{m(m-1)}{n}$. This allows us to prove the following theorem, generalizing de Finetti to consistent sequences derived from any $\mathrm{FinSet}^{\mathrm{op}}$-representation.

Theorem 4.4.15 (De Finetti for $\mathrm{FinSet}^{\mathrm{op}}$-representations)**.** *Let $\mathtt{V} = \{(\mathbb{V}_n), (\tau^{\mathrm{op}})\}$ be an* $\mathrm{FinSet}^{\mathrm{op}}$ *representation, and let (Ω_n) satisfy $\tau^{\mathrm{op}}(f)\Omega_n \subseteq \Omega_m$ for all $f\colon [m] \to [n]$. Consider the map from $\varinjlim_{\mathscr{D}(\mathtt{V})} \mathcal{P}(\Omega_n)^{\mathsf{S}_n}$ to $\varprojlim_{\mathscr{Z}(\mathtt{V})} \mathcal{P}(\Omega_m)^{\mathsf{S}_m}$ sending*

$$\left(\mathrm{sym}_N^{\mathscr{D}(\mathtt{V})}\mu_n\right)_{n|N} \mapsto \left(\mathrm{Law}(\tau^{\mathrm{op}}(F_{n,m})X)\right)_m. \tag{4.25}$$

where $X \sim \mu_n$ is independent of the uniformly random map $F_{n,m}\colon [m] \to [n]$. This map has a dense image in total variation. More precisely, for any $\mu_n \in \mathcal{P}(\mathbb{V}_n)^{\mathsf{S}_n}$ we have

$$\|\mathrm{Law}(\tau^{\mathrm{op}}(\iota_{n,m})X) - \mathrm{Law}(\tau^{\mathrm{op}}(F_{n,m})X)\|_{\mathrm{TV}} \le \frac{m(m-1)}{n},$$

where $X \sim \mu_n$ is independent of $F_{n,m}$.

The condition that $\tau^{\mathrm{op}}(f)\Omega_n \subseteq \Omega_m$ for all $f\colon [m] \to [n]$ in Theorem 4.4.15 is equivalent to the sets (Ω_n) satisfying (i) each Ω_n is S_n-invariant; (ii) $\tau^{\mathrm{op}}(\iota_{N,n})\Omega_N \subseteq \Omega_n$ for $n \le N$; and (iii) $\tau^{\mathrm{op}}(d_{n,N})\Omega_n \subseteq \Omega_N$ for $n|N$. In other words, such sequences of sets (Ω_n) are precisely the ones which are group-invariant, embed into each other in $\mathscr{D}(\mathtt{V})$, and project into each other in $\mathscr{Z}(\mathtt{V})$. This is a consequence of the following lemma, showing that any map between finite sets factorizes into an injection and an equipartition.

Lemma 4.4.16 (Factorization of maps between finite sets)**.** *For any map $f\colon [m] \to [n]$ and any $k \geq \max_{i\in[n]} |f^{-1}(i)|$, there exists $g \in \mathsf{S}_{nk}$ satisfying $f = d_{n,nk}g\iota_{nk,m}$.*

Finally, we conclude this section with an example illustrating that the error bounds obtained in Theorem 4.4.15 are tight.

Example 4.4.17 (Tightness of rates in Theorem 4.4.15)**.** *To see that the $1/n$ rate in Theorem 4.4.15 is tight, let $X \sim \mathrm{Ber}(1/2)^{\otimes n}$, and note that $\zeta^\star_{2,n}X \sim \mathrm{Ber}(1/2)^{\otimes 2}$ while $\rho(F_{n,2})X \sim (1-1/n)\mathrm{Ber}(1/2)^{\otimes 2} + (1/n)\mathrm{Unif}\{(0,0)^\top, (1,1)^\top\}$, since with probability $1-1/n$ the map $\rho(F_{n,2})$ picks out two distinct coordinates from X, while with probability $1/n$ the two coordinates are equal and equally likely to be 0 or 1. The total variation distance between the two laws is therefore $2/n$.*

Freely-Described Measures Over $\mathscr{D}$-Sequences

In this section we present our dual de Finetti theorem giving a dense subset of freely-described measures over $\mathscr{D}$-sequences in terms of freely-symmetrized measures over $\mathscr{Z}$-sequences. To this end, we consider the adjoint action $\rho(f)^\star$ of maps between finite sets. This adjoint action preserves the order of compositions, so $\rho(f_1\circ f_2)^\star = \rho(f_1)^\star\circ\rho(f_2)^\star$ and once again we have $\rho(\mathrm{id}_{[n]})^\star = \mathrm{id}_{\mathbb{V}_n}$. We call such an assignment of linear maps to maps between finite sets an FinSet-representation.

Definition 4.4.18 (FinSet-representation)**.** *An* FinSet-action *on a sequence of vector spaces $(\mathbb{V}_n)_{n\in\mathbb{N}}$ is an assignment of a linear map $\tau(f)\colon \mathbb{V}_n \to \mathbb{V}_m$ for each map $f\colon [n] \to [m]$ between finite sets satisfying $\tau(\mathrm{id}_{[n]}) = \mathrm{id}_{\mathbb{V}_n}$ and $\tau(f_1\circ f_2) = \tau(f_1)\circ\tau(f_2)$ whenever the composition is well-defined. A sequence of vector spaces endowed with an* FinSet-*action* $\mathtt{V} = \{(\mathbb{V}_n), (\tau)\}$ *is called an* FinSet-representation.

As the adjoint $\rho^\star$ of the standard $\mathrm{FinSet}^{\mathrm{op}}$-action ρ is an FinSet-action, we denote the standard FinSet-representation on $(\mathbb{R}^n)$ from Section 4.4 by $\mathtt{R}^\star = \{(\mathbb{R}^n), (\rho^\star)\}$. More broadly, note that if $\mathtt{V} = \{(\mathbb{V}_n), (\tau)\}$ is an FinSet-representation, then $\mathtt{V}^\star = \{(\mathbb{V}_n), (\tau^\star)\}$ is an $\mathrm{FinSet}^{\mathrm{op}}$-representation (and vice versa), where $\tau^\star$ is the adjoint of τ. Consequently, any FinSet-representation $\mathtt{V}$ yields the consistent sequences $\mathscr{Z}(\mathtt{V}^\star)$ and $\mathscr{D}(\mathtt{V}^\star)$, by appealing to (4.23). As with $\mathrm{FinSet}^{\mathrm{op}}$-representations, any standard construction $\mathcal{F}$ can be applied to the FinSet-representation $\mathtt{R}^\star$ to obtain an $\mathrm{FinSet}^{\mathrm{op}}$-representation $\mathcal{F}(\mathtt{R}^\star)$, and we again have that $\mathscr{Z}(\mathcal{F}(\mathtt{R}^\star)^\star) = \mathcal{F}(\mathscr{Z})$ and $\mathscr{D}(\mathcal{F}(\mathtt{R}^\star)^\star) = \mathcal{F}(\mathscr{D})$.

We proceed to present our dual de Finetti theorem. The architecture of the proof is similar to Section 4.4, which we again highlight by deferring the proofs and some technical lemmas to Section 4.4 below. The following is the key result we will need, which is the analog of Lemma 4.4.14 above.

Lemma 4.4.19. *Let* $F_{m,n}\colon [n] \to [m]$ *be a uniformly random map and let* $\Psi_{m,n}\colon [n] \to [m]$ *be a uniformly random equipartition, for any* $m|n$*. Then*

$$W_1(\mathrm{Law}(F_{m,n}), \mathrm{Law}(\Psi_{m,n})) = \inf_{\substack{(F,\Psi)\ \text{random} \\ F\overset{d}{=}F_{m,n},\ \Psi\overset{d}{=}\Psi_{m,n}}} \mathbb{E}\left[\frac{|F \neq \Psi|}{n}\right] \le \sqrt{\frac{m(m-1)}{n/m}}, \quad (4.26)$$

where $|F \neq \Psi| = |\{i \in [n] : F(i) \neq \Psi(i)\}|$*.*

Lemma 4.4.19 gives a weaker bound than Lemma 4.4.14. Indeed, not only do we get an $n^{-1/2}$ rate here as opposed to n^{-1} in Lemma 4.4.14, but also the present bound is in Wasserstein distance whereas the previous one is in total variation. In fact, we have $\|\mathrm{Law}(F_{m,n}) - \mathrm{Law}(\Psi_{m,n})\|_{\mathrm{TV}} \not\to 0$ as $n/m \to \infty$, since the distribution of the fiber sizes of the two maps satisfy

$$\left\|\mathrm{Multinomial}(n, m, \mathbb{1}_m/m) - \delta_{(n/m)\mathbb{1}_m}\right\|_{\mathrm{TV}} = 1 - 2^{-n}\binom{n}{n/m, \ldots, n/m} \to 1. \quad (4.27)$$

Thus, the two random maps in Lemma 4.4.19 are only close when compared in a specific metric between such maps, namely, $\mathrm{dist}(f, \psi) = |f \neq \psi|/n$. As a consequence, our dual de Finetti theorem only applies to FinSet-representations on which the FinSet action is appropriately Lipschitz-continuous with respect to this metric. The following theorem is the FinSet analogue of Theorem 4.4.15.

Theorem 4.4.20 (Dual de Finetti for FinSet-representations)**.** *Let* $\mathtt{V} = \{(\mathbb{V}_n), (\tau)\}$ *be an* FinSet*-representation endowed with norms* $\|\cdot\|$ *on each* $\mathbb{V}_n$ *such that for some* $L > 0$ *we have*

$$\min_{g\in \mathsf{S}_n} \|(\tau(f) - \tau(h))gx\| \le L\frac{|f \neq h|}{n}\|x\|, \quad (4.28)$$

for any $x \in \mathbb{V}_n$*, any* $f, h\colon [n] \to [m]$*, and any* n, m*. Let* (Ω_n) *be a sequence of* compact *sets satisfying* $\tau(f)\Omega_m \subseteq \Omega_n$ *for all* $f\colon [m] \to [n]$*. Consider the map from* $\varinjlim_{\mathscr{I}(\mathtt{V}^\star)} \mathcal{P}(\Omega_n)^{\mathsf{S}_n}$ *to* $\varprojlim_{\mathscr{D}(\mathtt{V}^\star)} \mathcal{P}(\Omega_m)^{\mathsf{S}_m}$ *sending*

$$(\mathrm{sym}_N^{\mathscr{I}(\mathtt{V}^\star)}\mu_n)_{N\ge n} \mapsto \Big(\mathrm{Law}(\tau(F_{m,n})X)\Big)_m \quad (4.29)$$

where $X \sim \mu_n$ *is independent of the uniformly random map* $F_{m,n}\colon [n] \to [m]$*. This map has a dense image in* W_1*-distance with respect to* $\|\cdot\|$*. More precisely, for any* $\mu_n \in \mathcal{P}(\mathbb{V}_n)^{\mathsf{S}_n}$ *and any* $m|n$ *we have*

$$W_1(\mathrm{Law}(\tau(d_{m,n})X), \mathrm{Law}(\tau(F_{m,n})X)) \le L\sqrt{\frac{m(m-1)}{n/m}}\mathbb{E}_{\mu_n}[\|X\|],$$

where $X \sim \mu_n$ is independent of $F_{m,n}$.

Once again, Lemma 4.4.16 shows that the condition that $\tau(f)\Omega_m \subseteq \Omega_n$ for all $f\colon [m] \to [n]$ in Theorem 4.4.20 is equivalent to the sets (Ω_n) being S_n-invariant, embedding into each other in $\mathscr{Z}(\mathtt{V}^\star)$, and projecting into each other in $\mathscr{D}(\mathtt{V}^\star)$.

To apply the above theorem to relate freely-symmetrized measures on $\mathscr{Z}$-sequences to freely-described measures on the corresponding $\mathscr{D}$-sequences, we require a choice of norm satisfying (4.28). We show in Lemma 4.4.23 in Section 4.4 that the ℓ_1 norm on $(\mathbb{R}^n)$ satisfies (4.28) with $L = 2$ for the standard FinSet-action $\rho^\star$. Moreover, we extend this ℓ_1 norm to any vector spaces derived from $(\mathbb{R}^n)$ via a standard construction $\mathcal{F}$ from Section 4.2, and we prove in Lemma 4.4.21 that these extended norms also satisfy (4.28) for the associated FinSet-action. Defining these extensions for any standard construction $\mathcal{F}$ proceeds in the usual way. In particular, for any normed spaces $\mathbb{V}$ and $\mathbb{U}$, we endow $\mathbb{V}\times\mathbb{U}$ and $\mathbb{V}\otimes\mathbb{U}$ with the following norms:

$$\|(v,u)\|_{\mathbb{V}\times\mathbb{U}} = \|v\|_{\mathbb{V}} + \|u\|_{\mathbb{U}} \qquad \|X\|_{\mathbb{V}\otimes\mathbb{U}} = \inf\left\{\sum_i \|v_i\|_{\mathbb{V}}\|u_i\|_{\mathbb{U}} : X = \sum_i v_i\otimes u_i\right\} \tag{4.30}$$

The norm on $\mathbb{V}^{\otimes d}$ restricts to a norm on $\mathrm{Sym}^d\mathbb{V}$. Further, if $\{e_i^{(\mathbb{V})}\}_{i\in[n]}$ and $\{e_j^{(\mathbb{U})}\}_{j\in[m]}$ are bases for $\mathbb{V}$ and $\mathbb{U}$, then $\{e_i^{(\mathbb{V})}\}_{i\in[n]} \cup \{e_j^{(\mathbb{U})}\}_{j\in[m]}$ and $\{e_i^{(\mathbb{V})} \otimes e_j^{(\mathbb{U})}\}_{i\in[n],j\in[m]}$ are extended bases for $\mathbb{V}\times\mathbb{U}$ and for $\mathbb{V}\otimes\mathbb{U}$, respectively. With these notions in hand, the following lemma shows that (4.28) is satisfied by the extended ℓ_1 norm on the vector spaces underlying $\mathcal{F}(\mathtt{R}^\star)$ for any standard construction $\mathcal{F}$.

Lemma 4.4.21. *Let $\mathcal{F}(\mathtt{R}^\star) = \{(\mathbb{V}_n), (\rho^\star)\}$ be obtained from the standard* FinSet-*representation* $\mathtt{R}^\star$ *via a standard construction $\mathcal{F}$ of degree $d_{\mathcal{F}}$, and let $\rho^\star$ be the induced* FinSet-*action. Let $\|\cdot\|_{\ell_1}$ be the ℓ_1 norm on $(\mathbb{V}_n)$ derived from the ℓ_1 norm on $(\mathbb{R}^n)$ via (4.30) and let $\{e_\alpha\}_{\alpha\in\mathcal{A}_n}$ be a basis for $(\mathbb{V}_n)$ that is extended from the canonical one for $\mathbb{R}^n$. Then $\|x\|_1 = \sum_{\alpha\in\mathcal{A}_n}|x_\alpha|$ where $x = \sum_{\alpha\in\mathcal{A}_n} x_\alpha e_\alpha$. Further, for each $n\in\mathbb{N}$ and $x\in\mathbb{V}_n$, we have $\|\rho(f)^\star x\|_1 \le \|x\|_1$ and*

$$\min_{g\in\mathsf{S}_n}\|(\rho(f)^\star - \rho(h)^\star)gx\|_1 \le \frac{1}{|\mathsf{S}_n|}\sum_{g\in\mathsf{S}_n}\|(\rho(f)^\star-\rho(h)^\star)gx\|_1 \le 2d_{\mathcal{F}}\frac{|f\neq h|}{n}\|x\|_1, \tag{4.31}$$

for all $f,h\colon [n]\to[m]$.

For example, if we endow $\mathbb{R}^n$ with the ℓ_1 norm, we get the entrywise ℓ_1 norms on $\mathbb{R}^{d\times n}$ and $(\mathbb{R}^n)^{\otimes d}$ for all d. These norms satisfy (4.31) with $d_{\mathcal{F}} = 1$ and d, respectively. Combining Lemma 4.4.21 with Theorem 4.4.20, we obtain Theorem 4.4.9

giving a dense subset of freely-described measures over any $\mathscr{D}$-sequence in terms of freely-symmetrized measures over the corresponding $\mathscr{L}$-sequences. Finally, we conclude this section with an example illustrating that the error bounds obtained in Theorem 4.4.20 are tight.

Example 4.4.22 (Tightness of rates in Theorem 4.4.20)**.** *To see that the use of the Wasserstein distance is necessary and that the rate of $1/\sqrt{n}$ in Theorem 4.4.20 is sharp, let $X_{2n} = \frac{1}{2n}\mathbb{1}_{2n}$ be a deterministic invariant vector. Then $\delta^{\star}_{2n,2}X_{2n} = \frac{1}{2}\mathbb{1}_2$ while $\rho(F_{2,2n})^{\star}X_{2n} = \left(\frac{B}{2n}, 1 - \frac{B}{2n}\right)^{\top}$ where $B \sim \mathrm{Bin}(2n, 1/2)$. Then their laws do not converge in total variation by (4.27), while the W_1 distance between their laws is $\frac{1}{n}\mathbb{E}|B - n| \sim 1/\sqrt{\pi n}$ by [9].*

Proofs for Section 4.4

We now prove the results stated earlier in the section, as well as a few additional technical lemmas needed for these proofs.

Proofs for Section 4.4: We begin by proving the invariances of a uniformly random map.

Proof (Lemma 4.4.13). Note that $F_{m,n}(i)$ is a uniformly random element of $[m]$, independent for different i. If $\phi_{n,k}$ is injective, then each $i \in [k]$ is mapped under $F_{m,n} \circ \phi_{n,k}$ to an independent and uniformly random element of $[m]$, proving the first claim. If $\psi_{k,nk}$ is an equipartition, then a uniformly random element of $[nk]$ is equally likely to fall in each of the equally-sized fibers of $\psi_{k,nk}$. Therefore, each $i \in [m]$ is mapped by $\psi_{k,nk}F_{nk,m}$ to an independent and uniformly random element of $[k]$, proving the second claim. □

Next, we prove the total variation bound between random maps and random injections.

Proof (Lemma 4.4.14). Identify a map $[m] \to [n]$ as an element of $[n]^m$. Then $F_{n,m}$ corresponds to the element of $[n]^m$ obtained by sampling m elements from $[n]$ uniformly with replacement, while $\Phi_{n,m}$ corresponds to sampling without replacement. The total variation between them is bounded by (4.24) in [208, Eq. (2.3)]. □

We now use the above lemma to prove our generalized de Finetti theorem for $\mathrm{FinSet}^{\mathrm{op}}$-representations.

Proof (Theorem 4.4.15). First, two measures defining the same freely-symmetrized element have the same images under the given map since $\tau^{\mathrm{op}}(F_{nk,m})\mathrm{sym}_{\mathsf{S}_{nk}}\tau^{\mathrm{op}}(d_{n,nk}) = \frac{1}{|\mathsf{S}_{nk}|}\sum_{g\in\mathsf{S}_{nk}}\tau^{\mathrm{op}}(d_{n,nk}gF_{nk,m}) \overset{d}{=} \tau^{\mathrm{op}}(F_{n,m})$ by Lemma 4.4.13. Similarly, the images of the given map are indeed freely-described since $\tau^{\mathrm{op}}(\iota_{M,m})\tau^{\mathrm{op}}(F_{n,M}) = \tau^{\mathrm{op}}(F_{n,M}\iota_{M,m}) \overset{d}{=} \tau^{\mathrm{op}}(F_{n,m})$ by the same lemma.

Second, we prove the given total variation estimate for a fixed-dimensional exchangeable measure. To this end, if $\Phi_{n,m}\colon [m] \to [n]$ is a uniformly random injection, note that $\Phi_{n,m} \overset{d}{=} G\iota_{n,m}$ where $G \sim \mathrm{Haar}(\mathsf{S}_n)$. Therefore, if $\mu_n \in \mathcal{P}(\Omega_n)^{\mathsf{S}_n}$ is S_n-invariant and $X_n \sim \mu_n$ is independent of $\Phi_{n,m}$, then $\tau^{\mathrm{op}}(\Phi_{n,m})X_n \overset{d}{=} \tau^{\mathrm{op}}(\iota_{n,m})X_n$. Since $\mathrm{Law}(\tau^{\mathrm{op}}(\Phi_{n,m})X_n)$ and $\mathrm{Law}(\tau^{\mathrm{op}}(F_{n,m})X_n)$ are pushforwards of $\mathrm{Law}(\Phi_{n,m}) \otimes \mu_n$ and $\mathrm{Law}(F_{n,m}) \otimes \mu_n$, respectively, and since the total variation distance between pushforwards can only decrease, Lemma 4.4.14 gives

$$\|\mathrm{Law}(\tau^{\mathrm{op}}(\Phi_{n,m})X_n)-\mathrm{Law}(\tau^{\mathrm{op}}(F_{n,m})X_n)\|_{\mathrm{TV}} \leq \|\mathrm{Law}(\Phi_{n,m})-\mathrm{Law}(F_{n,m})\|_{\mathrm{TV}} \leq \frac{m(m-1)}{n}, \tag{4.32}$$

as claimed.

Finally, we show that the image of the given map is dense in total variation. Let $(\mu_m) \in \varprojlim_{\mathscr{Z}(\mathbb{V})} \mathcal{P}(\Omega_m)^{\mathsf{S}_m}$. For each $n \in \mathbb{N}$ consider the image of $(\mathrm{sym}_N^{\mathscr{D}(\mathbb{V})}\mu_n)_{n|N} \in \varinjlim_{\mathscr{D}(\mathbb{V})} \mathcal{P}(\Omega_n)^{\mathsf{S}_n}$ under the given map, which yields the sequence of freely-described measures $(\mu_m^{(n)} = \mathrm{Law}(\tau^{\mathrm{op}}(F_{n,m})X_n))_m$ where $X_n \sim \mu_n$ is independent of $F_{n,m}$. Since (μ_m) is freely-described over $\mathscr{Z}(\mathbb{V})$, for each $n \geq m$ we have $\mu_m = \tau^{\mathrm{op}}(\iota_{n,m})\mu_n$. Therefore, for each fix $m \in \mathbb{N}$ we have

$$\|\mu_m^{(n)}-\mu_m\|_{\mathrm{TV}} = \|\mathrm{Law}(\tau^{\mathrm{op}}(F_{n,m})X_n)-\mathrm{Law}(\tau^{\mathrm{op}}(\iota_{n,m})X_n)\|_{\mathrm{TV}} \leq \frac{m(m-1)}{n} \xrightarrow{n\to\infty} 0. \tag{4.33}$$

This yields the claimed density. □

To obtain Theorem 4.4.5 from Theorem 4.4.15, we need to reformulate the condition that $\tau^{\mathrm{op}}(f)\Omega_n \subseteq \Omega_m$ for all $f\colon [m] \to [n]$. This was done above using Lemma 4.4.16 which we now prove.

Proof (Lemma 4.4.16). Denote the fibers of f by $f^{-1}(i) = \{n_{i,1} < \ldots < n_{i,k_i}\}$ for $i \in [n]$. Since $k \geq \max_i k_i$ we have $m = \sum_{i=1}^n k_i \leq kn$. Define the injection $\phi\colon [m] \hookrightarrow [nk]$ by $\phi(n_{i,\ell}) = (i-1)k+\ell$, which is well-defined since $1 \leq \ell \leq k_i \leq k$. Since ϕ is injective, there exists $g \in \mathsf{S}_{nk}$ satisfying $\phi = g\iota_{nk,m}$, and by construction $f = d_{n,nk} \circ \phi = d_{n,nk}g\iota_{nk,m}$. □

Proofs for Section 4.4: We begin by proving the Wasserstein bound between random maps and random equipatitions.

Proof (Lemma 4.4.19). Consider the following coupling of $F_{m,n}$ and $\Psi_{m,n}$. Let $(N_1, \ldots, N_m) \sim \text{Multinomial}(n, m, \mathbb{1}_m/m)$ be a sample from a multinomial distribution. Let $N_{(1)} \geq \ldots \geq N_{(m)}$ be these values in sorted order. Next, split $[n]$ into consecutive intervals of length $N_{(1)}, \ldots, N_{(m)}$ and define $F_0 \colon [n] \to [m]$ to take value i on the ith interval. Formally,

$$F_0(j) = i \quad \text{if } j \in \left\{ \sum_{\ell=1}^{i-1} N_{(\ell)} + 1, \ldots, \sum_{\ell=1}^{i} N_{(\ell)} \right\}. \tag{4.34}$$

Finally, let $G \sim \text{Haar}(\mathsf{S}_n)$ and set $F = F_0 \circ G$ and $\Psi = d_{m,n} \circ G$. Note that $F \overset{d}{=} F_{m,n}$ since $(|F_{m,n}^{-1}(1)|, \ldots, |F_{m,n}^{-1}(m)|) \overset{d}{=} (N_1, \ldots, N_m)$ and both F and $F_{m,n}$ give a uniformly random ordered partition of $[n]$ with these part sizes. We also have $\Psi \overset{d}{=} \Psi_{m,n}$ since equipartitions are precisely maps of the form $d_{m,n} g$ for some $g \in \mathsf{S}_n$.

For the above coupling, we have

$$|F \neq \Psi| = |F_0 \neq d_{m,n}| = \sum_{i=1}^{m} |F_0^{-1}(i) \setminus d_{m,n}^{-1}(i)| \leq \sum_{i=1}^{m} \sum_{\ell=1}^{i} (N_{(\ell)} - \tfrac{n}{m}) \leq m \sum_{i=1}^{m} |N_i - \tfrac{n}{m}|.$$

Indeed, observe that $\sum_{j=1}^{i} N_{(j)} \geq in/m$ for all i, otherwise we must have $N_{(\ell)} < n/m$ for all $\ell \geq i$ and $\sum_{j=1}^{m} N_{(j)} < in/m + \sum_{\ell=i+1}^{m} N_{(\ell)} < n$, a contradiction. Therefore, we have

$$F_0^{-1}(i) \setminus d_{m,n}^{-1}(i) = \left\{ \max\left(in/m, \sum_{\ell=1}^{i-1} N_{(\ell)} \right) + 1, \ldots, \sum_{\ell=1}^{i} N_{(\ell)} \right\},$$

so

$$|F_0^{-1}(i) \setminus d_{m,n}^{-1}(i)| = \sum_{\ell=1}^{i} N_{(\ell)} - \max\left(in/m, \sum_{\ell=1}^{i-1} N_{(\ell)} \right) \leq \sum_{\ell=1}^{i} (N_{(\ell)} - \tfrac{n}{m}).$$

Finally, using the variance of the coordinates of the multinomial distribution, we get

$$\mathbb{E}\left[\frac{|F \neq \Psi|}{n} \right] \leq \frac{1}{n/m} \sum_{i=1}^{m} \mathbb{E}[|N_i - \tfrac{n}{m}|^2]^{1/2} = \frac{m}{n/m} \sqrt{n(1/m)(1-1/m)} = \sqrt{\frac{m(m-1)}{n/m}}, \tag{4.35}$$

yielding the claimed bound. $\square$

Using the above estimate, we can then prove our dual de Finetti theorem.

Proof (Theorem 4.4.20). Two measures defining the same freely-symmetrized element map to the same image since $\tau(F_{m,N})\mathrm{sym}_{\mathsf{S}_N}\tau(\iota_{N,n}) \overset{d}{=} F_{m,n}$ by Lemma 4.4.13. Similarly, the image of the given map is freely-described since $\tau(d_{m,mk})\tau(F_{mk,n}) \overset{d}{=} \tau(F_{m,n})$ by the same lemma.

Next, for any $\mu_n \in \mathcal{P}(\Omega_n)^{\mathsf{S}_n}$, sample $X_n \sim \mu_n$ and independently $F_{m,n}$, and let $G_* \in \operatorname{argmin}_{g\in\mathsf{S}_n} \|(\tau(d_{m,n}) - \tau(F_{m,n}))gX_n\|$. Then $(\tau(d_{m,n})G_*X_n, \tau(F_{m,n})G_*X_n)$ is a coupling of $\tau(d_{m,n})X_n$ and $\tau(F_{m,n})X_n$ since $gX_n \overset{d}{=} X_n$ for any $g \in \mathsf{S}_n$, so

$$\begin{aligned} W_1(\mathrm{Law}(\tau(d_{m,n})X_n), \mathrm{Law}(\tau(F_{m,n})X_n)) &\le \mathbb{E}[\|(\tau(d_{m,n}) - \tau(F_{m,n}))g_*X\|] \\ &= \mathbb{E}\left[\min_{g\in\mathsf{S}_n} \|(\tau(d_{m,n}) - \tau(F_{m,n}))gX\|\right] \\ &\le L\mathbb{E}\left[\|X\|\frac{|d_{m,n} \neq F_{m,n}|}{n}\right] \le L\sqrt{\frac{m(m-1)}{n/m}}\mathbb{E}_{X\sim\mu_n}\|X\|. \end{aligned}$$

Finally, for any $(\mu_m) \in \varprojlim_{\mathscr{D}(\mathbb{V})} \mathcal{P}(\Omega_m)^{\mathsf{S}_m}$ consider the sequence of freely-symmetrized elements $(\mathrm{sym}_N^{\mathscr{I}(\mathbb{V})}\mu_{n!})$ and their images $(\mathrm{Law}(\tau(F_{m,n!})X_{n!}))_m$ under the given map. For any fixed m and all $n \ge m$ we have

$$W_1(\mu_m, \mathrm{Law}(\tau(F_{m,n!})X_{n!})) = W_1(\mathrm{Law}(\tau(d_{m,n!})X_{n!}), \mathrm{Law}(\tau(F_{m,n!})X_{n!})) \le L\sqrt{\frac{m(m-1)}{n!/m}}\max_{x\in\Omega_m}\|x\|,$$

which vanishes as $n \to \infty$, showing that the image of the given map is indeed dense. □

To instantiate this theorem on $(\mathbb{R}^n)$ and sequences derived from it, we prove that the ℓ_1 norm satisfies a stronger version of (4.28).

Lemma 4.4.23. *Endow $(\mathbb{R}^n)$ with its standard* FinSet *action and ℓ_1 norms. Let $n, m \in \mathbb{N}$ be arbitrary and consider any $x \in \mathbb{R}^n$. Then $\|\rho(f)^\star x\|_1 \le \|x\|_1$ and*

$$\min_{g\in\mathsf{S}_n} \|(\rho(f)^\star - \rho(h)^\star)gx\|_1 \le \frac{1}{|\mathsf{S}_n|}\sum_{g\in\mathsf{S}_n} \|(\rho(f)^\star - \rho(h)^\star)gx\|_1 \le 2\frac{|f \neq h|}{n}\|x\|_1, \tag{4.36}$$

for all $f, h\colon [n] \to [m]$. In particular (4.28) is satisfied with $L = 2$.

Proof. The first claim follows since

$$\|\rho(f)^\star x\|_1 = \sum_{i=1}^m \left|\sum_{j\in f^{-1}(i)} x_j\right| \le \sum_{i=1}^m \sum_{j\in f^{-1}(i)} |x_j| = \|x\|_1.$$

The first inequality in (4.36) simply bounds the minimum over S_n by the average. For the second inequality, we have

$$\|(\rho(f)^\star - \rho(h)^\star)gx\|_1 = \sum_{i=1}^m \left| \sum_{j\in f^{-1}(i)} x_{g(j)} - \sum_{j\in h^{-1}(i)} x_{g(j)} \right| \le \sum_{i=1}^m \sum_{j\in f^{-1}(i)\triangle h^{-1}(i)} |x_{g(j)}|$$
$$= 2\sum_{j\in\{f\neq h\}} |x_{g(j)}|,$$

where the first equality follows from the definition of the FinSet action and the last equality follows since each $j \in \{f \neq h\}$ satisfies $j \in f^{-1}(f(j))\triangle h^{-1}(f(j))$ and $j \in f^{-1}(h(j))\triangle h^{-1}(h(j))$. Since S_n acts transitively on subsets of $[n]$ of size $|f \neq h|$, we get

$$\frac{1}{|\mathsf{S}_n|}\sum_{g\in\mathsf{S}_n} \|(\rho(f)^\star - \rho(h)^\star)gx\|_1 = 2\frac{1}{\binom{n}{|f\neq g|}} \sum_{\substack{S\subseteq[n]\\|S|=|f\neq h|}} \sum_{j\in S} |x_j| = 2\frac{\binom{n-1}{|f\neq h|-1}}{\binom{n}{|f\neq h|}}\|x\|_1 = 2\frac{|f\neq h|}{n}\|x\|_1,$$

where the second equality follows since each $j \in [n]$ is contained in $\binom{n-1}{|f\neq h|-1}$ subsets of size $|f \neq h|$. This proves the claim. $\square$

We now extend the ℓ_1 norm to any sequence $(\mathcal{F}(\mathbb{R}^n))$ using (4.30), and show that it is also an ℓ_1 norm with respect to the bases obtained from the same definition. Further, this extended ℓ_1 norm satisfies similar estimates as the ℓ_1 norm itself.

Proof (Lemma 4.4.21). We induct on the constructions $\mathcal{F}$. For $\mathcal{F} = \mathrm{id}$ this is trivial. Suppose that the claim holds for $(\mathcal{F}_1(\mathbb{R}^n))$ and $(\mathcal{F}_2(\mathbb{R}^n))$, with bases $(e_\alpha)_{\alpha\in\mathcal{A}_n}$ and $(e_\beta)_{\beta\in\mathcal{B}_n}$. For any $(x,y) \in \mathcal{F}_1(\mathbb{R}^n)\times\mathcal{F}_2(\mathbb{R}^n)$, we have $(x,y) = \sum_{\alpha\in\mathcal{A}_n} x_\alpha(e_\alpha, 0) + \sum_{\beta\in\mathcal{B}_n} y_\beta(0, e_\beta)$, and $\|(x,y)\| = \|x\| + \|y\| = \sum_{\alpha\in\mathcal{A}_n}|x_\alpha| + \sum_{\beta\in\mathcal{B}_n}|y_\beta|$. Next, write $X \in \mathcal{F}_1(\mathbb{R}^n)\otimes\mathcal{F}_2(\mathbb{R}^n)$ as $X = \sum_{\alpha,\beta} X_{\alpha,\beta}e_\alpha\otimes e_\beta$. On the one hand, $\|X\| \le \sum_{\alpha,\beta}|X_{\alpha,\beta}|$ by definition of $\|X\|$. On the other hand, for any $\epsilon > 0$ let $X = \sum_i x_i \otimes y_i$ such that $\sum_i \|x\|\|y\| \le \|X\|+\epsilon$. Then $X_{\alpha,\beta} = \sum_i x_{i,\alpha}y_{i,\beta}$ and $\sum_{\alpha,\beta}|X_{\alpha,\beta}| \le \sum_{i,\alpha,\beta}|x_{i,\alpha}|\cdot|y_{i,\beta}| = \sum_i \|x\|\|y\| \le \|X\| + \epsilon$ which proves the claim as $\epsilon > 0$ was arbitrary.

To obtain estimates on these extended ℓ_1 norms, we again induct on the constructions $\mathcal{F}$ we apply to $(\mathbb{R}^n)$. Suppose the result is true for a degree k sequence $(\mathbb{V}_n = \mathcal{F}_1(\mathbb{R}^n))$ and for a degree ℓ sequence $(\mathbb{U}_n = \mathcal{F}_2(\mathbb{R}^n))$. We prove that it holds for $(\mathbb{V}_n \times \mathbb{U}_n)$ and $(\mathbb{V}_n \otimes \mathbb{U}_n)$, which have degrees $\max\{k,\ell\}$ and $k+\ell$, respectively. The latter also implies that the result holds for $(\mathrm{Sym}^\ell\mathbb{V}_n)$ for any ℓ. For the direct product, we have

$$\|\rho(f)^\star(v,u)\|_1 = \|(\rho(f)^\star v, \rho(f)^\star u)\|_1 = \|\rho(f)^\star v\|_1 + \|\rho(f)^\star u\|_1 \le \|v\|_1 + \|u\|_1 = \|(v,u)\|_1,$$

and similarly $\|(\rho(f)^\star - \rho(h)^\star)g(v,u)\|_1 = \|(\rho(f)^\star - \rho(h)^\star)gv\|_1 + \|(\rho(f)^\star - \rho(h)^\star)gu\|_1$ so

$$\begin{aligned}
&\frac{1}{|\mathsf{S}_n|}\sum_{g\in\mathsf{S}_n}\|(\rho(f)^\star - \rho(h)^\star)g(v,u)\|_1\\
&= \frac{1}{|\mathsf{S}_n|}\sum_{g\in\mathsf{S}_n}(\|(\rho(f)^\star - \rho(h)^\star)gv\|_1 + \|(\rho(f)^\star - \rho(h)^\star)gu\|_1)\\
&\le 2\frac{|f\neq h|}{n}(k\|v\|_1 + \ell\|u\|_1) \le 2\max\{k,\ell\}\frac{|f\neq h|}{n}\|(v,u)\|_1.
\end{aligned}$$

For the tensor product, if $X \in \mathbb{V}_n \otimes \mathbb{U}_n$ can be written as $X = \sum_i v_i \otimes u_i$ then

$$\begin{aligned}
\|\rho(f)^\star X\|_1 \le \sum\nolimits_i \|\rho(f)^\star(v_i\otimes u_i)\|_1 &= \sum\nolimits_i \|(\rho(f)^\star v_i)\otimes(\rho(f)^\star u_i)\|_1\\
&= \sum\nolimits_i \|\rho(f)^\star v_i\|_1\|\rho(f)^\star u_i\|_1 \le \sum\nolimits_i \|v_i\|_1\|u_i\|_1,
\end{aligned}$$

so taking the infimum over such representations of X yields the first claim. Next, we have

$$\begin{aligned}
&\|(\rho(f)^\star - \rho(h)^\star)g(v_i\otimes u_i)\|_1 = \|(\rho(f)^\star gv_i)\otimes(\rho(f)^\star gu_i) - (\rho(h)^\star gv_i)\otimes(\rho(h)^\star gu_i)\|_1\\
&\le \|((\rho(f)^\star - \rho(h)^\star)gv_i)\otimes(\rho(f)^\star gu_i)\|_1 + \|(\rho(h)^\star gv_i)\otimes((\rho(f)^\star - \rho(h)^\star)gu_i)\|_1\\
&\le \|(\rho(f)^\star - \rho(h)^\star)gv_i\|_1\|u_i\|_1 + \|v_i\|_1\|(\rho(f)^\star - \rho(h)^\star)gu_i\|_1.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\frac{1}{|\mathsf{S}_n|}\sum_{g\in\mathsf{S}_n}\|(\rho(f)^\star - \rho(h)^\star)gX\|_1 \le \sum\nolimits_i \frac{1}{|\mathsf{S}_n|}\sum_{g\in\mathsf{S}_n}\|(\rho(f)^\star - \rho(h)^\star)gv_i\|_1\|u_i\|_1\\
&+\sum\nolimits_i \frac{1}{|\mathsf{S}_n|}\sum_{g\in\mathsf{S}_n}\|v_i\|_1\|(\rho(f)^\star - \rho(h)^\star)gu_i\|_1 \le 2(k+\ell)\frac{|f\neq h|}{n}\sum\nolimits_i \|v_i\|_1\|u_i\|_1,
\end{aligned}$$

and taking the inf over such representations of X yields the second claim. □

Proofs for Section 4.4: We prove Theorems 4.4.5 and 4.4.9 as a consequence of Theorems 4.4.15 and 4.4.20.

Proof (Theorem 4.4.5). Let $\mathcal{F}(\mathtt{R})$ be the $\mathrm{FinSet}^{\mathrm{op}}$-representation derived from the construction $\mathcal{F}$, where $\mathtt{R}$ is the standard $\mathrm{FinSet}^{\mathrm{op}}$-representation. Next we recall that the consistent sequences $\mathscr{Z}(\mathcal{F}(\mathtt{R}))$ and $\mathscr{D}(\mathcal{F}(\mathtt{R}))$ from (4.23) are equal to $\mathcal{F}(\mathscr{Z})$ and $\mathcal{F}(\mathscr{D})$, respectively. Instantiating Theorem 4.4.15 on the $\mathrm{FinSet}^{\mathrm{op}}$-representation $\mathcal{F}(\mathtt{R})$ yields the desired conclusion. □

Proof (Theorem 4.4.9). The proof follows the same strategy as that of Theorem 4.4.5. For a choice of norm in Theorem 4.4.20, we use the induced ℓ_1 norm of Lemma 4.4.21. This yields the desired conclusion. □

4.5 Bounds for Any-Dimensional Polynomial Problems

In this section, we use Theorems 4.4.5 and 4.4.9 from Section 4.4 to construct lower bounds on the limiting optimal value of a freely-described POP via a freely-symmetrized POP and we prove associated rates of convergence. There is some asymmetry in these two theorems, which leads to different rates for freely-described problems over $\mathscr{Z}$-sequences and over $\mathscr{D}$-sequences. Nevertheless, the construction of the bounds is the same for the two cases, and we seek to provide as unified a treatment as possible. We begin by abstracting away the common features of these two cases into a single definition.

Definition 4.5.1 (De Finetti maps)**.** *Let $(\mathscr{U},\mathscr{L})$ be an* FinSet*-pair on $(\mathbb{V}_n)$ as in Definition 4.3.5 and let $F_{m,n}\colon [n]\to[m]$ be a uniformly random map. Define the* de Finetti maps *on $(\mathscr{U},\mathscr{L})$ to be the random linear maps $L_{m,n}\colon \mathbb{V}_n\to\mathbb{V}_m$ given by*

1. *$L_{m,n}=\rho(F_{n,m})$ if $(\mathscr{U},\mathscr{L})=(\mathcal{F}(\mathscr{D}),\mathcal{F}(\mathscr{Z}))$,*
2. *$L_{m,n}=\rho(F_{m,n})^\star$ if $(\mathscr{U},\mathscr{L})=(\mathcal{F}(\mathscr{Z}),\mathcal{F}(\mathscr{D}))$,*

for some sequence of standard constructions $\mathcal{F}$. Here we have used the action ρ of maps between finite sets on $(\mathbb{V}_n)$ derived in Section 4.4 from the standard action on $(\mathbb{R}^n)$.

We construct bounds for freely-described problems over $\mathscr{U}$ in terms of freely-symmetrized problems over $\mathscr{L}$ using the random maps $L_{m,n}$ and their associated de Finetti theorems. In particular, the generalized de Finetti theorems 4.4.5 and 4.4.9 from Section 4.4 can be summarized as follows. For any FinSet-pair $(\mathscr{U},\mathscr{L})$ and any FinSet-compatible sequence of sets (Ω_n), the map from $\varinjlim_{\mathscr{L}}\mathcal{P}(\Omega_n)^{\mathsf{S}_n}$ to $\varprojlim_{\mathscr{U}}\mathcal{P}(\Omega_n)^{\mathsf{S}_n}$ defined by:

$$(\mathrm{sym}_N^{\mathscr{L}}\mu_m)_{m\leq_{\mathscr{L}}N}\mapsto\big(\mathrm{Law}(L_{m,n}^\star X)\big)_n$$

has a dense image, where $X\sim\mu_m$ is independent of $L_{m,n}$. Further, we have nonasymptotic error bounds (all in the appropriate topology).

Section 4.5 presents our main result on obtaining freely-symmetrized bounds for freely-described POPs. These lower bounds are interpreted in terms of any-dimensional certificates of nonnegativity in Section 4.5. Finally, Section 4.5 gives the proofs of some supporting results that are used in the proof of the main theorem in Section 4.5. This section is organized so that it is possible to proceed directly to Section 4.6 after Section 4.5.

Freely-Symmetrized Bounds for Freely-Described Problems

The goal of this section is to prove the following theorem describing a procedure for constructing freely-symmetrized lower bounds for freely-described problems. We use de Finetti maps to identify freely-symmetrized POPs whose optimal values yield lower bounds on the given freely-describes POPs, and then to derive associated rates of convergence for these bounds by applying Theorems 4.4.5 and 4.4.9 from Section 4.4.

Theorem 4.5.2. *Let* $(\mathscr{U}, \mathscr{L})$ *be an* FinSet*-pair of degree* D*, let* $(\Omega_n \subseteq \mathbb{V}_n)$ *be* FinSet*-compatible and compact, and let* $(p_n) \in \varprojlim_{\mathscr{U}} \mathbb{R}[\mathbb{V}_n]_{\leq d}^{\mathsf{S}_n}$ *be a freely-described polynomial of degree* d*. Consider the freely-described POP* $u_n = \inf_{x \in \Omega_n} p_n(x)$ *with its limiting optimal value* $u_\infty = \inf_n u_n$*. We have the following results.*

(Dual cost) *There is a freely-symmetrized element* $(\mathrm{sym}_n^{\mathscr{L}} q_k)$ *satisfying*

$$p_n(x) = \mathbb{E}[q_k(L_{k,n}x)], \quad \text{for all } x \in \mathbb{V}_n \text{ and all } n \in \mathbb{N}. \tag{4.37}$$

This freely-symmetrized element is unique, meaning that if q_{k_1} *and* q_{k_2} *are two polynomials satisfying* (4.37)*, then* $\mathrm{sym}_n^{\mathscr{L}} q_{k_1} = \mathrm{sym}_n^{\mathscr{L}} q_{k_2}$ *whenever* $n \geq_{\mathscr{L}} k_1, k_2$*. In terms of the above* q_k*, we have*

$$u_\infty = \inf_{(\mu_n) \in \varprojlim_{\mathscr{L}} \mathcal{P}(\Omega_n)^{\mathsf{S}_n}} \mathbb{E}_{\mu_k} q_k. \tag{4.38}$$

(Freely-symmetrized bounds) *Let* q_k *be as above and consider the freely-symmetrized sequence of problems*

$$\ell_n = \inf_{x \in \Omega_n} \mathrm{sym}_n^{\mathscr{L}} q_k,$$

for $k \leq_{\mathscr{L}} n$*. Then* $\ell_n \leq \ell_N \leq u_\infty$ *for all* $k \leq_{\mathscr{L}} n \leq_{\mathscr{L}} N$*, and denoting by* $\ell_\infty = \sup_n \ell_n$ *their limiting optimal value, we have* $\ell_\infty = u_\infty$*.*

(Rate of convergence I) *If* $\mathscr{U}$ *is a* $\mathscr{D}$*-sequence and* $k \leq n$*, then*

$$u_n - \ell_n \leq \frac{k(k-1)}{n} \|q_k\|_{\Omega_k}, \tag{4.39}$$

where $\|q_k\|_{\Omega_k} = \sup_{x \in \Omega_k} |q_k(x)|$*.*

(Rate of convergence II) *If* $\mathscr{U}$ *is a* $\mathscr{Z}$*-sequence and the underlying vector spaces* $(\mathbb{V}_n)$ *are endowed with the 1-norms of Lemma 4.4.21, then*

$$u_n - \ell_n \leq 2D\sqrt{\frac{k(k-1)}{n/k}} \|q_k'\|_{\Omega_k} \max_{x \in \Omega_k} \|x\|_1, \tag{4.40}$$

for all $k|n$ where $\|q_k'\|_{\Omega_k} = \sup_{x\in\Omega_k}\|\nabla q_k\|_\infty$ and $\|\cdot\|_\infty$ is the dual norm to $\|\cdot\|_1$ with respect to the same extended basis as $\|\cdot\|_1$ (see Lemma 4.5.6).

In summary, given an FinSet-pair $(\mathscr{U},\mathscr{L})$ and a freely-described POP defined over a consistent sequence $\mathscr{U}$, we obtain a reformulation (4.38) involving optimization over a sequence of measures that is freely-described over $\mathscr{L}$; this is the promised generalization of (4.3) from the example of Section 4.1. Truncating the sequence of measures in this reformulation then yields the desired lower bounds. For freely-described problems (u_n) over $\mathscr{D}$-sequences, we get freely-symmetrized lower bounds (ℓ_n) over the corresponding $\mathscr{X}$-sequence satisfying

$$\ldots \leq \ell_n \leq \ell_{n+1} \leq \ldots \ell_\infty = u_\infty \leq \ldots \leq u_{nk} \leq u_n \leq \ldots,$$

for all $n,k\in\mathbb{N}$, converging at a rate $u_n - \ell_n \lesssim 1/n$. For freely-described problems (u_n) over $\mathscr{X}$-sequences, we get freely-symmetrized lower bounds (ℓ_n) over the corresponding $\mathscr{D}$-sequence satisfying

$$\ldots \leq \ell_n \leq \ell_{nk} \leq \ldots \ell_\infty = u_\infty \leq \ldots \leq u_{n+1} \leq u_n \leq \ldots,$$

for all $k|n$, converging at a rate $u_{nk} - \ell_{nk} \lesssim 1/\sqrt{n}$ for fixed k as $n\to\infty$. We remark that if (Ω_n) are not compact, the optimal values (ℓ_n) are still monotonically increasing lower bounds for u_∞ as above, but may not converge to u_∞. We also note that while u_n is not a lower bound on u_∞ in general, one could imagine computing u_n and subtracting the right-hand-sides of (4.39) or (4.40) and thereby obtaining lower bounds on u_∞. However, such an approach leads to more conservative lower bounds than ℓ_n by construction, as we illustrate in several examples in Section 4.6.

As Theorem 4.5.2 shows, the main step in our framework is the construction of the fixed polynomial q_k satisfying (4.37) for a given freely-described polynomial (p_n). We explain how to do so for various FinSet-pairs arising in applications in Section 4.6.

We state two intermediate propositions needed to prove Theorem 4.5.2 followed by a proof of the theorem. The proofs of the propositions themselves are deferred to Section 4.5. The first proposition pertains to an isomorphism between freely-symmetrized elements in $\mathscr{L}$ and freely-described elements in $\mathscr{U}$ induced by the de Finetti maps, which facilitates the identification of the dual cost in Theorem 4.5.2.

Proposition 4.5.3. *Let $(\mathscr{U},\mathscr{L})$ be an* FinSet*-pair over a sequence of vector spaces $(\mathbb{V}_n)$. The map from $\varinjlim_{\mathscr{L}} \mathbb{V}_n^{\mathsf{S}_n}$ to $\varprojlim_{\mathscr{U}} \mathbb{V}_n^{\mathsf{S}_n}$ which sends $(\mathrm{sym}_n^{\mathscr{L}} v_k)_{k\leq_{\mathscr{L}} n} \mapsto (\mathbb{E}L_{k,n}^\star v_k)_n$ is an isomorphism, where $L_{k,n}$ is the appropriate de Finetti map on $(\mathscr{U},\mathscr{L})$.*

One consequence of this isomorphism is that projections and embeddings between invariants become isomorphisms starting from some dimension, which also implies that the spaces of freely-described and freely-symmetrized elements over such $\mathscr{V}$ are isomorphic to fixed-dimensional spaces of invariants $\mathbb{V}_k^{\mathsf{S}_k}$ for all large enough k. Formally, for each $k \in \mathbb{N}$ let $\pi_k \colon \varprojlim_{\mathscr{V}} \mathbb{V}_n^{\mathsf{S}_n} \to \mathbb{V}_k^{\mathsf{S}_k}$ be the map sending $(v_n) \mapsto v_k$ and $\varsigma_k \colon \mathbb{V}_k^{\mathsf{S}_k} \to \varinjlim_{\mathscr{V}} \mathbb{V}_n^{\mathsf{S}_n}$ be the map sending $v_k \mapsto (\mathrm{sym}_n^{\mathscr{V}} v_k)_{n \geq_{\mathscr{L}} k}$. We then have the following strengthening of Proposition 4.2.6.

Proposition 4.5.4. *If $\mathscr{V}$ is a $\mathscr{X}$- or a $\mathscr{D}$-sequence of degree $d_{\mathscr{V}}$, then* $\dim \varprojlim \mathbb{V}_n^{\mathsf{S}_n} = \dim \varinjlim \mathbb{V}_n^{\mathsf{S}_n} = \dim \mathbb{V}_{d_{\mathscr{V}}}^{\mathsf{S}_{d_{\mathscr{V}}}}$. *Moreover, there exists $k \in \mathbb{N}$ such that the maps π_n and ς_n are isomorphisms for all $n \geq k$, and the map $\mathrm{sym}_{\mathsf{S}_N} \varphi_{N,n} \colon \mathbb{V}_n^{\mathsf{S}_n} \to \mathbb{V}_N^{\mathsf{S}_N}$ and $\varphi_{N,n}^{\star} \colon \mathbb{V}_N^{\mathsf{S}_N} \to \mathbb{V}_n^{\mathsf{S}_n}$, are isomorphisms whenever $N \geq n \geq k$.*

The results [139, Prop. 2.9, Thm. 2.11] show that when $\mathscr{V}$ is a $\mathscr{X}$-sequence, we can take $k = d_{\mathscr{V}}$ in Proposition 4.5.4. We conjecture that this is true for $\mathscr{D}$-sequences as well:

Conjecture: When $\mathscr{V}$ is a $\mathscr{D}$-sequence of degree $d_{\mathscr{V}}$, the maps π_n and ς_n are isomorphisms for all $n \geq k$, and the maps $\mathrm{sym}_{\mathsf{S}_N} \varphi_{N,n}$ and $\varphi_{N,n}^{\star}$ are isomorphisms of invariants whenever $n|N$ and $n \geq d_{\mathscr{V}}$.

We prove this conjecture for the special case of $\mathscr{V} = \mathscr{D}$ and $\mathrm{Sym}^2\mathscr{D}$ in Sections 4.6 and 4.6, respectively. We are now in a position to prove Theorem 4.5.2.

Proof (Theorem 4.5.2). The existence and uniqueness of a freely-symmetrized element $(\mathrm{sym}_n^{\mathscr{L}} q_k) \in \varinjlim_{\mathscr{L}} \mathbb{R}[\mathbb{V}_n]_{\leq d}^{\mathsf{S}_n}$ satisfying (4.37) follows by applying Proposition 4.5.3 to the FinSet-pair $(\mathrm{Sym}^d \mathscr{U}, \mathrm{Sym}^d \mathscr{L})$, and observing that if $q \in \mathbb{R}[\mathbb{V}_k]$ then $(L_{k,n}^{\star} q)(x) = q(L_{k,n} x)$.

To obtain (4.38), note that

$$
\begin{aligned}
u_\infty &= \inf_n \inf_{x \in \Omega_n} \mathbb{E} q_k(L_{k,n} x) \\
&= \inf_{(\mu_n) \in \varprojlim_{\mathscr{L}} \mathcal{P}(\Omega_n)^{\mathsf{S}_n}} \mathbb{E}_{\mu_k} q_k \quad \text{s.t. } (\mu_n = \mathrm{Law}(L_{n,m} x))_n \text{ for some } x \in \Omega_m.
\end{aligned}
$$

Since the objective is linear in (μ_n) and continuous in the weak topology on $\mathcal{P}(\Omega_k)^{\mathsf{S}_k}$, we can further optimize over limits of mixtures of such freely-described measures, which yields all of $\varprojlim_{\mathscr{L}} \mathcal{P}(\Omega_n)^{\mathsf{S}_n}$ by Theorems 4.4.5 and 4.4.9 applied to the reversed FinSet-pair $(\mathscr{L}, \mathscr{U})$.

Next, the optimal values ℓ_n are increasing by (4.16). They are lower bounds for u_∞ because whenever $k \preceq_{\mathscr{L}} n$ we have $\mathrm{sym}_n^{\mathscr{L}} q_k = \mathrm{sym}_{\mathsf{S}_n}(q_k \circ \psi_{n,k}^\star)$ so

$$\ell_n = \inf_{x\in\Omega_n} \mathrm{sym}_{\mathsf{S}_n}(q_k \circ \psi_{n,k}^\star) = \inf_{\mu\in\psi_{n,k}^\star \mathcal{P}(\Omega_n)^{\mathsf{S}_n}} \mathbb{E}_\mu q_k \leq \inf_{(\mu_n)\in \varprojlim_{\mathscr{L}} \mathcal{P}(\Omega_n)^{\mathsf{S}_n}} \mathbb{E}_{\mu_k} q_k = u_\infty, \tag{4.41}$$

where the inequality follows since if $(\mu_n) \in \varprojlim_{\mathscr{L}} \mathcal{P}(\Omega_n)^{\mathsf{S}_n}$ then $\mu_n \in \mathcal{P}(\Omega_n)^{\mathsf{S}_n}$ and $\mu_k = \psi_{n,k}^* \mu_n$.

Finally, we prove the claim convergence rates, which also imply that $\ell_\infty = u_\infty$. For each $x \in \Omega_n$ define $\nu_n = \mathrm{sym}_n \delta_x$ to be the distribution of Gx where $G \sim \mathrm{Haar}(\mathsf{S}_n)$ is a uniformly random permutation. Note that whenever $k \preceq_{\mathscr{L}} n$ we have

$$|p_n(x) - \mathrm{sym}_n^{\mathscr{L}} q_k(x)| = \left| \mathop{\mathbb{E}}_{X\sim L_{k,n}\nu_n} q_k(X) - \mathop{\mathbb{E}}_{X\sim\psi_{n,k}^\star \nu_n} q_k(X) \right| \leq \begin{cases} \|L_{k,n}\nu_n - \psi_{n,k}^\star \nu_n\|_{\mathrm{TV}} \|q_k\|_{\Omega_k}, \\ W_1(L_{k,n}\nu_n, \psi_{n,k}^\star \nu_n) \|q_k'\|_{\Omega_k}, \end{cases} \tag{4.42}$$

since q_k is $\|q_k'\|_{\Omega_k}$-Lipschitz with respect to $\|\cdot\|_1$ on Ω_k as $\|\cdot\|_\infty$ is the dual norm to $\|\cdot\|_1$. If $\mathscr{U}$ is a $\mathscr{D}$-sequence then $\mathscr{L}$ is a $\mathscr{Z}$-sequence an Theorem 4.4.5 gives $\|L_{k,n}\nu_n - \psi_{n,k}^\star \nu_n\|_{\mathrm{TV}} \leq \frac{k(k-1)}{n}$, while if $\mathscr{U}$ is a $\mathscr{Z}$-sequence then $\mathscr{L}$ is a $\mathscr{D}$-sequence and Theorem 4.4.9 gives $W_1(L_{k,kn}\nu_{nk}, \psi_{nk,k}^\star \nu_{nk}) \leq 2D\sqrt{\frac{k(k-1)}{n}} \|x\|_1$. Since

$$u_n - \ell_n = \inf_{x\in\Omega_n} p_n(x) - \inf_{x\in\Omega_n} \mathrm{sym}_n q_k(\psi_{n,k}^\star x) \leq \sup_{x\in\Omega_n} |p_n(x) - \mathrm{sym}_n q_k(\psi_{n,k}^\star x)|,$$

this proves both claimed rates of convergence. □

Any-Dimensional Certificates of Nonnegativity

We can rewrite u_∞ and our lower bounds ℓ_n for it in terms of any-dimensional certificates of nonnegativity. Suppose $\mathscr{U}$ is a $\mathscr{Z}$- or $\mathscr{D}$-consistent sequence, the sets $(\Omega_n \subseteq \mathbb{V}_n)$ are FinSet-compatible, and $u_n = \inf_{x\in\Omega_n} p_n(x)$ is a freely-described problem over $\mathscr{U}$ of degree at most d. Define the nonnegativity cones

$$\mathcal{K}_n = \{p \in \mathbb{R}[\mathbb{V}_n]_{\leq d}^{\mathsf{S}_n} : p|_{\Omega_n} \geq 0\}, \tag{4.43}$$

where we suppress the dependence of $\mathcal{K}_n$ on the sets (Ω_n) and degree d, held fixed throughout this section. We can rewrite each optimal value u_n in the sequence as the greatest upper bound on p_n over Ω_n, namely,

$$u_n = \sup_{\gamma\in\mathbb{R}} \gamma \quad \text{s.t. } p_n - \gamma \in \mathcal{K}_n. \tag{4.44}$$

The limiting optimal value can similarly be written as

$$\begin{aligned} u_\infty &= \sup_{\gamma\in\mathbb{R}} \gamma \quad \text{s.t. } p_n - \gamma \in \mathcal{K}_n \text{ for all } n \in \mathbb{N}, \\ &= \sup_{\gamma\in\mathbb{R}} \gamma \quad \text{s.t. } (p_n - \gamma)_n \in \bigcap_{n\in\mathbb{N}} \pi_n^{-1}(\mathcal{K}_n), \end{aligned} \tag{4.45}$$

where $\pi_k \colon \varprojlim_{\mathscr{U}} \mathbb{R}[\mathbb{V}_n]_{\leq d}^{\mathsf{S}_n} \to \mathbb{R}[\mathbb{V}_k]_{\leq d}^{\mathsf{S}_k}$ sends $(p_n) \mapsto p_k$. The latter intersection is precisely the collection of freely-described polynomials that are nonnegative over (Ω_n) in all dimensions, denoted

$$\varprojlim_{\mathscr{U}} \mathcal{K}_n = \left\{ (p_n) \in \varprojlim_{\mathscr{U}} \mathbb{R}[\mathbb{V}_n]_{\leq d}^{\mathsf{S}_n} : p_n|_{\Omega_n} \geq 0 \text{ for all } n \right\} = \bigcap_n \pi_n^{-1}(\mathcal{K}_n), \tag{4.46}$$

which is a convex cone in the finite-dimensional vector space $\varprojlim_{\mathscr{U}} \mathbb{R}[\mathbb{V}_n]_{\leq d}^{\mathsf{S}_n}$.

Producing lower bounds on u_∞ amounts to optimizing over subcones of $\varprojlim_{\mathscr{U}} \mathcal{K}_n$, and one family of such subcones is furnished by Theorem 4.5.2. Indeed, if $h_m \in \mathcal{K}_m$ is a fixed-dimensional polynomial nonnegative over Ω_m, then its image under the de Finetti map $(\mathbb{E}h_m \circ L_{m,n})_n \in \varprojlim_{\mathscr{U}} \mathcal{K}_n$ is a freely-described polynomial nonnegative in all dimensions, meaning that that $\mathbb{E}h_m \circ L_{m,n}|_{\Omega_n} \geq 0$ for all $n \in \mathbb{N}$. Therefore, for each fixed $m \in \mathbb{N}$ we get the lower bound

$$u_\infty \geq \sup_{\gamma\in\mathbb{R}} \gamma \quad \text{s.t. } (p_n - \gamma) = (\mathbb{E}h_m \circ L_{m,n}) \text{ for some } h_m \in \mathcal{K}_m. \tag{4.47}$$

We now argue that the lower bound in (4.47) is precisely ℓ_m from Theorem 4.5.2. Indeed, recall from Theorem 4.5.2 that $(p_n) = (\mathbb{E}q_k \circ L_{k,n})$ for a unique freely-symmetrized element $(\mathrm{sym}_n^{\mathscr{L}} q_k)$. Therefore, if $m \geq_{\mathscr{L}} k$ then

$$\begin{aligned} (p_n - \gamma) = (\mathbb{E}h_m \circ L_{m,n}) &\iff (\mathbb{E}(q_k - \gamma) \circ L_{k,n}) = (\mathbb{E}h_m \circ L_{m,n}) \\ &\iff (\mathrm{sym}_n^{\mathscr{L}}(q_k - \gamma)) = (\mathrm{sym}_n^{\mathscr{L}} h_m) \\ &\iff \mathrm{sym}_m^{\mathscr{L}} q_k - \gamma = h_m, \end{aligned}$$

where the first equivalence follows by definition of q_k, the second from the fact that the de Finetti map defines an isomorphism as in Proposition 4.5.3, and the third from Definition 4.2.3 and the assumption that $m \geq_{\mathscr{L}} k$. We also note that $\mathrm{sym}_m q_k - \gamma = h_m$ for some $h_m \in \mathcal{K}_m$ precisely when $\mathrm{sym}_m q_k - \gamma \in \mathcal{K}_m$. Thus, the lower bound in (4.47) becomes

$$\ell_m = \sup_{\gamma\in\mathbb{R}} \gamma \quad \text{s.t. } \mathrm{sym}_m q_k - \gamma \in \mathcal{K}_m. \tag{4.48}$$

Note that the only difference between the expression (4.48) for ℓ_m and the expression (4.44) for u_n is that p_n is replaced by $\mathrm{sym}_m q_k$, again highlighting the fact that we only modify the cost function but keep the same constraints in our construction of lower bounds.

Similarly, any convex relaxation of ℓ_m obtained by replacing $\mathcal{K}_m$ with a tractable subcone of it, including SOS [173] and relative entropy-based relaxations [161], can be interpreted in this way as searching over lower bounds γ that can be certified by a subcone of freely-described polynomials nonnegative in all dimensions.

Finally, taking the greatest lower bound that can be certified by letting the dimension m vary above, we obtain the lower bound

$$u_\infty \geq \ell_\infty = \sup_{\gamma\in\mathbb{R}} \gamma \quad \text{s.t. } (p_n - \gamma) = (\mathbb{E}h_m \circ L_{m,n}) \text{ for some } (\mathrm{sym}_n^{\mathscr{L}} h_m) \in \bigcup_n \varsigma_n(\mathcal{K}_n),$$

where $\varsigma_k\colon \mathbb{R}[\mathbb{V}_k]_{\leq d}^{\mathsf{S}_k} \to \varinjlim_{\mathscr{L}} \mathbb{R}[\mathbb{V}_n]_{\leq d}^{\mathsf{S}_n}$ sends $h_m \mapsto (\mathrm{sym}_n^{\mathscr{L}} q_m)_{n\geq_{\mathscr{L}} m}$. Just like our reformulation of u_∞ in (4.45), the bound ℓ_∞ involves a limiting cone in a finite-dimensional space of polynomials, this time the freely-symmetrized nonnegativity cone

$$\varinjlim_{\mathscr{L}} \mathcal{K}_n = \left\{ (\mathrm{sym}_n^{\mathscr{L}} h_m)_{m\leq_{\mathscr{L}} n} \in \varinjlim_{\mathscr{L}} \mathbb{R}[\mathbb{V}_n]_{\leq d}^{\mathsf{S}_n} : h_m|_{\Omega_m} \geq 0 \right\} = \bigcup_n \varsigma_n(\mathcal{K}_n), \tag{4.49}$$

contained in $\varinjlim_{\mathscr{L}} \mathbb{R}[\mathbb{V}_n]_{\leq d}^{\mathsf{S}_n}$.

The de Finetti map $(\mathrm{sym}_n^{\mathscr{L}} h_m) \mapsto (\mathbb{E}h_m \circ L_{m,n})$ sends the cone $\varinjlim_{\mathscr{L}} \mathcal{K}_n$ into the cone $\varprojlim_{\mathscr{U}} \mathcal{K}_n$. The fact that $u_\infty = \ell_\infty$ can then be derived from our present geometric perspective by proving that the image of this map is dense. We proceed to prove this, along with a few basic geometric properties of the two limit cones.

Theorem 4.5.5. *Suppose $(\mathscr{U}, \mathscr{L})$ is an* FinSet*-pair, let (Ω_n) be* FinSet*-compatible, fix degree $d \in \mathbb{N}$, and denote by $\mathcal{K}_n$ the nonnegativity cones (4.43).*

(Density) *The isomorphism $(\mathrm{sym}_n^{\mathscr{L}} q_k) \mapsto (\mathbb{E}q_k \circ L_{k,n})$ from Proposition 4.5.3 maps $\varinjlim_{\mathscr{L}} \mathcal{K}_n$ into a dense subset of $\varprojlim_{\mathscr{U}} \mathcal{K}_n$.*

(Monotonicity) *We have*

$$\pi_n^{-1}(\mathcal{K}_n) \supset \pi_N^{-1}(\mathcal{K}_N) \ \text{ if } n \preceq_{\mathscr{U}} N, \quad \text{and} \quad s_n(\mathcal{K}_n) \subseteq s_N(\mathcal{K}_N) \ \text{ if } n \preceq_{\mathscr{L}} N. \tag{4.50}$$

(Geometry) *The cones* $\varprojlim_{\mathscr{U}} \mathcal{K}_n$ *and* $\varinjlim_{\mathscr{L}} \mathcal{K}_n$ *are convex, pointed, and have nonempty interiors. The former is also closed.*

For the density result, closures are taken with respect to the usual unique Hausdorff TVS topology on a finite-dimensional space (e.g., induced by a norm on the coefficients in a basis).

Proof. Note that if $q_k|_{\Omega_k} \geq 0$ then $\mathbb{E}q_k(L_{k,n}x) \geq 0$ for all $x \in \Omega_n$ since $L_{k,n}x \in \Omega_k$ almost surely. This shows that $\varinjlim_{\mathscr{L}} \mathcal{K}_n$ is indeed mapped into $\varprojlim_{\mathscr{U}} \mathcal{K}_n$. Moreover, we argue next that this image is dense. If $(p_n) \in \varprojlim_{\mathscr{U}} \mathcal{K}_n$ then for each $i \in \mathbb{N}$ we have $(\mathrm{sym}_n^{\mathscr{L}} p_i)_{n \geq_{\mathscr{L}} i} \in \varinjlim_{\mathscr{L}} \mathcal{K}_n$, which is mapped to $(\mathbb{E}p_i \circ L_{i,n}) \in \varprojlim_{\mathscr{U}} \mathcal{K}_n$. By the proof of Proposition 4.5.3, we have $\lim_{i\to\infty} \mathbb{E}p_i \circ L_{i,n} = p_n$ for each $n \in \mathbb{N}$, which implies $(\mathbb{E}p_i \circ L_{i,n}) \to (p_n)$ in the usual topology on the finite-dimensional space $\varprojlim_{\mathscr{U}} \mathbb{R}[\mathbb{V}_n]_{\leq d}^{\mathsf{S}_n}$. This proves the claimed density.

For monotonocity, note that if $(p_n) \in \varprojlim_{\mathscr{U}} \mathbb{R}[\mathbb{V}_n]_{\leq d}^{\mathsf{S}_n}$ satisfies $p_N|_{\Omega_N} \geq 0$ and $n \preceq_{\mathscr{U}} N$, then $p_n|_{\Omega_n} = p_N \circ \varphi_{N,n}|_{\Omega_n} \geq 0$ since $\varphi_{N,n}\Omega_n \subseteq \Omega_N$. Similarly, if $q_k|_{\Omega_k} \geq 0$ and $k \preceq_{\mathscr{L}} n$ then $\mathrm{sym}_n^{\mathscr{L}} q_k|_{\Omega_n}(x) = \mathbb{E}_{G\sim\mathrm{Haar}(\mathsf{S}_n)}[q_k(\psi_{n,k}^{\star} Gx)] \geq 0$ since $\psi_{n,i}^{*} Gx \in \Omega_i$ almost surely by compatibility of (Ω_n).

Since π_n and s_n are isomorphisms for all large n by Proposition 4.5.4 and as $\mathcal{K}_n$ is closed, convex, and has nonempty interior, we conclude that $\pi_n^{-1}(\mathcal{K}_n)$ and $s_n(\mathcal{K}_n)$ are both closed, convex, and have nonempty interiors. Therefore, the expressions (4.46) and (4.49) imply that $\varprojlim_{\mathscr{U}} \mathcal{K}_n$ and $\varinjlim_{\mathscr{L}} \mathcal{K}_n$ are both convex and pointed, that the former is closed, and that the latter has nonempty interior. Since $\varinjlim_{\mathscr{L}} \mathcal{K}_n$ is mapped by an isomorphism to a subset of $\varprojlim_{\mathscr{U}} \mathcal{K}_n^{\mathsf{S}_n}$, we conclude that the latter also has nonempty interior. $\square$

The last two results in Theorem 4.5.5 were proved for $\mathscr{U} = \mathscr{I}$ in [3, Thm. 3.2] and for $\mathscr{U} = \mathscr{D}$ in [2, §1.1]. Our proof for the former case is significantly simpler thanks to the first result in Theorem 4.5.5, and our proof applies more generally to any FinSet-pair.

Proofs for Section 4.5

In this section, we prove Propositions 4.5.3 and 4.5.4 from Section 4.5. To prove the former proposition, we require a technical lemma guaranteeing that certain norms of vectors in a freely-described sequence over a $\mathscr{I}$-sequence are constant.

Lemma 4.5.6. *Let $\mathscr{V} = \mathcal{F}(\mathscr{Z})$ be a $\mathscr{Z}$-sequence of degree d and let $(e_\alpha^{(n)})_{\alpha\in\mathcal{A}_n}$ be the bases for it obtained by extending the standard bases for $(\mathbb{R}^n)$ as in Lemma 4.4.21.*

(Permutation sequence) *There are inclusions $\mathcal{A}_n \subseteq \mathcal{A}_N$ such that $\varphi_{N,n}(e_\alpha^{(n)}) = e_\alpha^{(N)}$ and $\varphi_{N,n}^\star e_\alpha^{(N)} = e_\alpha^{(n)}$ if $\alpha \in \mathcal{A}_n$ and 0 otherwise, for all $n \leq N$. Also, there is an action of S_n on the index sets $\mathcal{A}_n$ such that $g \cdot e_\alpha^{(n)} = e_{g\cdot\alpha}^{(n)}$ for all $\alpha \in \mathcal{A}_n$ and all n.*

(Isometries on invariants) *Define $\|v\|_\infty = \max_{\alpha\in\mathcal{A}_n} |v_\alpha|$ where $v = \sum_{\alpha\in\mathcal{A}_n} v_\alpha e_\alpha^{(n)}$. Then this is the dual norm to the one from Lemma 4.4.21, and $\|\varphi_{N,n}^\star v_N\|_\infty = \|v_N\|_\infty$ whenever $v_N \in \mathbb{V}_N^{\mathsf{S}_N}$, and $d \leq n \leq N$.*

Consistent sequences admitting such bases are called *permutation modules* in [139, §2.4].

Proof. The first claim is trivial for $(\mathbb{R}^n)$ and follows straightforwardly for general $\mathscr{Z}$-sequences by inducting on the constructions $\mathcal{F}$. Next, observe that invariants in $\mathbb{V}_n$ are in bijection with S_n-orbits in $\mathcal{A}_n$, so that the number such orbits is constant for $n \geq d$ by Proposition 4.2.6. It was further shown in [139, Prop. 2.15(b)] that these orbits are generated by $\mathcal{A}_d$ for all $n \geq d$. Thus, if $O_1, \dots O_m$ are the S_d-orbits in $\mathcal{A}_d$, then $\mathsf{S}_n O_1, \dots, \mathsf{S}_n O_m$ are the S_n-orbits in $\mathcal{A}_n$ for all $n \geq d$.

If $v_N \in \mathbb{V}_N^{\mathsf{S}_N}$ then it can be written in the above basis via $v_N = \sum_{i=1}^m c_i \sum_{\alpha\in\mathsf{S}_N O_i} e_\alpha^{(N)}$ and $\varphi_{N,n}^\star v_N = \sum_{i=1}^m c_i \sum_{\alpha\in\mathsf{S}_n O_i} e_\alpha^{(n)}$. This shows that $\|v_N\|_\infty = \|\varphi_{N,n}^\star v_N\|_\infty = \max_{i\in[m]} |c_i|$. The fact that this ℓ_∞ norm is the dual norm to the norm from Lemma 4.4.21 follows from the fact that the latter is the ℓ_1 norm in this same basis. □

We now prove that the de Finetti maps induce isomorphisms between freely-symmetrized and freely-described elements.

Proof (Proposition 4.5.3). First, observe that the given map is well-defined and yields freely-described elements. Indeed, stated in terms of de Finetti maps, Lemma 4.4.13 states that $L_{n,m}^\star \psi_{n,k} \stackrel{d}{=} L_{k,m}^\star$ whenever $k \preceq_{\mathscr{L}} n$ and $\varphi_{m,k}^\star L_{n,m}^\star \stackrel{d}{=} L_{n,k}^\star$ whenever $k \preceq_{\mathscr{U}} m$. The former property shows that two polynomials which correspond to the same freely-symmetrized elements have the same images, and the latter shows that the images of our map are indeed freely-described over $\mathscr{U}$.

Second, it suffices to prove the claim when $\mathscr{U} = \mathcal{F}(\mathscr{Z})$ is a $\mathscr{Z}$-sequence and $\mathscr{L} = \mathcal{F}(\mathscr{D})$ is the corresponding $\mathscr{D}$-sequence, as the opposite case follows from this one via duality by Proposition 4.2.5. Third, we show that the given map is surjective, which suffices for the proof as $\dim \varinjlim_{\mathscr{L}} \mathbb{V}_n^{\mathsf{S}_n} \leq \dim \varprojlim_{\mathscr{U}} \mathbb{V}_n^{\mathsf{S}_n}$ by Proposition 4.2.6. To this end, fix a freely-described element $(v_n) \in \varprojlim_{\mathscr{U}} \mathbb{V}_n^{\mathsf{S}_n}$ and let $\Omega_n = \overline{\bigcup_{m\in\mathbb{N}} \operatorname{supp}(L_{m,n}^{\star} v_m)}$, which is FinSet-compatible by construction. Note also that these sets are compact, as if we endow $(\mathbb{V}_n)$ with the ℓ_∞ norm from Lemma 4.5.6 and let d be the degree of $\mathcal{F}$, we have $\|L_{m,n}^{\star} v_m\|_\infty \leq \|v_m\|_\infty = \|v_d\|_\infty$ by Lemmas 4.4.21 and 4.5.6.

Then (v_n) can be identified with the freely-described measure $(\delta_{v_n}) \in \varprojlim_{\mathscr{U}} \mathcal{P}(\Omega_n)^{\mathsf{S}_n}$ over $\mathscr{U}$. By Theorem 4.4.5, the random variables $L_{k,n}^{\star} v_k$ converge in total variation to δ_{v_n} for any fixed n as $k \to \infty$. Taking expectations and using the compactness of (Ω_n), we conclude that $\|\mathbb{E} L_{k,n}^{\star} v_k - v_n\|_\infty \to 0$ for each n as $k \to \infty$. Since $(\mathbb{E} L_{k,n}^{\star} v_k)$ is the image under the given map of $(\operatorname{sym}_n^{\mathscr{L}} v_k)_{k|n}$, this shows that the given map has a dense image. Since its image is a finite-dimensional space it must be further closed. Thus, the given map is surjective. □

Combining the isomorphism from Proposition 4.5.3 with Proposition 4.2.6 yields Proposition 4.5.4.

Proof (Proposition 4.5.4). This result was proved for $\mathscr{Z}$-sequences with $k = d_{\mathscr{V}}$ in [139, Prop. 2.9, Thm. 2.11]. Suppose $\mathscr{V} = \mathcal{F}(\mathscr{D})$ is a $\mathscr{D}$-sequence and let $\mathscr{V}' = \mathcal{F}(\mathscr{Z})$ be the corresponding $\mathscr{Z}$-sequence, so that $(\mathscr{V}, \mathscr{V}')$ is an FinSet-pairs. Since the underlying vectors spaces are the same for $\mathscr{V}$ and for $\mathscr{V}'$, we have $\dim \varinjlim_{\mathscr{V}'} \mathbb{V}_n^{\mathsf{S}_n} = \dim \mathbb{V}_n^{\mathsf{S}_n} = \dim \mathbb{V}_{d_{\mathscr{V}}}^{\mathsf{S}_{d_{\mathscr{V}}}}$ for all $n \geq d_{\mathscr{V}}$ by the above result of [139]. Applying Proposition 4.5.3, we conclude that $\dim \varprojlim_{\mathscr{V}} \mathbb{V}_n^{\mathsf{S}_n} = \dim \mathbb{V}_{d_{\mathscr{V}}}^{\mathsf{S}_{d_{\mathscr{V}}}}$. By Proposition 4.2.6 we also have $\dim \varinjlim_{\mathscr{V}} \mathbb{V}_n^{\mathsf{S}_n} = \dim \mathbb{V}_{d_{\mathscr{V}}}^{\mathsf{S}_{d_{\mathscr{V}}}}$.

Denote $D = \dim \mathbb{V}_{d_{\mathscr{V}}}^{\mathsf{S}_{d_{\mathscr{V}}}}$. Suppose $\{(\operatorname{sym}_n^{\mathscr{V}} v_{k_i})_{n\geq k_i}\}_{i=1}^{D}$ is a basis for $\varinjlim_{\mathscr{V}} \mathbb{V}_n^{\mathsf{S}_n}$. If $k \in \mathbb{N}$ satisfies $k_i | k$ for all $i \in [D]$ and $k \geq d_{\mathscr{V}}$, note that $\{\operatorname{sym}_n^{\mathscr{V}} v_{k_i}\}_{i\in[D]}$ forms a basis for $\mathbb{V}_n^{\mathsf{S}_n}$ whenever $k|n$. Indeed, this is a collection of $D = \dim \mathbb{V}_n^{\mathsf{S}_n}$ linearly independent vectors, since any linear relation between them in $\mathbb{V}_n^{\mathsf{S}_n}$ also holds in $\varinjlim_{\mathscr{V}} \mathbb{V}_n^{\mathsf{S}_n}$. This shows that $\operatorname{sym}_N \varphi_{N,n} \colon \mathbb{V}_n^{\mathsf{S}_n} \to \mathbb{V}_N^{\mathsf{S}_N}$ is an isomorphism whenever $k|n$ and $n|N$. Its adjoint is simply the projection $\varphi_{N,n}^{\star}$, hence the latter is also an isomorphism. □

4.6 Illustrations of Our Framework

In this section we present a systematic framework for obtaining bounds on freely-described POPs by leveraging Theorem 4.5.2 in Section 4.5. We describe the steps of our method here and we illustrate our approach in four consistent sequences arising in different applications.

Concretely, suppose that $(\mathscr{U}, \mathscr{L})$ is an FinSet-pair of degree D and consider freely-described problems $u_n = \inf_{x \in \Omega_n} p_n(x)$ over $\mathscr{U}$, where $\deg p_n = d$ and the constraints (Ω_n) are FS-compatible (i.e., embed into each other in $\mathscr{U}$ and project into each other in $\mathscr{L}$). We construct freely-symmetrized lower bounds for such problems by carrying out the following steps.

(Basis for $\mathscr{L}$) Fix a $k \geq Dd$. Let $q_k^{(1)}, \ldots, q_k^{(m)}$ be a basis for $\mathbb{R}[\mathbb{V}_k]_{\leq d}^{\mathsf{S}_k}$.

(Basis for $\mathscr{U}$) Derive $p_n^{(i)} = \mathbb{E}[q_k^{(i)} \circ L_{k,n}]$ to which the above basis elements map under the de Finetti map. We prove that the collection $(p_n^{(1)}), \ldots, (p_n^{(m)})$ forms a basis for freely-described polynomials over $\mathscr{U}$ of degree at most d, by proving that the $p_n^{(i)}$ form a basis for invariants for each $n \geq Dd$.

(Dual cost) Express the freely-described cost in this basis as $p_n = \sum_{i=1}^m c^{(i)} p_n^{(i)}$. The dual costs for $n \succeq_{\mathscr{L}} k$ are given by $q_n = \sum_{i=1}^m c^{(i)} \mathrm{sym}_n^{\mathscr{L}} q_k^{(i)}$.

When $\mathscr{L}$ is a $\mathscr{Z}$-sequence, the collection $(\mathrm{sym}_n^{\mathscr{L}} q_k^{(1)})_{n \succeq_{\mathscr{L}} k}, \ldots, (\mathrm{sym}_n^{\mathscr{L}} q_k^{(m)})_{n \succeq_{\mathscr{L}} k}$ is a basis for freely-symmetrized polynomials over $\mathscr{L}$ of degree at most d by [139, Prop. 2.9, Thm. 2.11]. This proves that $(p_n^{(1)}), \ldots, (p_n^{(m)})$ form a basis for freely-described polynomials over $\mathscr{U}$ of degree at most d by Proposition 4.5.3. On the other hand, when $\mathscr{L}$ is a $\mathscr{D}$-sequence, Proposition 4.5.4 only guarantees that the $p_n^{(i)}$ form a basis when k is chosen sufficiently large in step 1. Nevertheless, we prove using case-specific arguments that the $p_n^{(i)}$ obtained this way indeed form a basis, for *any* $k \geq Dd$. This proves our conjectured strengthening of Proposition 4.5.4 for the $\mathscr{D}$-sequences $\mathscr{D}$ and $\mathrm{Sym}^2\mathscr{D}$ studied below.

The above steps yield the freely-symmetrized bounds $\ell_n = \inf_{x \in \Omega_n} q_n(x)$ for $n \succeq_{\mathscr{L}} k$. When $\mathscr{L}$ is a $\mathscr{Z}$-sequence, the partial order $\preceq_{\mathscr{L}}$ is just the standard order on $\mathbb{N}$. Consequently, it suffices to just perform the above derivation for $k = Dd$ to form the lower bounds for all dimensions $n \geq k$. On the other hand, when $\mathscr{L}$ is a $\mathscr{D}$-sequence, the partial order $\preceq_{\mathscr{L}}$ is the divisibility order, and performing the above three steps for $k = Dd$ only yields the bounds ℓ_n for $k|n$. If we seek a lower bound ℓ_n when

$k \nmid n$, we need to perform the preceding computations by setting $k = n$ in the first step (or for some other $k' \geq Dd$ such that $k'|n$).

On a related but distinct point, explicitly forming the polynomials $\mathrm{sym}_n^{\mathscr{L}} q_k^{(i)}$ can also be inconvenient when $\mathscr{L}$ is a $\mathscr{D}$-sequence, in which case embeddings do not map monomials to monomials. It is then more convenient to directly perform the preceding derivations with $k = n$ in the first step. As we illustrate in this section, in several applications it is possible to choose convenient bases $(q_k^{(i)})$ in the first step which are explicitly given as simple functions of k, and to perform the above three steps for them to obtain $p_n^{(i)}$ as a function of k for all k; see Sections 4.6 and 4.6.

All the code producing the numerical results in this section is available at: `https://github.com/eitangl/anyDimPolyOpt`.

Polynomial Mean-Field Games

Model problem. We consider freely-described problems over $\mathscr{D}^{d_g}$ arising as the values of potential mean-field games. Mean-field games are limits of games with a large number of identical players; see [36] for an introduction. Formally, in a potential mean-field game with a compact strategy set $\Theta \subseteq \mathbb{R}^{d_g}$, we consider symmetric games in a growing number n of players with symmetric costs $C_1^{(n)}, \dots, C_n^{(n)}$ satisfying $C_{\sigma(i)}^{(n)}(x_{\sigma(1)}, \dots, x_{\sigma(n)}) = C_i(x_1, \dots, x_n)$ for all $\sigma \in \mathsf{S}_n$ and all $x_1, \dots, x_n \in \Theta$. The game is collaborative in that the players jointly incur a cost $\frac{1}{n}\sum_{i=1}^n C_i^{(n)}(x)$ that they seek to minimize, and we are interested in the minimum value that they can achieve in the limit as $n \to \infty$. Under suitable assumptions [36, Thm. 2.1], there exists a single function $C \colon \Theta \times \mathcal{P}(\Theta) \to \mathbb{R}$ such that $\sup_{x \in \Theta^n} |C_i^{(n)}(x) - C(x_i, \frac{1}{n}\sum_j \delta_{x_j})| \to 0$ for all $i \in [n]$ as $n \to \infty$. The limiting value of the game becomes an optimization problem over measures $\inf_{\mu \in \mathcal{P}(\Theta)} \mathbb{E}_\mu[C(X, \mu)]$ where $X \sim \mu$. We restrict ourselves to polynomial costs of the form $f(x, \mathbb{E}_\mu[X^{\alpha_1}], \dots, \mathbb{E}_\mu[X^{\alpha_m}])$ for a polynomial $f : \mathbb{R}^{d_g} \times \mathbb{R}^m \to \mathbb{R}$ and some fixed multi-indices $\alpha_1, \dots, \alpha_m \in \mathbb{N}^d \setminus \{0\}$. We then have the following sequence of freely-described POPs, similar to the one considered in Section 4.1

$$u_n = \inf_{(x_1, \dots, x_n) \in \Theta^n} \frac{1}{n}\sum_{i=1}^n f\left(x_i, \frac{1}{n}\sum_{j=1}^n x_j^{\alpha_1}, \dots, \frac{1}{n}\sum_{j=1}^n x_j^{\alpha_m}\right), \tag{4.51}$$

with the associated limiting optimal value of

$$u_\infty = \inf_{\mu \in \mathcal{P}(\Theta)} \mathbb{E}_\mu\Big[f(X, \mathbb{E}_\mu[X^{\alpha_1}], \dots, \mathbb{E}_\mu[X^{\alpha_m}])\Big].$$

Our objective is to obtain lower bounds on u_∞ by identifying the corresponding freely-symmetrized POP over $\mathscr{Z}^{d_g}$. Note that the vector spaces underlying $\mathscr{D}^{d_g}$ and $\mathscr{Z}^{d_g}$ are $\mathbb{V}_n = \mathbb{R}^{d_g \times n}$.

Basis for $\mathscr{L} = \mathscr{Z}^{d_g}$. Consider a list of multi-indices $\Lambda = (\alpha_1 \geq \ldots \geq \alpha_{\mathrm{len}(\Lambda)})$ sorted lexicographically with $\alpha_i \in \mathbb{N}^{d_g} \setminus \{0\}$ for all i. For $k \geq \mathrm{len}(\Lambda)$, we define the monomial means

$$\bar{m}_k^{(\Lambda)}(x_1, \ldots, x_k) = \mathrm{sym}_k^{\mathscr{Z}^{d_g}} x_1^{\alpha_1} \cdots x_{\mathrm{len}(\Lambda)}^{\alpha_{\mathrm{len}(\Lambda)}} = \frac{1}{\binom{k}{\mathrm{len}(\Lambda)}} \sum_{1 \leq i_1 < \cdots < i_{\mathrm{len}(\Lambda)} \leq k} x_{i_1}^{\alpha_1} \cdots x_{i_{\mathrm{len}(\Lambda)}}^{\alpha_{\mathrm{len}(\Lambda)}}, \tag{4.52}$$

whose degree is $|\Lambda| = \sum_{i=1}^{\mathrm{len}(\Lambda)} |\alpha_i|$. The collection $\{\bar{m}_k^{(\Lambda)}\}_{|\Lambda| \leq d}$ forms a basis for $\mathbb{R}[\mathbb{V}_k]_{\leq d}^{\mathsf{S}_k}$ for any $k \geq d$, as can be seen by writing any polynomial in the monomial basis and symmetrizing. Moreover, we have $\bar{m}_K^{(\Lambda)} = \mathrm{sym}_K^{\mathscr{L}} \bar{m}_k^{(\Lambda)}$ whenever $k \leq K$ by construction. Thus, $(\bar{m}_n^{(\Lambda)})_{n \geq d}$ form a basis for freely-symmetrized polynomials of degree at most d over $\mathscr{L}$, as proved more generally in [139].

Basis for $\mathscr{U} = \mathscr{D}^{d_g}$. To compute $\mathbb{E}[\bar{m}_k^{(\Lambda)} \circ L_{k,n}]$ for the de Finetti map $L_{k,n}$, observe that if $x_1, \ldots, x_n \in \mathbb{R}^{d_g}$ then $\mathrm{Law}(L_{k,n} \cdot (x_1, \ldots, x_n)) = \mu_x^{\otimes k}$ where $\mu_x = \frac{1}{n} \sum_{i=1}^n \delta_{x_i} \in \mathcal{P}(\mathbb{R}^{d_g})$. Therefore, from (4.52)

$$\mathbb{E}[\bar{m}_k^{(\Lambda)}(L_{k,n} \cdot (x_1, \ldots, x_n))] = \mathbb{E}_{\mu_x^{\otimes k}}[x_1^{\alpha_1} \cdots x_{\mathrm{len}(\Lambda)}^{\alpha_{\mathrm{len}(\Lambda)}}] = \prod_{i=1}^{\mathrm{len}(\Lambda)} \mathbb{E}_{\mu_x}[x^{\alpha_i}] = \prod_{i=1}^{\mathrm{len}(\Lambda)} \left(\frac{1}{n} \sum_{j=1}^n x_j^{\alpha_i} \right).$$

By defining the power means as

$$\bar{s}_n^{(\Lambda)}(x_1, \ldots, x_n) = \prod_{i=1}^{\mathrm{len}(\Lambda)} \bar{s}_n^{(\alpha_i)}(x_1, \ldots, x_n), \quad \text{where} \quad \bar{s}_n^{(\alpha_i)}(x_1, \ldots, x_n) = \frac{1}{n} \sum_{j=1}^n x_j^{\alpha_i},$$

we have that $\mathbb{E}[\bar{m}_k^{(\Lambda)} \circ L_{k,n}] = \bar{s}_n^{(\Lambda)}$. We conclude by Proposition 4.5.3 that $\{(\bar{s}_n^{(\Lambda)})\}_{|\Lambda| \leq d}$ is a basis for freely-described polynomials of degree d over $\mathscr{D}^{d_g}$.

Dual cost. Write the freely-described cost (p_n) as $p_n = \sum_\Lambda c^{(\Lambda)} \bar{s}_n^{(\Lambda)}$. Then the dual cost is given by $q_n = \sum_\Lambda c^{(\Lambda)} \bar{m}_n^{(\Lambda)}$ for all $n \geq d$. The lower bounds on the freely-described problems (4.51) are then given by $\ell_n = \inf_{X \in \Theta^n} q_n(X)$, which satisfy $\ell_n \leq \ell_{n+1}$ and $u_n - \ell_n \lesssim 1/n$ by Theorem 4.5.2.

FinSet**-compatible sets.** The sequences of constraint sets $(\Omega_n \subseteq \mathbb{R}^{d_g \times n})$ for which our framework is applicable consist of S_n-invariant compact sets satisfying $(\underbrace{x_1, \ldots, x_1}_{k \text{ times}}, \ldots, \underbrace{x_n, \ldots, x_n}_{k \text{ times}}) \in \Omega_{nk}$ and $(x_1, \ldots, x_{n-1}) \in \Omega_{n-1}$ if $(x_1, \ldots, x_n) \in \Omega_n$. An example is a product set $\Omega_n = \Theta^n$ for any compact $\Theta \subseteq \mathbb{R}^{d_g}$. For non-examples, neither the sequence of Frobenius norm balls in $\mathbb{R}^{d_g \times n}$ nor the sequence of row stochastic matrices in $\mathbb{R}^{d_g \times n}$ stochastic matrices are FinSet-compatible.

Example 4.6.1. *For $d_g = 1$ and $\Theta = [-1, 1]$, consider the polynomial cost $f(x, \mu_1, \mu_2) = 5x\mu_1 - 2x^2\mu_1 - 2x\mu_2 - \mu_1$, where $\mu_k = \mathbb{E}_\mu[t^k]$. This is a mean-field extension of the non-convex-concave two-player game studied in [174, Ex. 3.2], where we replace one of the players by a growing number of symmetric players and take their payoff function to be our cost. Then the value of the game (4.51) becomes*

$$u_n = \inf_{x \in [-1,1]^n} 5\bar{s}_n^{(1,1)}(x) - 4\bar{s}_n^{(2,1)}(x) - \bar{s}_n^{(1)}(x),$$

which is once again neither convex nor concave. We then get the lower bounds for $n \geq 3$

$$\ell_n = \inf_{x \in [-1,1]^n} 5\bar{m}_n^{(1,1)}(x) - 4\bar{m}_n^{(2,1)}(x) - \bar{m}_n^{(1)}(x).$$

These are precisely the optimization problems from Example 1.3.4 in Section 3.1. We numerically compute and plot u_n and ℓ_n for $n \in [3, 21]$ in Figure 1.1. Specifically, in the left panel of Figure 1.1 we obtain upper bounds on u_n and intervals containing ℓ_n; the upper bounds on u_n, ℓ_n are computed using a nonconvex solver (Matlab's `fmincon`), while the lower bounds on ℓ_n are computed using a degree-4 sums-of-squares relaxation. In the right panel of Figure 1.1 we plot the error bars containing the differences $u_n - \ell_n$ against the theoretical bound on these differences from Theorem 4.5.2, using $k = 3$ and $\|q_3\|_{[-1,1]^3} \leq 10$ to obtain (1.4). Note that the upper bounds u_n are not monotonically decreasing. This does not contradict (4.13), which only guarantees $u_{nk} \leq u_n$.

Inequalities in Symmetric Functions

Model problem. Dually to the previous section, we consider freely-described problems over $\mathscr{U} = \mathscr{X}$. Freely-described polynomials over $\mathscr{X}$ are precisely the extensively-studied symmetric functions [209, Chap. 7], and certifying inequalities between them which hold in all dimensions is therefore of interest. Any homogeneous symmetric polynomial is a linear combination of products of power sums $s_n^{(m)}(x) = \sum_{i=1}^n x_i^m$ by [209, §7.7]. Moreover, certifying nonnegativity of a homogeneous polynomial is equivalent to certifying that its minimum over a norm ball is nonnegative. Thus, we consider freely-described problems of the form

$$u_n = \inf_{\substack{x \in \mathbb{R}^n \\ \|x\|_1 \leq 1}} \sum_\lambda c^{(\lambda)} s_n^{(\lambda)}(x), \tag{4.53}$$

where for each partition $\lambda = (\lambda_1 \geq \cdots \geq \lambda_k > 0)$ we set $s_n^{(\lambda)}(x) = \prod_{i=1}^k s_n^{(\lambda_i)}(x)$ as the product of power sums. It was shown in [4] that the problem of deciding if a given inequality in power sums holds in all dimensions is undecidable. Here we systematically construct lower bounds for symmetric functions which hold in all dimensions in terms of freely-symmetrized problems over $\mathscr{D}$.

Basis for $\mathscr{L} = \mathscr{D}$. Consider any partition $\lambda = (\lambda_1 \geq \ldots \geq \lambda_{\mathrm{len}(\lambda)} > 0)$ where $\mathrm{len}(\lambda)$ denotes the length of λ and $|\lambda| = \sum_i \lambda_i$. We view $\lambda \in \mathbb{N}^k$ for any $k \geq \mathrm{len}(\lambda)$ by zero-padding λ, and let S_k act on λ by coordinate permutations. The monomial sums are defined by sums over orbits of such partitions for $x \in \mathbb{R}^k$:

$$m_k^{(\lambda)}(x) = \sum_{\lambda' \in \mathsf{S}_k \lambda} x^{\lambda'}. \tag{4.54}$$

For example, for each $m \in \mathbb{N}$ we have $m_k^{(m)} = s_k^{(m)}$ and the m'th elementary symmetric polynonial $m_k^{(1^m)}(x) = \sum_{i_1 < \ldots < i_m} x_{i_1} \cdots x_{i_m}$, where $(1^m) = (1, \ldots, 1)$ is the partition consisting of m ones. Note that $\deg m_k^{(\lambda)} = |\lambda|$. For each $k \geq \mathrm{len}(\lambda)$ the monomial sums $m_k^{(\lambda)}$ are scalar multiples of the monomial means $\bar{m}_k^{(\lambda)}$ in (4.52). In particular, for each $k \geq d$ the monomial sums $\{m_k^{(\lambda)}\}_{|\lambda| \leq d}$ form a basis for $\mathbb{R}[x_1, \ldots, x_k]_{\leq d}^{\mathsf{S}_k}$. Proposition 4.5.4 does not, however, guarantee that the corresponding freely-symmetrized elements $(\mathrm{sym}_n^{\mathscr{L}} m_k^{(\lambda)})$ form a basis. Therefore, the images $(\mathbb{E}[m_k^{(\lambda)} \circ L_{k,n}])_n$ under the de Finetti maps are not *a priori* guaranteed to form a basis for freely-described polynomials. Both collections turn out to be bases, however, as we show below.

Basis for $\mathscr{U} = \mathscr{Z}$. We proceed to derive $\mathbb{E}[m_k^{(\lambda)} \circ L_{k,n}]$ for each k, and to prove that the collection of these as λ varies over partitions with $|\lambda| \leq d$ forms a basis for freely-described polynomials over $\mathscr{Z}$. To state the resulting expressions, we recall the refinement partial order on partitions. Specifically, for two partitions μ, λ, we set $\lambda \leq_R \mu$ if λ is a *refinement* of μ, meaning that there exists a surjection $f\colon [\mathrm{len}(\lambda)] \to [\mathrm{len}(\mu)]$, denoted $f(\lambda) = \mu$, such that $\sum_{j \in f^{-1}(i)} \lambda_j = \mu_i$ for each $i \in [\mathrm{len}(\mu)]$. For example, $(1,1,1) \leq_R (2,1) \leq_R (3)$ are all the refinments of (3). Define the integer $R_{\lambda,\mu}$ to be the number of such surjections f. The refinement partial order and integers $R_{\lambda,\mu}$ are classical objects of study, e.g., they relate power sums and monomial sums via [209, Prop. 7.7.1]

$$s_n^{(\lambda)}(x) = \sum_{\lambda \leq_R \mu} R_{\lambda,\mu} m_n^{(\mu)}(x), \tag{4.55}$$

for any $n \geq \mathrm{len}(\lambda)$. Remarkably, these integers $R_{\lambda,\mu}$ also appear when applying our de Finetti map to monomial sums.

Proposition 4.6.2. *For any $k \geq \operatorname{len}(\lambda)$, we have that $p_n^{(k,\lambda)}(x) = \mathbb{E}[m_k^{(\lambda)}(L_{k,n}x)]$ for $x \in \mathbb{R}^n$ is given by*

$$p_n^{(k,\lambda)}(x) = \sum_{\mu \leq_R \lambda} \frac{\lambda!}{\mu!} \cdot \frac{(k)_{\operatorname{len}(\lambda)}}{k^{\operatorname{len}(\mu)}} \cdot \frac{R_{\mu,\lambda}}{R_{\lambda,\lambda}} \cdot m_n^{(\mu)}(x), \tag{4.56}$$

where $(k)_\ell = \frac{k!}{(k-\ell)!}$ and $\lambda! = \prod_{i=1}^{\operatorname{len}(\lambda)} \lambda_i!$. Moreover, whenever $k, n \geq d$ the collection $\{p_n^{(k,\lambda)}\}_{|\lambda|=d}$ forms a basis for $\mathbb{R}[\mathbb{V}_n]_d^{\mathsf{S}_n}$.

Proof. We have $L_{k,n}x = \sum_{i=1}^k \left(\sum_{j \in F_{k,n}^{-1}(i)} x_j\right) e_i$ where $F_{k,n} \colon [n] \to [k]$ is uniformly random, by definition of the de Finetti map in Section 4.5. Then

$$p_n^{(k,\lambda)}(x) = \frac{(k)_{\operatorname{len}(\lambda)}}{R_{\lambda,\lambda}} \mathbb{E}\left[\prod_{i=1}^{\operatorname{len}(\lambda)} \left(\sum_{j \in L_{k,n}^{-1}(i)} x_j\right)^{\lambda_i}\right] = \frac{(k)_{\operatorname{len}(\lambda)}}{R_{\lambda,\lambda}} \cdot \frac{1}{k^n} \sum_{f \colon [n] \to [k]} \prod_{i=1}^{\operatorname{len}(\lambda)} \left(\sum_{j \in f^{-1}(i)} x_j\right)^{\lambda_i},$$

where the first equality follows since $m_k^{(\lambda)}$ is a sum of $(k)_{\operatorname{len}(\lambda)}/R_{\lambda,\lambda}$ monomials by [209, Cor. 7.7.2, Prop. 7.8.3], and the second equality follows from the summands being permutations of each other and getting mapped to the same image. To simplify the above expression, note that

$$\prod_{i=1}^{\operatorname{len}(\lambda)} \left(\sum_{j \in f^{-1}(i)} x_j\right)^{\lambda_i} = \prod_{i=1}^{\operatorname{len}(\lambda)} \sum_{\substack{\mu^{(i)} \in \mathbb{N}^{f^{-1}(i)} \\ |\mu^{(i)}| = \lambda_i}} \frac{\lambda_i!}{\mu^{(i)}!} x^{\mu^{(i)}} = \sum_{\substack{\mu' \in \mathbb{N}^n \\ f(\mu') = \lambda}} \frac{\lambda!}{\mu'!} x^{\mu'},$$

where the first equality is just the multinomial theorem. For the second equality, split a multi-index $\mu' \in \mathbb{N}^n$ indexed by $[n]$ into multi-indices indexed by the fibers of f, i.e., $\mu^{(i)} = (\mu'_j)_{j \in f^{-1}(i)}$ for each $i \in [k]$. Now note that $x^{\mu'} = \prod_{i=1}^k x^{\mu^{(i)}}$ appears on the right-hand side if and only if $|\mu^{(i)}| = \lambda_i$, or more concisely, iff $f(\mu') = \lambda$, and each such μ' appears $\prod_{i=1}^{\operatorname{len}(\lambda)} \frac{\lambda_i}{\mu^{(i)}!} = \frac{\lambda!}{\mu'!}$ times. Finally,

$$\sum_{f \colon [n] \to [k]} \prod_{i=1}^{\operatorname{len}(\lambda)} \left(\sum_{j \in f^{-1}(i)} x_j\right)^{\lambda_i} = \sum_{\mu \leq_R \lambda} k^{n - \operatorname{len}(\mu)} R_{\mu,\lambda} \frac{\lambda!}{\mu!} \sum_{\mu' \in \mathsf{S}_n \mu} x^{\mu'} = \sum_{\mu \leq_R \lambda} k^{n - \operatorname{len}(\mu)} \frac{\lambda!}{\mu!} R_{\mu,\lambda} m_n^{(\mu)}(x),$$

where the first equality follows since if μ is the unique partition obtained as a permutation of a multi-index μ', then $\mu'! = \mu!$ and there are $R_{\mu,\lambda} k^{n - \operatorname{len}(\mu)}$ maps $f \colon [n] \to [k]$ such that $f(\mu') = \lambda$. Combining the two displayed equations above yields (4.56).

To see that $\{p_n^{(k,\lambda)}\}_{|\lambda|=d}$ is a basis for $\mathbb{R}[\mathbb{V}_n]^{\mathsf{S}_n}$, recall that $\{m_n^{(\lambda)}\}_{|\lambda|=d}$ forms such a basis for $n \geq d$. Moreover, the expression (4.56) shows that the linear map sending $(m_n^{(\lambda)})_\lambda$ to $(p_n^{(k,\lambda)})_\lambda$ is given by a matrix $M_{\lambda,\mu} = \frac{\lambda!}{\mu!} \cdot \frac{(k)_{\mathrm{len}(\lambda)}}{k^{\mathrm{len}(\mu)}} \cdot \frac{R_{\mu,\lambda}}{R_{\lambda,\lambda}}$, which is upper-triangular with respect to the refinement order and with the nonzero entries $M_{\lambda,\lambda} = \frac{(k)_{\mathrm{len}(\lambda)}}{k^{\mathrm{len}(\lambda)}}$ on the diagonal. It is therefore invertible, proving that $\{p_n^{(k,\lambda)}\}_\lambda$ is a basis. $\square$

Since $(p_n^{(k,\lambda)})_n$ is freely-described for each k, λ and since $\{p_n^{(k,\lambda)}\}_{|\lambda|=d}$ is a basis for invariant homogeneous polynomials of degree d for each n large enough, we conclude that $\{(p_n^{(k,\lambda)})_n\}_{|\lambda|=d}$ is a basis for freely-described homogeneous polynomials of degree d for each $k \geq d$. These bases do not appear to have been studied before.

Dual cost. For each $k \geq d$, expand $p_n = \sum_\lambda c^{(k,\lambda)} p_n^{(k,\lambda)}$ in the kth basis from Proposition 4.6.2. We can obtain $c^{(k,\lambda)}$ by first changing basis from power sums to monomial sums using (4.55), and then changing to the bases $\{p_n^{(k,\lambda)}\}$ using (4.56). The k'th dual cost is then given by $q_k = \sum_\lambda c^{(k,\lambda)} m_k^{(\lambda)}$, and the lower bounds on (4.53) are given by $\ell_n = \inf_{\substack{x \in \mathbb{R}^n \\ \|x\|_1 \leq 1}} q_n(x)$. We then have $\ell_n \leq \ell_{nk}$ for all n, k and $u_{nk} - \ell_{nk} \lesssim 1/\sqrt{n}$ for fixed k and growing n.

FinSet**-compatible sets.** The sequences of constraint sets ($\Omega_n \subseteq \mathbb{R}^n$) for which our framework is applicable consist of S_n-invariant compact sets satisfying $(x_1 + \ldots + x_k, \ldots, x_{nk-k+1} + \ldots + x_{nk}) \in \Omega_n$ if $x \in \Omega_{nk}$ and $(x, 0) \in \Omega_{n+1}$ if $x \in \Omega_n$. These assumptions are satisfied by the sequence of simplices $\Omega_n = \Delta^{n-1}$ and ℓ_1-balls, but not by ℓ_p-balls for any $p > 1$.

Example 4.6.3 (Quadratic symmetric functions)**.** *Any degree-2 symmetric polynomial can be written in the power sums basis as*

$$p_n(x) = \alpha_1 s_n^{(2)}(x) + \alpha_2 s_n^{(1,1)}(x).$$

For each $k \geq 2$, our bases (4.56) are given by

$$p_n^{(k,(1,1))}(x) = \frac{k(k-1)}{k^2} m_n^{(1,1)}(x) = \frac{1}{2}\left(1 - \frac{1}{k}\right)(s_n^{(1,1)}(x) - s_n^{(2)}(x)),$$
$$p_n^{(k,(2))}(x) = m_n^{(2)}(x) + \frac{2}{k} m_n^{(1,1)}(x) = \left(1 - \frac{1}{k}\right) s_n^{(2)}(x) + \frac{1}{k} s_n^{(1,1)}(x).$$

This gives a linear system we can invert to represent p_n in terms of our basis functions, yielding

$$p_n = (\alpha_1 + \alpha_2) p_n^{(k,(2))} + 2\left(\alpha_2 - \frac{1}{k-1}\alpha_1\right) p_n^{(k,(1,1))}.$$

The dual cost is then given by

$$q_n = (\alpha_1+\alpha_2)m_n^{(2)} + 2\left(\alpha_2 - \frac{1}{n-1}\alpha_1\right)m_n^{(1,1)} = \left(\frac{n\alpha_1}{n-1}\right)s_n^{(2)} + \left(\frac{\alpha_2(n-1)-\alpha_1}{n-1}\right)s_n^{(1,1)}.$$

As a simple example of our lower bounds, consider

$$u_n = \inf_{x\in\Delta^{n-1}} s_n^{(2)}(x) = \frac{1}{n},$$

for which $u_\infty = 0$. *For all* $n \geq 2$, *we get the lower bounds*

$$\ell_n = \inf_{x\in\Delta^{n-1}} \frac{n}{n-1}s_n^{(2)}(x) - \frac{1}{n-1}s_n^{(1,1)}(x) = 0.$$

For higher degree problems, we automate the derivation of q_n from p_n, as the relevant change-of-basis matrices can be formed and inverted computationally. The code to do so is available on our GitHub, using the code of [116] to form the matrices $R_{\mu,\lambda}$. We utilize this code below to prove lower bounds that hold in all dimensions for one of the examples of [3].

Example 4.6.4. *In [3, Thm. 3.6], the authors show that*

$$p_n = 4s_n^{(4)} - \frac{139}{20}s_n^{(3,1)} + 4s_n^{(2,2)} - 5s_n^{(2,1,1)} + 4s_n^{(1^4)}, \tag{4.57}$$

is a symmetric function which is nonnegative for all n but is not a sum of squares for any $n \geq 4$. *Since* p_n *is homogeneous of degree 4, certifying that* $p_n \geq 0$ *for all n amounts to certifying* $u_\infty \geq 0$ *where* $u_n = \inf_{\|x\|_1\leq 1} p_n(x)$. *Our framework yields finite-dimensional polynomial bounds* $\ell_n = \inf_{\|x\|_1\leq 1} q_n(x)$ *where*

$$q_n = \tfrac{n^3}{(n-1)(n-2)(n-3)}\Big[4(1+3\tfrac{1}{n})s_n^{(4)} - \tfrac{1}{20}(139 - 97\tfrac{1}{n} + 960\tfrac{1}{n^2})s_n^{(3,1)} + 4(1 - 9\tfrac{1}{n} + 9\tfrac{1}{n^2})s_n^{(2,2)}$$
$$- \tfrac{1}{20}(100 - 757\tfrac{1}{n} - 69\tfrac{1}{n^2})s_n^{(2,1,1)} + \frac{1}{10}(2-\tfrac{1}{n})(20 - 85\tfrac{1}{n} + 3\tfrac{1}{n^2})s_n^{(1^4)}\Big],$$

for $n \geq 4$. *Running both a nonconvex solver (Matlab's* `fmincon`*) and a degree-6 sums-of-squares relaxation, we find the lower bounds* ℓ_n *in Table 4.1 up to the displayed digits. Note that we are only guaranteed that* $\ell_{nk} \geq \ell_n$ *but the values we obtain increase monotonically with n.*

Finally, we return to Example 1.3.5 from Section 3.1, prove that the symmetric function there is not a sum-of-squares (SOS), and apply our framework to certify its nonnegativity.

n	4	5	6	7	8
ℓ_n	-0.8403	-0.4023	-0.2541	-0.1817	-0.1396

Table 4.1: Lower bounds over ℓ_1-ball for (4.57). Here $u_\infty = 0$.

Example 4.6.5. *Note that the symmetric function ($p_n = f(s_1, s_2, s_4, s_6)$) in (1.5) is not SOS in any dimension $n \geq 2$, since for $n = 2$ it reduces to $p_2(x) = x_1^4x_2^2 + x_1^2x_2^4 - \frac{3}{2}x_1^2x_2^2 + 1$. If this were SOS then the coefficient of $x_1^2x_2^2$ would be positive [173, Exer. 3.97], a contradiction. Similarly, if p_n admitted an SOS decomposition for $n > 2$ then restricting this decomposition to vectors of the form $(x_1, x_2, 0, \ldots, 0)$ would yield an SOS decomposition for p_2, a contradiction.*

Applying our framework to the freely-described problems $u_n = \inf_{x\in\mathbb{R}^n} p_n(x)$, we get the lower bounds $\ell_n = \inf_{x\in\mathbb{R}^n} q_n(x)$ with

$$q_n = \frac{n}{2(n-1)}\Bigg[m_n^{(4,2)} + m_n^{(4,1^2)} + 2m_n^{(3,2,1)} + \frac{6(n-2)}{n-3}m_n^{(3,1^3)} + 9m_n^{(2^3)} + \frac{3(3n-10)}{n-3}m_n^{(2^2,1^2)}$$
$$+ \frac{3(7n-18)}{(n-3)}m_n^{(2,1^4)} + \frac{45(n-2)(n-6)}{(n-3)(n-5)}m_n^{(1^6)} - 3m_n^{(2^2)} - 3m_n^{(2,1^2)} - \frac{9(n-2)}{n-3}m_n^{(1^4)}\Bigg] + 1.$$

for $n = 2, 4$ and $n \geq 6$. In particular, for $n = 2$ any monomial sum $m_n^{(\lambda)}$ with $\mathrm{len}(\lambda) \geq 3$ vanishes, so we get $q_2(x) = m_2^{(4,2)}(x) - 3m_2^{(2^2)}(x) + 1$, which is precisely the Motzkin polynomial.

We remark that Proposition 4.6.2 only guarantees the existence of $q_k \in \mathbb{R}[\mathbb{V}_k]$ satisfying $(p_n) = (\mathbb{E}q_k \circ L_{k,n})$ when $k \geq \deg p_n$. The denominators of the coefficients of q_k in the above examples vanish precisely for those $k < \deg p_n$ for which there is no such q_k.

Inequalities in Graph Homomorphism Densities

Model problem. We consider the problem of minimizing linear expressions in weighted graph homomorphism densities as in Example 1.3.3, which are freely-described problems over $\mathrm{Sym}^2\mathscr{D}$. A graph X on n vertices with weights on both its edges and vertices is represented by its adjacency matrix, which we also denote $X \in \mathbb{S}^n$ by abuse of notation, where $X_{i,i}$ is the weight on vertex i and $X_{i,j}$ is the weight on edge $\{i, j\}$. An unweighted graph H, possibly containing multiple edges and self-loops, is denoted $H \in \mathbb{S}^k(\mathbb{N})$, and the homomorphism number of H in X is defined by [146, Eq. (5.3)]

$$\hom(H,X) = \sum_{f\colon [k]\to[n]} \prod_{\substack{1\le i\le j\le k\\ H_{i,j}\neq 0}} X_{f(i),f(j)}^{H_{i,j}}, \tag{4.58}$$

which is a polynomial over $\mathbb{S}^n$. Note that $\hom(H;X)$ is unchanged if we relabel the vertices or append extra isolated vertices to either H or X; the former operation corresponds to conjugating the adjacency matrices by permutation matrices, and the latter to zero-padding the adjacency matrices.[5] The homomorphism density of H in X is defined as:

$$t(H,X) = \frac{\hom(H,X)}{n^k}, \tag{4.59}$$

where n^k is the number of all maps $[k]\to[n]$. This is once again a polynomial over $\mathbb{S}^n$. The homomorphism density only depends on the isomorphism class $[H]$ of H with all isolated vertices removed, and we denote the number of non-isolated vertices in H by $|V(H)|$ and the number of edges by $|H| = \sum_{i\le j} H_{i,j}$. Certifying inequalities in homomorphism densities over unweighted graphs amounts to considering the sequence of problems

$$u_n = \inf_{X\in\mathbb{S}^n(\{0,1\})} \sum\nolimits_j c_j t(H_j;X), \tag{4.60}$$

for some unweighted graphs H_j, and certifying that $u_\infty \ge 0$. We proceed to show that these problems are freely-described over $\mathrm{Sym}^2\mathscr{D}$. Conversely, we show that any such freely-described problem can be written as (4.60) for more general constraint sets Ω_n and for some $H_j \in \mathbb{S}^k(\mathbb{N})$.

Basis for $\mathscr{L} = \mathrm{Sym}^2\mathscr{Z}$. Define the injective homomorphism number $\mathrm{inj}(H,X)$ similarly to (4.58) except that we only sum over injective maps $[k]\hookrightarrow[n]$, and define $t_{\mathrm{inj}}(H,X) = \frac{\mathrm{inj}(H,X)}{(n)_k}$ to be the fraction of injective maps $[k]\hookrightarrow[n]$ that are homomorphisms, where $(n)_k = \frac{n!}{(n-k)!}$ is the total number of these injective maps. For $k\ge |V(H)|$, consider the monomial $X^H = \prod_{1\le i\le j\le k} X_{i,j}^{H_{i,j}} \in \mathbb{R}[\mathbb{S}^k]_{|H|}$ for $X\in\mathbb{S}^k$. Recalling the action ρ of maps between finite sets from Section 4.4, we have the following polynomial by symmetrizing X^H

$$\begin{aligned} t_{\mathrm{inj}}(H;X) = \frac{(k-|V(H)|)!}{k!} \sum_{f\colon [|V(H)|]\hookrightarrow[k]} (\rho(f)X)^H = \frac{1}{k!}\sum_{g\in\mathsf{S}_k} ((\zeta_{k,|V(H)|}^{\otimes 2})^\star g\cdot X)^H \\ = \mathrm{sym}_k^{\mathrm{Sym}^2\mathscr{Z}} X^H, \end{aligned} \tag{4.61}$$

[5] Some authors set $\hom(K_1;X) = n$ for $X\in\mathbb{S}^n$. We do *not* do so here, as otherwise we get different results depending on how many isolated vertices we append to a given graph. Our (4.58) gives $\hom(K_1;X) = 1$ for all X.

since a map $f\colon [|V(H)|] \to [k]$ is injective precisely when $f = g\iota_{k,|V(H)|}$ for some $g \in \mathsf{S}_k$, and $\rho(\iota_{k,|V(H)|}) = (\zeta_{k,|V(H)|}^{\otimes 2})^\star$. Thus, the polynomials $t_{\mathrm{inj}}(H;\cdot)$ are symmetrizations of monomials, in analogy to the monomials means from Section 4.6. For $k \geq |V(H)|$ denote by $q_k^{[H]}(X) = t_{\mathrm{inj}}(H;X)$ the restriction of $t_{\mathrm{inj}}(H;X)$ to $X \in \mathbb{S}^k$, which only depends on the isomorphism class $[H]$ of H. Observe that the collection $\{q_k^{[H]}\}_{|H|\leq d}$ as $[H]$ ranges over isomorphism classes of graphs with at most d edges forms a basis for $\mathbb{R}[\mathbb{S}^k]_{\leq d}^{\mathsf{S}_k}$ for all $k \geq 2d$. Indeed, any monomial over $\mathbb{R}[\mathbb{S}^k]_{\leq d}$ is of the form X^H for some graph H with $|H| \leq d$, which must have at most $2d$ non-isolated vertices, so symmetrizations of these monomials form a basis for S_k-invariants. Furthermore, (4.61) shows that $q_K^{[H]} = \mathrm{sym}_K^{\mathrm{Sym}^2\mathscr{D}} q_k^{[H]}$ for all $k \leq K$. Thus, the collection $\{(q_k^{[H]})_{k\geq 2d}\}_{|H|\leq d}$ is a basis for freely-symmetrized elements over $\mathrm{Sym}^2\mathscr{D}$ of degree $\leq d$, again recovering a special case of [139].

Basis for $\mathscr{U} = \mathrm{Sym}^2\mathscr{D}$. To derive $\mathbb{E}[q_k^{[H]} \circ L_{k,n}]$, observe that $(L_{k,n}X)_{i,j} = X_{F_{n,k}(i),F_{n,k}(j)}$ for $X \in \mathbb{S}^n$, where $F_{n,k}(1),\dots,F_{n,k}(k)$ are sampeld iid uniformly from $[n]$. Therefore,

$$\begin{aligned}\mathbb{E}[t_{\mathrm{inj}}(H;L_{k,n}X)] &= \mathbb{E}[\mathrm{sym}_{\mathsf{S}_k}(L_{k,n}X)^H] = \mathbb{E}[(L_{k,n}X)^H] = \frac{1}{n^k}\sum_{f\colon [k]\to[n]}(\rho(f)X)^H\\ &= t(H;X),\end{aligned}$$

where the first equality was shown above, and the second follows since $gL_{k,n} \overset{d}{=} L_{k,n}$ for all $g \in \mathsf{S}_k$. Denoting by $p_n^{[H]}(X) = t(H;X)$ the restriction of $t(H;X)$ to $X \in \mathbb{S}^n$, we conclude that $(p_n^{[H]})$ is freely-described over $\mathrm{Sym}^2\mathscr{D}$ for each H, and that the collection $\{(p_n^{[H]})_n\}_{|H|\leq d}$ forms a basis for such freely-described polynomials of degree d by Proposition 4.5.3.

Remark 4.6.6 (Connection to graphons)**.** *The above conclusions could have been derived using the connection between our generalized de Finetti theorem and the theory of graphons explained in Example 4.4.7. Briefly, a graphon is a symmetric, bounded, measurable function $W\colon [0,1]^2 \to \mathbb{R}$, and we can define the homomorphism densities $t(H;W)$ by an integral analog of (4.58), see [146, §7.2]. For any $X \in \mathbb{S}^n$, we can define a step graphon W_X as in Example 4.4.7, and note that $t(H;X) = t(H;W_X)$, which shows that $t(H;X)$ is freely-described over* $\mathrm{Sym}^2\mathscr{D}$ *as X and $X \otimes \mathbb{1}\mathbb{1}^\top$ define the same step graphon. Furthermore, note that $L_{k,n}X \overset{d}{=} (W_X(x_i,x_j))_{i,j\in[k]}$ where $x_1,\dots,x_k$ are sampled iid uniformly from $[0,1]$. The definition of $t(H;W)$ also shows that $\mathbb{E}[t_{\mathrm{inj}}(H;L_{k,n}X)] = \mathbb{E}[(L_{k,n}X)^H] = t(H;X)$.*

Dual cost. For a freely-described problem expressed in terms of homomorphism densities as in (4.60), form the dual cost $q_n(X) = \sum_j c_j t_{\mathrm{inj}}(H_j; X)$ for $n \geq V = \max_j |V(H_j)|$. The lower bounds on (4.60) are then given by $\ell_n = \inf_{X \in \mathbb{S}^n(\{0,1\})} q_n(X)$, which satisfy $\ell_n \leq \ell_{n+1}$ and $u_n - \ell_n \leq \frac{V(V-1)\sum_j |c_j|}{n}$ by Theorem 4.5.2, since $|t_{\mathrm{inj}}(H; X)| \leq 1$ for $X \in \mathbb{S}^n(\{0,1\})$. This is the convergence bound from Example 1.3.6 in Section 3.1.

FinSet-**compatible sets** The sequences of constraint sets ($\Omega_n \subseteq \mathbb{S}^n$) to which our framework is applicable consists of S_n-invariant sets satisfying $\zeta^\star_{n,n-1} X \zeta_{n,n-1} \in \Omega_{n-1}$ and $\delta_{nk,n} X \delta^\star_{nk,n} \in \Omega_{nk}$ whenever $X \in \Omega_n$, where ζ and δ are the zero-padding and duplication entries from Section 4.2. Examples of constraints satisfying the above hypotheses include entrywise constraints, such as the set of simple graphs $\{X \in \mathbb{S}^n : X_{i,j} \in \{0,1\}, X_{i,i} = 0\}$, and H-free simple graphs for any graph H. Non-examples include sets specified by spectral constraints such as $\lambda_{\max}(X) \leq 1$ or $\mathrm{Tr}(X) \leq 1$, as well as their normalized counterparts $\lambda_{\max}(X)/n \leq 1$ or $\overline{\mathrm{Tr}}(X) \leq 1$. In particular, the maximum eigenvalue constraint is not FinSet-compatible because $\lambda(X \otimes \mathbb{1}_k \mathbb{1}_k^\top) = k\lambda(X)$, and the latter is not compatible because extracting a principal submatrix might increase the average and the maximum eigenvalues.

Example 4.6.7 (Goodman's Theorem)**.** *Goodman's theorem states that $t(K_3; X) - 2t(K_2 \sqcup K_2; X) + t(K_2; X) \geq 0$ for all simple graphs X of all sizes, or equivalently, that $u_n = \inf_{X \in \mathbb{S}^n(\{0,1\})} t(K_3; X) - 2t(K_2 \sqcup K_2; X) + t(K_2; X)$ is equal to zero for all n (since $X = K_n$ gives objective value zero). Our framework yields the lower bounds $\ell_n = \inf_{X \in \mathbb{S}^n(\{0,1\})} t_{\mathrm{inj}}(K_3; X) - 2t_{\mathrm{inj}}(K_2 \sqcup K_2; X) + t_{\mathrm{inj}}(K_2; X)$ for $n \geq 4$. That is, the costs in the lower bounds are simply given by replacing t by t_{inj}. By Turan's theorem, the minimizer is attained at the complete bipartite graph $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$, so*

$$\ell_n = \begin{cases} -\frac{n}{2(n^2-4n+3)} & n \text{ even}, \\ -\frac{n+1}{2(n^2-2n)} & n \text{ odd}. \end{cases}$$

We thus have $u_n - \ell_n = \frac{1}{2n} + o(\frac{1}{n})$, while the bound from Theorem 4.5.2 gives $u_n - \ell_n \leq \frac{48}{n}$.

Example 4.6.8. *The* Ramsey multiplicity *of a graph H is defined as* $\mathrm{mult}(H) = u_\infty$ *where $u_n = \inf_{X \in \mathbb{S}^n(\{0,1\})} t(H; X) + t(H; 1 - X)$. Burr and Rosta [32] conjectured that* $\mathrm{mult}(H) = 2^{1-|H|}$ *for all graphs H, generalizing a conjecture of Erdös [69] for complete graphs. Both conjectures turned out to be false. Burr and Rosta's conjecture was disproved by Sidorenko [204], who showed that the triangle with a pendant* *edge H, the graph on $V(H) = [4]$ with edges $E(H) = \{\{1,2\},\{2,3\},\{2,4\}\{3,4\}\}$, has* $\mathrm{mult}(H) \leq 3/4 - 4\sqrt{2}/9 < 1/8$. *Since then, there has been much follow-up work yielding better upper bounds on Ramsey multiplicities, especially for complete graphs [171]. We use our framework to produce* lower *bounds on the Ramsey multiplicity of Sidorenko's counterexample above. Our bounds read $\ell_n = \inf_{X\in\mathbb{S}^n(\{0,1\})} t_{\mathrm{inj}}(H;X) + t_{\mathrm{inj}}(H;1-X)$, and we compute ℓ_n by exhaustively searching over all simple graphs with at most 10 vertices, obtaining the bounds in Table 4.2.*

n	≤ 6	7, 8	9, 10
ℓ_n	0	0.0286	0.0476

Table 4.2: Lower bounds for Ramsey multiplicity from Example 4.6.8.

Inequalities in Graph Homomorphism Numbers

Model problem. We study the problem of minimizing linear expressions in homomorphism *numbers*, which are freely-described problems over $\mathrm{Sym}^2\mathscr{Z}$. For example, consider certifying nonnegativity of a linear expression $\sum_j c_j \mathrm{hom}(H_j;X)$ in homomorphism numbers over all weighted graphs X, where hom is defined in (4.58). If $|H_j| = d$ for all j, i.e., all graphs H_j have the same number of edges d, then proving nonnegativity of the above expression is equivalent to considering

$$u_n = \inf_{\substack{X\in\mathbb{S}^n\\ \|X\|_1\leq 1}} \sum\nolimits_j c_j \mathrm{hom}(H_j;X), \tag{4.62}$$

and certifying $u_\infty \geq 0$. We apply our framework to derive lower bounds on u_∞ in terms of freely-symmetrized problems over $\mathrm{Sym}^2\mathscr{D}$. The development in this section parallels that in Section 4.6, as we observe below.

Basis for $\mathscr{L} = \mathrm{Sym}^2\mathscr{D}$. If $H \in \mathbb{S}^k(\mathbb{N})$ denote by $X^H = \prod_{1\leq i\leq j\leq k} X_{i,j}^{H_{i,j}} \in \mathbb{R}[\mathbb{S}^k]_{|H|}$ as before. We now view H as an element of $\mathbb{S}^n(\mathbb{N})$ for any $n \geq k$ by zero-padding it, and define the polynomial over $\mathbb{S}^n$

$$m_n^{[H]}(X) = \sum_{G\in\mathsf{S}_n H} X^G,$$

which is the sum over all distinct monomials X^G obtained by relabeling the vertices of H by elements from $[n]$. These are the analogs of monomial sums from Section 4.6, with isomorphism classes of graphs $[H]$ replacing partitions λ; indeed, the monomial sums of Section 4.6 can be obtained by considering graphs with only self-loops whose adjacency matrix is $\mathrm{diag}(\lambda)$. Note that $m_n^{[H]}(X)$ counts the number of times H appears as a subgraph of X up to isomorphism (i.e., $m_k^{[H]}(H) = 1$, for instance), so

$$m_n^{[H]}(X) = |\mathrm{Aut}(H)|^{-1}\mathrm{inj}(H;X), \tag{4.63}$$

where $|\mathrm{Aut}(H)|$ is the number of automorphisms of H without isolated vertices, which only depends on $[H]$. Since $m_k^{[H]}(X)$ is a multiple of $t_{\mathrm{inj}}(H;X)$ for any $k \geq |V(H)|$, we conclude that the collection $\{m_k^{[H]}\}_{|H|\leq d}$ as $[H]$ ranges over isomorphism classes of graphs with at most d edges forms a basis for $\mathbb{R}[\mathbb{S}^k]_{\leq d}^{\mathsf{S}_k}$ for all $k \geq 2d$, as the same is true for t_{inj} as shown in Section 4.6.

Basis for $\mathscr{U} = \mathrm{Sym}^2\mathscr{Z}$ In order to derive $\mathbb{E}[m_k^{[H]} \circ L_{k,n}]$, we generalize the refinement partial order on partitions from Section 4.6 to graphs and prove an analog of Proposition 4.6.2. This refinement partial order on graphs does not seem to have been studied before. If $X \in \mathbb{S}^n$ and $f\colon [n] \to [k]$, let $P_f = (f^{-1}(1), \ldots, f^{-1}(k))$ be the ordered partition of $[n]$ induced by f and denote $X/P_f = \rho(f)X = \sum_{i,j}(\sum_{(k,m)\in f^{-1}(i,j)} X_{k,m})E_{i,j}$ where $E_{i,j} = e_ie_i^\top$ if $i = j$ and $E_{i,j} = e_ie_j^\top + e_je_i^\top$ otherwise. The reason for this notation is that fX is the adjacency matrix of the quotient graph X/P_f obtained by identifying all vertices of X in each partition of P_f, with the weight of an edge $\{P_i, P_j\}$ being the sum of the weights in X of the edges $\{k, m\}$ with $k \in P_i$ and $m \in P_j$. See Figure 4.1 for an example, and see [146, Eq. (6.2)] for an example usage of this quotient. We denote $G \leq_R H$ if G is a

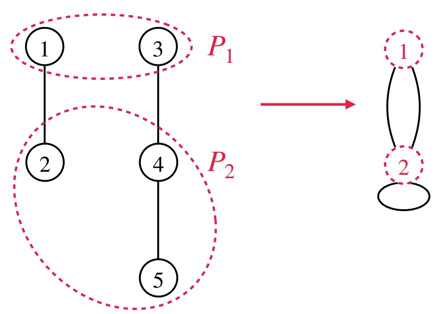


Figure 4.1: Example of graph quotient. Note that quotients of a simple graph need not be simple.

refinement of H, meaning that there is a surjective $f\colon V(G) \to V(H)$ such that $H = G/P_f$. For example, we show all refinements of K_3 in Figure 4.2. Note that if $G \leq_R H$ then $\sigma G \leq_R \tau H$ for any permutations σ, τ, so we also write $[G] \leq_R [H]$. If $G \leq_R H$ and neither G nor H have isolated vertices, we let $R_{[G],[H]}$ be the number of such surjective maps f. Note that this extends our notation for partitions in Section 4.6 upon identifying partitions with diagonal matrices. In particular, note the graph analog of the relation (4.55) between power sums and monomial sums

$$\hom(H;X) = \sum_{[H]\le_R[G]} |\mathrm{Aut}(G)|^{-1} R_{[H],[G]} \mathrm{inj}(G;X) = \sum_{[H]\le_R[G]} R_{[H],[G]} m_{[G]}(X), \tag{4.64}$$

which follows since each homomorphism $f\colon H \to X$ can be factored into two homomorphisms $H \to H/P_f \hookrightarrow X$, and conversely every composition of a quotient map $H \to G$ with an injective homomorphism $G \hookrightarrow X$ gives a homomorphism $H \to X$ unique up to automorphisms of H. This shows that the homomorphism numbers are the graph analogs of power sums. Indeed, if $H = (k)$ is the graph on a single vertex with k self loops, then $\hom(H;X) = \sum_{i=1}^n X_{i,i}^k$. Using the

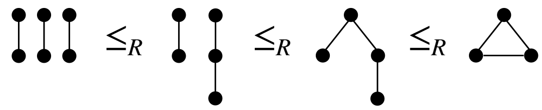


Figure 4.2: All refinements of K_3. Note that refinements of a simple graph are simple.

above refinement order and the associated integers $R_{[G],[H]}$, we are ready to derive $\mathbb{E}[m_k^{[H]} \circ L_{k,n}]$ in analogy with Proposition 4.6.2.

Proposition 4.6.9. *For any $k \ge |V(H)|$, we have that $p_n^{(k,[H])}(x) = \mathbb{E}[m_k^{[H]}(L_{k,n}X)]$ for $X \in \mathbb{S}^n$ is given by*

$$p_n^{(k,[H])}(X) = \sum_{[G]\le_R[H]} \frac{H!}{G!} \cdot \frac{(k)_{|V(H)|}}{k^{|V(G)|}} \cdot \frac{R_{[G],[H]}}{R_{[H],[H]}} m_n^{[G]}(X), \tag{4.65}$$

where $H! = \prod_{i\le j} H_{i,j}!$. As $[H]$ ranges over classes of graphs with at most d edges, each of the collections $\{(p_n^{(k,[H])})_n\}_{|H|\le d}$, $\{(m_n^{[H]})_n\}_{|H|\le d}$, $\{(\mathrm{inj}(H;\cdot)|_{\mathbb{S}^n})_n\}_{|H|\le d}$, and $\{(\hom(H;\cdot)|_{\mathbb{S}^n})_n\}_{|H|\le d}$ forms a basis for freely-described polynomials of degree d over $\mathrm{Sym}^2\mathscr{X}$.

Observe that (4.65) reduces to (4.56) when $H = \mathrm{diag}(\lambda)$ consists only of self loops. In fact, this entire proposition and its proof reduce to Proposition 4.6.2 in this way.

Proof. Observe that if $X \in \mathbb{S}^n$, then $L_{k,n}X \overset{d}{=} X/P_{n,k}$ where $P_{n,k}$ is a uniformly random ordered partition of $[n]$ into k parts. Indeed, we have $P_{n,k} \overset{d}{=} P_{L_{n,k}}$ since ordered partitions of $[n]$ into k parts are in bijection with maps $[n] \to [k]$, hence a uniformly random element of the former is induced by a uniformly random element of the latter. We then have $L_{k,n}X = X/P_{L_{k,n}} \overset{d}{=} X/P_{n,k}$ by definition. Therefore, for any $k \geq |V(H)|$ we have

$$p_n^{(k,[H])}(X) = |\mathsf{S}_k H| \cdot \mathbb{E}[(X/P_{n,k})^H] = \frac{(k)_{|V(H)|}}{R_{[H],[H]}} \cdot \frac{1}{k^n} \sum_{f\colon [n]\to[k]} \prod_{i_1\leq i_2} \left(\sum_{(j_1,j_2)\in f^{-1}(i_1,i_2)} X_{j_1,j_2} \right)^{H_{i_1,i_2}},$$

since $|\mathsf{S}_k H| = (k)_{|V(H)|}/|\mathrm{Aut}(H)|$ and $|\mathrm{Aut}(H)| = R_{[H],[H]}$. Here $f^{-1}(i_1,i_2) = \{(j_1,j_2) \in [n]^2 : f(j_1) = i_1, f(j_2) = i_2\}$. Now observe that

$$\prod_{i_1\leq i_2} \left(\sum_{(j_1,j_2)\in f^{-1}(i_1,i_2)} X_{j_1,j_2} \right)^{H_{i_1,i_2}} = \prod_{i_1\leq i_2} \sum_{\substack{G^{(i_1,i_2)}\in\mathbb{N}^{f^{-1}(i_1,i_2)} \\ |G^{(i_1,i_2)}|=H_{i_1,i_2}}} \frac{H_{i,j}!}{G^{(i_1,i_2)}!} X^{G^{(i_1,i_2)}} = \sum_{\substack{G\in\mathbb{S}^n(\mathbb{N}) \\ G/P_f=H}} \frac{H!}{G!} X^G,$$

where $\mathbb{N}^{f^{-1}(i_1,i_2)}$ consists of vectors indexed by $f^{-1}(i_1,i_2)$ with entries in $\mathbb{N}$. The first equality above is just the multinomial theorem. To see the second equality, note the bijection between $G \in \mathbb{N}^{n\times n} \cap \mathbb{S}^n$ such that $G/P_f = H$ and sequences $(G^{(i_1,i_2)} \in \mathbb{N}^{f^{-1}(i_1,i_2)})_{i_1\leq i_2}$ with $|G^{(i_1,i_2)}| = H_{i_1,i_2}$ given by $G \mapsto G^{(i_1,i_2)}_{j_1,j_2} = G_{j_1,j_2}$ and $(G^{(i_1,i_2)}) \mapsto G_{j_1,j_2} = G^{(i_1,i_2)}_{j_1,j_2}$ if $(f(j_1), f(j_2)) = (i_1,i_2)$ or (i_2,i_1). Under this bijection, $X^G = \prod_{j_1\leq j_2} X^{G_{j_1,j_2}} = \prod_{i_1\leq i_2} X^{G^{(i_1,i_2)}}$, which proves the second equality above. Next,

$$\begin{aligned}
\sum_{f\colon [n]\to[k]} \sum_{\substack{G\in\mathbb{S}^n(\mathbb{N}) \\ G/P_f=H}} \frac{H!}{G!} X^G &= \sum_{\substack{G\in\mathbb{S}^n(\mathbb{N}) \\ [G]\leq_R [H]}} k^{n-|V(G)|} R_{[G],[H]} \frac{H!}{G!} X^G \\
&= \sum_{[G]\leq_R [H]} k^{n-|V(G)|} R_{[G],[H]} \frac{H!}{G!} m_{[G]}^{(n)}(X),
\end{aligned}$$

where the first equality follows since for each $G \leq_R H$ there are $R_{[G],[H]} k^{n-|V(G)|}$ maps $f\colon [n] \to [k]$ satisfying $G/P_f = H$ (note that there are $n - |V(G)|$ isolated vertices above), and the second follows since $|V(G)|$, $R_{[G],[H]}$ and $G!$ only depend on G via $[G]$. Combining the above three displayed equations we get (4.65).

For the last claim, it is clear that the monomial sums $\{(m_n^{[H]})_n\}_{|H|\leq d}$ form a basis for freely-described polynomials, since $\{m_n^{[H]}\}_{|H|\leq d}$ is a basis for invariant polynomials over $\mathbb{S}^n$ for each n. The fact that the collection $\{(p_n^{(k,[H])})_n\}_{|H|\leq d}$ forms a basis follows from (4.65) expressing $p_n^{(k,[H])}$ in terms of $m_n^{[H]}$, since the matrix sending $(m_n^{[H]})_{|H|\leq d}$ to $(p_n^{(k,[H])})_{|H|\leq d}$ is invertible. Indeed, this matrix M has entries $M_{[G],[H]} = \frac{H!}{G!} \cdot \frac{(k)_{|V(H)|}}{k^{|V(G)|}} \cdot \frac{R_{[G],[H]}}{R_{[H],[H]}}$ and is therefore upper-triangular in refinement order with nonzero entries on the diagonal $M_{[H],[H]} = \frac{(k)_{|V(H)|}}{k^{|V(H)|}}$. Thus, for each

$n \geq 2d$ the collection $\{p_n^{(k,[H])}\}_{|H|\leq d}$ is a basis for invariant polynomials of degree at most d, showing that the freely-described polynomials $\{(p_n^{(k,[H])})_n\}_{|H|\leq d}$ form a basis as well. Further, we have $\mathrm{inj}(H;G) = |\mathrm{Aut}(H)|m_n^{[H]}(G)$ for all H, G, hence $\{(\mathrm{inj}(H;\cdot)|_{\mathbb{S}^n})_n\}_{|H|\leq d}$ also gives a basis. The relation (4.64) shows that the matrix expressing homomorphism numbers in terms of injective numbers is upper triangular in refinement order and has nonzero entries on the diagonal $R_{[H],[H]} = |\mathrm{Aut}(H)| \geq 1$. It is therefore invertible so we conclude that $\{(\mathrm{hom}(H;\cdot)|_{\mathbb{S}^n})_n\}_{|H|\leq d}$ is also a basis. $\square$

We are not aware of prior studies on the refinement partial order on graphs, or of graph functions of the form (4.65). In particular, it would be of interest to efficiently compute the integers $R_{[G],[H]}$ given two graphs, in analogy to the code [116] for the diagonal case.

Dual cost. For each $k \geq 2d$, expand the freely-described cost as $p_n = \sum_{[H]} c^{(k,[H])}p_n^{(k,[H])}$ in the k'th basis from Proposition 4.6.9. If p_n is given in terms of (injective) homomorphism numbers as in (4.62), we can obtain $c^{(k,[H])}$ by first changing basis to $\{(m_n^{[H]})\}$ using (4.64) and (4.63), and then changing to the bases $\{(p_n^{(k,[H])})\}$ using (4.65). The k'th dual cost is then given by $q_k = \sum_{[H]} c^{(k,[H])}m_k^{[H_k]}$, and the lower bounds on (4.62) are given by $\ell_n = \inf_{\substack{X\in\mathbb{S}^n\\ \|X\|_1\leq 1}} q_n(X)$. These satisfy $\ell_n \leq \ell_{nk}$ for all nk and $u_{nk} - \ell_{nk} \lesssim 1/\sqrt{n}$ for fixed k and growing n. Similar to Section 4.6, performing the preceding steps for some k in principle yields the dual costs via symmetrization for all q_{nk}; however, the sequence of monomial sums $(m_k^{[H]})$ is not freely-symmetrized over $\mathrm{Sym}^2\mathscr{D}$, so it is not in general true that $m_{nk}^{([H])} = \mathrm{sym}_{nk}^{\mathrm{Sym}^2\mathscr{D}} m_k^{([H])}$. As discussed at the beginning of Section 4.6, we have an explicit characterization of $(m_k^{([H])})$ as a function of k in (4.63), so it is instead more convenient to directly perform the preceding three steps starting from the monomial sums $m_{nk}^{([H])}$ to obtain q_{nk}.

FinSet**-compatible sets.** The sequence of constraint sets $(\Omega_n \subseteq \mathbb{S}^n)$ to which our framework applies consists of S_n-invariant sets satisfying $\zeta_{n+1,n}X\zeta_{n+1,n}^\star \in \Omega_{n+1}$ if $X \in \Omega_n$ and $\delta_{nk,n}^\star X\delta_{nk,n} \in \Omega_n$ if $X \in \Omega_{nk}$, where ζ and δ are the zero-padding and duplication entries from Section 4.2. As in Section 4.6, these assumptions are satisfied by the simplices $\{X \in \mathbb{S}^n : X \geq 0, \mathbb{1}^\top X\mathbb{1} = 1\}$, but not (for example) by their spectral analogues $\{X \in \mathbb{S}^n : X \succeq 0, \mathrm{Tr}(X) = 1\}$.

Example 4.6.10. *Let* $G_1 = K_2 \sqcup K_2 \sqcup K_2 \leq_R G_2 = K_2 \sqcup P_3 \leq_R G_3 = P_4 \leq_R G_4 = K_3$

the chain in Figure 4.2. Then

$$R_{[G_1],[G_2]} = 48, \quad R_{[G_2],[G_3]} = 8, \quad R_{[G_3],[G_4]} = 6, \quad R_{[G_4],[G_4]} = 6,$$

and since this is a chain in refinement order, we have $R_{[G_i],[G_j]} = \prod_{k=j+1}^{j} R_{[G_{k-1}],[G_k]}$ *for* $i < j$. *Since these are all simple graphs, we also have* $G_i! = 1$ *for all* i. *Therefore, we have*

$$p_n^{(k,[K_3])} = \frac{k(k-1)(k-2)}{k^3}\left(m_n^{([K_3])} + \frac{1}{k}m_n^{([P_4])} + \frac{8}{k^2}m_n^{([K_2\sqcup P_3])} + \frac{384}{k^3}m_n^{([K_2\sqcup K_2\sqcup K_2])}\right).$$

Example 4.6.11. *In Example 1.3.7 from Section 3.1, we consider the freely-described problem* (1.8)*, where we seek the largest* $\gamma \in \mathbb{R}$ *satisfying* $\mathrm{inj}(P_3; X) - \mathrm{inj}(K_2 \sqcup K_2; X) \geq \gamma \mathrm{hom}(K_2; X)^2$ *for all* $X \geq 0$ *of all sizes. Applying Proposition 4.6.9, we obtain the lower bounds* (1.9) *for* $n \geq 4$. *We numerically compute upper bounds on* u_n *and intervals containing* ℓ_n *for* $n \in [4, 9]$, *with the results shown in Table 1.1. The upper bounds on* u_n, ℓ_n *are computed using a nonconvex solver (Matlab's* `fmincon`*), while the lower bounds on* ℓ_n *are computed using a degree-*4 *sums-of-squares relaxation.*

4.7 Discussion

We have studied sequences of POPs of growing dimension and presented a framework to produce lower bounds on their limiting optimal values in terms of finite-dimensional POPs. Our framework is based on reformulating the limiting optimal value as an optimization problem over measures projecting onto each other. To produce convergent sequences of finite-dimensional relaxations for such problems, we prove new generalizations of de Finetti's theorem approximating such sequences of measures by particularly simple ones. Our proofs are based on a new perspective on de Finetti theorems via representations of categories, which also underpins results from representation stability that we use throughout the paper. Finally, we demonstrated our bounds on a range of numerical experiments. There are a number of questions suggested by our work.

(Tightness of rates) Is the $1/\sqrt{n}$ rate given in Theorem 4.5.2 tight for freely-described problems over $\mathscr{L}$-sequences? In all of our examples we see a rate of $1/n$ or faster, though the $1/\sqrt{n}$ rate for the corresponding de Finetti theorem is indeed tight (see Section 4.4).

(Conjecture) Is our conjectured strengthening of Proposition 4.5.4 true? In particular, can our proofs of it for the special cases in Sections 4.6 and 4.6 be extended to an arbitrary $\mathscr{D}$-sequence?

(Beyond permutations) Our generalizations of de Finetti's theorem relied on FinSet-representations, and hence only applies to representations of the permutation groups S_n. How can we further generalize these theorems to sequences of representations of other groups, in particular, the other classical Weyl groups and Lie groups?

(Finite step convergence) For which freely-described problems do we have $\ell_m = u_\infty$ for some finite $m \in \mathbb{N}$, i.e., when does our hierarchy of lower bounds converge in finitely-many steps? When does it converge in one step? From our numerical experiments, it seems that this is the case for generic symmetric functions, for instance. Finite-step convergence would imply a finite-dimensional reformulation for u_∞.

(Semialgebraicity of limiting cones) When is the cone of nonnegative freely-described polynomials studied in Section 4.5 semialgebraic? This would imply the decidability of checking $u_\infty \geq 0$.

(Any-dimensional positivstellensätze) Are there any-dimensional analogs of fixed-dimensional Positivstellensätze, such as those due to Krivine, Schmüdgen, and Putinar? Can we obtain such results for all freely-described costs (p_n) or just a subset of them? How should the sequence of constraints (Ω_n) be described for such results to hold?

(Bases for symmetric functions) Classical bases for symmetric functions and the relations between them have rich connections to combinatorics and representation theory [151, 209]. It would be interesting to develop such connections for the de Finetti maps acting on these symmetric functions, as well as the bases $(p_n^{(k,\lambda)})$ and $(p_n^{(k,[H])})$ over vectors and graphs they produce (see Propositions 4.6.2 and 4.6.9).

Chapter 5

LIMITS OF WEIGHTED GRAPHS VIA RANDOM QUOTIENTS

5.1 Introduction

To set the stage, we define a random quotient of a graph on n vertices in terms of a uniformly random map $F_{k,n}\colon [n] \to [k]$ sending each $i \in [n]$ independently to a uniformly-random element $F_{k,n}(i) \in [k]$. Equivalently, this is a uniformly random ordered partition of $[n]$ into k parts, given by the sequence of (random) fibers $F_{k,n}^{-1}(j)$ for $j \in [k]$. We denote by $\Delta^{n\times n}$ the set of $n \times n$ matrices with nonnegative entries that sum to one.

Definition 5.1.1. *A sequence of weighted graphs* $(G_n \in \Delta^{n\times n})_{n=1}^{\infty}$ *with edge weights that sum to one* quotient-converges *if for each* $k \in \mathbb{N}$ *the sequence of random graphs* $(\rho(F_{k,n})G_n \in \Delta^{k\times k})_{n=1}^{\infty}$ *converges weakly. A general sequence of nonzero weighted graphs* $(G_n \in \mathbb{R}_+^{n\times n})_n$ quotient-converges *if* $(G_n/\mathbb{1}^\top G_n \mathbb{1} \in \Delta^{n\times n})_n$ *quotient-converges.*

Normalizing our graphs to have unit total edge weight does not change the structure of its hubs, which were defined above in terms of the *fraction* of edge weight incident on them. Unless specified otherwise, we will assume throughout that the total edge weight of our graphs equals one.

Denoting by $\rho(F_{n,k})^\star\colon \mathbb{R}^{n\times n} \to \mathbb{R}^{k\times k}$ the adjoint of $\rho(F_{n,k})$ with respect to the Frobenius inner product, note that $\rho(F_{n,k})^\star G_n$ is the induced subgraph on k vertices sampled with replacement from the vertex set $[n]$ of G_n, in terms of which graphon convergence is defined [146, §9]. This is a first sense in which our theory is dual to that of graphons. We remark that there is an alternative view of graphon limits via convergence of quotients [146, §12], but these latter quotients differ from ours. In particular, these latter quotients cannot detect hubs in a large graph, as we explain in Section 5.1 below.

Combinatorial Aspects of Quotient Convergence (Section 5.2)

As our first contribution, we analyze combinatorial aspects of quotient convergence in Section 5.2. Whereas previous notions of graph convergence were based on convergence of local "motifs" in the form of homomorphism densities (appropriately normalized), we show that our quotient convergence is equivalent to convergence of homomorphism *numbers*. More precisely, for a multigraph $H \in \mathbb{N}^{k\times k}$ possibly containing parallel edges and self-loops, and a general weighted graph $G \in \mathbb{R}^{n\times n}$ on n vertices, the homomorphism number and homomorphism density of H in G are given by [146, Eq. (5.3)]

$$\hom(H;G) = \sum_{f\colon [k]\to[n]} \prod_{i,j=1}^{k} G_{f(i),f(j)}^{H_{i,j}}, \qquad t(H;G) = \frac{1}{n^k}\hom(H;G). \tag{5.1}$$

The terminology comes from the fact that when H and G are simple graphs, then $\hom(H;G)$ is the number of graph homomorphisms from H to G and $t(H;G)$ counts the average number of times H occurs in a uniformly random subgraph of G. In particular, convergence of the homomorphism densities $t(H;\cdot)$ is used to define limits of dense graphs [20, 147].

In analogy, we define the *quotient density* $t_Q(H;G)$ of a multigraph $H \in \mathbb{N}^{k\times k}$ in a general weighted graph $G \in \mathbb{R}^{n\times n}$ as the average number of times H occurs in a uniformly random quotient of G:

$$t_Q(H;G) = \frac{1}{k^n} \sum_{f\colon [n]\to[k]} \prod_{i,j=1}^{k} (\rho(f)G)_{i,j}^{H_{i,j}}. \tag{5.2}$$

The role of homomorphism densities in graphon theory is played by quotient densities in our framework. We prove that a sequence of graphs is quotient-convergent if all of their quotient densities converge.

Theorem 5.1.2. *For a sequence of graphs $(G_n \in \Delta^{n\times n})_n$, the following are equivalent:*

1. *The sequence (G_n) is quotient-convergent.*
2. *For each multigraph $H \in \mathbb{N}^{k\times k}$, the sequence of quotient densities $(t_Q(H;G_n))_n$ converges.*
3. *For each multigraph $H \in \mathbb{N}^{k\times k}$, the sequence of homomorphism numbers $(\hom(H;G_n))_n$ converges.*

The proof of this result is given in Section 5.2, and is based on expressing quotient densities in terms of (injective) homomorphism numbers. This expression involves a partial order on the collection of multigraphs induced by the quotient operation.

For simple graphs $(G_n \in \mathbb{S}^n_{\mathrm{sim}})_n$, i.e., undirected graphs with weights $(G_n)_{i,j} \in \{0,1\}$ and with no self-loops so $\mathrm{diag}(G_n) = 0$, we conclude that $(G_n/2|E(G_n)|)_n$ is quotient-convergent if and only if $\left(\frac{\hom(H;G_n)}{|E(G_n)|^{|E(H)|}}\right)_n$ converges for all simple graphs $H \in \mathbb{S}^k_{\mathrm{sim}}$ (see Corollary 5.2.7). Thus, quotient-convergence of simple graphs is equivalent to convergence of a homomorphism-density-like quantity, but one that is differently normalized than usual. To the best of our knowledge, such a normalization based on the number of edges has not appeared previously in the literature, as normalizations based on the numbers of vertices k and n are used (see Section 5.1). As a consequence, we will see below in our discussion of limit objects (Corollary 5.3.7) that quotient-convergent sequences of both dense and degree-bounded sequences of simple graphs converge to the same (uninteresting) limit. In fact, quotient limits of simple graphs can only detect stars, which are vertices on which a positive fraction of edges are incident. More broadly, hub structure in weighted and directed graphs can consist not only of (weighted and directed) stars, but also of individual edges carrying a significant fraction of the edge weight in the graph. Such structure is modeled by our limit objects for quotient-convergent sequences, which we proceed to describe in detail.

Limit Objects for Quotient Convergence (Section 5.3)

Our second contribution consists of constructing analytic limit objects for quotient-convergent sequences together with a notion of distance between them that allows us to compare graphs of different sizes, akin to graphons and the cut metric for dense graph sequences. Note that appending isolated vertices to a graph $G \in \Delta^{n\times n}$ or relabeling its vertices does not change the distribution of its quotients $\rho(F_{k,n})G_n$ for each $k \in \mathbb{N}$. Therefore, the limit object we associate to G must be the same as that associated to any other graph differing from it by extra isolated vertices and vertex relabeling. We describe our limit objects in terms of certain random exchangeable measures on the square, which we define next.

Definition 5.1.3 (Random exchangeable measures)**.** *A* random exchangeable probability measure *on* $[0,1]^2$ *is a random measure*[1] M *satisfying* $\mathsf{M} \overset{d}{=} \mathsf{M} \circ (\sigma,\sigma)$ *for all (Lebesgue-)measure-preserving bijections* $\sigma\colon [0,1] \to [0,1]$*. We say that a sequence of random measures* (M_n) *converges if it converges weakly viewed as a sequence of random variables taking values in the space of probability measures on* $[0,1]^2$ *(see Section 5.1 for details).*

[1]We endow the space of probability measures on $[0,1]^2$ with the weak topology and consider Borel random variables on it.

Any finite graph $G \in \Delta^{n\times n}$ can be associated with the random exchangeable measure:

$$\mathsf{M}_G = \sum_{i,j=1}^{n} G_{i,j}\delta_{(T_i,T_j)}, \quad \text{where } T_1,\ldots,T_n \overset{iid}{\sim} \mathrm{Unif}([0,1]). \tag{5.3}$$

Note that relabeling the vertices of G or appending isolated vertices to it yields the same random measure in this way. More generally, Kallenberg showed that all random exchangeable measures on $[0,1]^2$ have the form [113]:

$$\mathsf{M} = \sum_{i,j\in\mathbb{N}} E_{i,j}\delta_{(T_i,T_j)} + \sum_{i\in\mathbb{N}}[\sigma_i(\delta_{T_i}\otimes\lambda) + \varsigma_i(\lambda\otimes\delta_{T_i})] + \theta\lambda^2 + \vartheta\lambda_D, \tag{5.4}$$

where (i) λ is the uniform measure on $[0,1]$, λ^2 is the uniform measure on $[0,1]^2$, and λ_D is the uniform measure on the diagonal $\{(x,x) : x\in[0,1]\}$; (ii) $T_1,T_2,\ldots \overset{iid}{\sim}$ $\mathrm{Unif}([0,1])$; and (iii) $E_{i,j},\sigma_i,\varsigma_i,\theta,\vartheta \geq 0$ may potentially be random but must be independent of the (T_i) and must satisfy $\sum_{i,j} E_{i,j} + \sum_i(\sigma_i+\varsigma_i) + \theta + \vartheta = 1$ almost surely. Moreover, the collection of extremal random exchangeable measures, i.e., those that are not mixtures of other such random measures, are precisely the ones for which the coefficients $E,\sigma,\varsigma,\theta,\vartheta$ in (5.4) can be taken to be deterministic. We call such extremal random measures *grapheurs*, and show that these are precisely limits of quotient-convergent graph sequences.

Definition 5.1.4. *A* grapheur *(graph edge measure) is an extremal random exchangeable measure, i.e., one of the form* (5.4) *with deterministic coefficients* $(E,\sigma,\varsigma,\theta,\vartheta)$. *We denote the collection of grapheurs by*

$$\begin{aligned}\mathcal{M} = \Big\{\mathsf{M} \text{ of the form (5.4)} : E \in \mathbb{R}_{\geq 0}^{\mathbb{N}\times\mathbb{N}},\ \sigma,\varsigma\in\mathbb{R}_{\geq 0}^{\mathbb{N}},\ \theta,\vartheta\in\mathbb{R}_{\geq 0} \text{ s.t.}\\ \textstyle\sum_{i,j} E_{i,j} + \sum_i(\sigma_i+\varsigma_i) + \theta + \vartheta = 1\Big\}.\end{aligned} \tag{5.5}$$

The space of grapheurs inherits the weak topology from the ambient space of random exchangeable measures.

The coefficients $(E,\sigma,\varsigma,\theta,\vartheta)$ describing a fixed grapheur $\mathsf{M}\in\mathcal{M}$ are essentially unique. Specifically, note that for any permutation $\pi\colon\mathbb{N}\to\mathbb{N}$, the coefficients $(E,\sigma,\varsigma,\theta,\vartheta)$ and $(\pi E\pi^\top,\pi\sigma,\pi\varsigma,\theta,\vartheta)$ define the same random exchangeable distribution via (5.4). Further, adding zero rows to E and corresponding zero entries to σ, or zero columns to E and corresponding zero entries to ς, does not change the resulting random measure. Kallenberg showed in [113, Prop. 3] that these are the only possible ambiguities in the coefficients. In particular, note that $\mathsf{M}_G\in\mathcal{M}$ is a grapheur for any finite graph G, and that $\mathsf{M}_G = \mathsf{M}_{G'}$ for two such graphs G, G' if and only if they differ by isolated vertices and vertex relabeling.

In analogy with the cut metric between graphons, we endow $\mathcal{M}$ with the following metric between grapheurs that metrizes their convergence:

$$W_{\square}(\mathsf{M}_1, \mathsf{M}_2) = \inf_{\substack{\text{random } (\mathsf{M}_1', \mathsf{M}_2') \\ \mathsf{M}_i' \overset{d}{=} \mathsf{M}_i \text{ for } i=1,2}} \sup_{\substack{S,T \subseteq [0,1] \\ \text{intervals}}} \mathbb{E}|\mathsf{M}_1'(S \times T) - \mathsf{M}_2'(S \times T)|, \tag{5.6}$$

where the infimum is over couplings $(\mathsf{M}_1', \mathsf{M}_2')$ of the two random measures M_1 and M_2. Here $\mathsf{M}_i' \overset{d}{=} \mathsf{M}_i$ means that the distribution of M_i' is equal to that of M_i, see Section 5.1 for more detail. We show that grapheurs are the limits of quotient-convergent sequences of graphs in the following precise sense.

Theorem 5.1.5. *A sequence of grapheurs* $(\mathsf{M}_n) \subseteq \mathcal{M}$ *converges weakly to a grapheur* $\mathsf{M} \in \mathcal{M}$ *if and only if* $\lim_n W_{\square}(\mathsf{M}_n, \mathsf{M}) = 0$. *Furthermore, the following statements hold.*

1. *A sequence of graphs* $(G_n \in \Delta^{n \times n})$ *is quotient-convergent if and only if there is a grapheur* $\mathsf{M} \in \mathcal{M}$ *satisfying* $\lim_n W_{\square}(\mathsf{M}, \mathsf{M}_{G_n}) = 0$.
2. *The space of grapheurs* $\mathcal{M}$ *is compact. In particular, any sequence of graphs* $(G_n \in \Delta^{n \times n})$ *has a quotient-convergent subsequence.*
3. *For every grapheur* $\mathsf{M} \in \mathcal{M}$ *there is a sequence* $(G_n \in \Delta^{n \times n})$ *satisfying* $\lim_n W_{\square}(\mathsf{M}, \mathsf{M}_{G_n}) = 0$.

In view of Theorem 5.1.5, we say that a sequence of graphs (G_n) converges to a grapheur $\mathsf{M} \in \mathcal{M}$ if $\lim_n \mathsf{M}_{G_n} = \mathsf{M}$ in $\mathcal{M}$. Theorem 5.1.5 states that any quotient-convergent sequence of graphs converges to a grapheur, and conversely each grapheur is the limit of a quotient-convergent sequence of graphs. Moreover, this theorem shows that the space of grapheurs endowed with the weak topology is compact, and that this compact topology is metrized by our $W_{\square}$ metric (5.6).

To prove Theorem 5.1.5, we relate random quotients of finite graphs to their associated grapheurs (5.3). For each grapheur $\mathsf{M} \in \mathcal{M}$ and $k \in \mathbb{N}$, define the random graph $\mathsf{G}_k[\mathsf{M}] \in \Delta^{k \times k}$ whose entries are given by:

$$(\mathsf{G}_k[\mathsf{M}])_{i,j} = \mathsf{M}(I_i^{(n)} \times I_j^{(n)}) \tag{5.7}$$

for $i, j \in [n]$, where $I_i^{(n)} = [(i-1)/n, i/n)$ for $i \in [n]$.[2] In words, we partition the square $[0,1]^2$ into a $k \times k$ uniform grid, and let the entries of $\mathsf{G}_k[\mathsf{M}]$ be the (random) measure under M of each square in this grid. We show in Section 5.3 that $\rho(F_{k,n})G \overset{d}{=} \mathsf{G}_k[\mathsf{M}_G]$ for any finite graph $G \in \Delta^{n\times n}$, and that a sequence of graphs (G_n) converges to the grapheur M precisely when the random quotients $(\rho(F_{k,n})G_n)_k$ converge weakly to $\mathsf{G}_k[\mathsf{M}]$ for each $k \in \mathbb{N}$. Similarly, we can extend quotient densities to grapheurs by setting

$$t_Q(H; \mathsf{M}) = \mathbb{E} \prod_{i,j=1}^{k} (\mathsf{G}_k[\mathsf{M}])_{i,j}^{H_{i,j}} = \int \prod_{i,j=1}^{k} \mu(I_i^{(k)} \times I_j^{(k)})^{H_{i,j}} \, d\mathsf{M}(\mu), \tag{5.8}$$

for any multigraph $H \in \mathbb{N}^{k\times k}$, analogously to the extension of homomorphism densities to graphons [146, §7.2]. In terms of quotient densities, the sequence (G_n) converges to the grapheur M if and only if $t_Q(H; G_n) \to t_Q(H; \mathsf{M})$ for all multigraphs H, see Section 5.3.

Having characterized our limit objects, we interpret in Section 5.3 each of the parameters $E, \sigma, \varsigma, \theta, \vartheta$ describing a grapheur (5.4) in terms of certain features of the large graphs converging to it. Informally, the array $E \in \mathbb{R}_{\geq 0}^{\mathbb{N}\times\mathbb{N}}$ specifies the connections between a discrete collection of hubs, with each entry corresponding to an edge with an asymptotically positive fraction of the total edge weight. The arrays $\sigma, \varsigma \in \mathbb{R}_{\geq 0}^{\mathbb{N}}$ specify directed stars centered at hubs, with each entry corresponding to the fraction of out-degree or in-degree of a hub from all the edges incident to it with asymptotically vanishing edge weights. Lastly, the parameters θ and ϑ specify the remaining edge weight that comes from all edges between vertices with asymptotically vanishing degrees and their self-loops. In particular, we say that a grapheur M does not have hubs if $E = 0$ and $\sigma = \varsigma = 0$, and we say that a sequence (G_n) of graphs does not contain hubs asymptotically if it quotient-converges to such a grapheur.

Finally, we expand on the duality between graphons and grapheurs. We have seen previously that the sampling operators used to define each notion of graph limit, namely, those extracting random subgraphs and random quotients, respectively, are duals of each other. We further show in Section 5.3 that the space of graphons endowed with the cut metric is dual to the space $\mathcal{M}$ endowed with the $W_\square$ metric (5.6) as follows.

[2]Note that $\sum_{i,j}(\mathsf{G}_k[\mathsf{M}])_{i,j} = \mathsf{M}([0,1)^2) = 1$ almost surely by (5.4).

Theorem 5.1.6 (Informal; see Theorem 5.3.10)**.** *There is a random pairing $\langle \mathsf{M}, W\rangle$ associating a scalar random variable to each grapheur M and graphon W satisfying the following properties. We have $\lim_n \mathsf{M}_n = \mathsf{M}$ in the $W_\square$ metric if and only if $\langle \mathsf{M}_n, W_G\rangle \to \langle \mathsf{M}, W_G\rangle$ weakly for each step graphon W_G associated to a finite simple graph G. We also have $\lim_n W_n \to W$ in the cut metric if and only if $\langle \mathsf{M}_G, W_n\rangle \to \langle \mathsf{M}_G, W\rangle$ weakly for each grapheur M_G associated to a finite undirected graph G with unit total edge weight and no self-loops.*

Edge Sampling and Property Testing (Section 5.4)

As our next contribution, we apply our framework to test properties of large graphs based on small summaries of them. Specifically, we focus on testing properties given by graph parameters that are continuous with respect to quotient convergence, including quotient densities, homomorphism numbers, clustering and centrality coefficients, see Section 5.4. While it would be natural to test such properties using random quotients, forming these requires access to the entire large graph, which limits their utility in practice. Motivated by this challenge, we develop an edge-based sampling perspective on quotient-convergence in Section 5.4. Specifically, we sample edges from a graph proportionally to their edge weights to form random graphs $\mathsf{G}^{(n)}$ with n edges, and extend this sampling procedure to any grapheur $\mathsf{M} \in \mathcal{M}$. We detail the general sampling procedure in Section 5.4.

We prove the following convergence results for the sequence of random graphs $(\mathsf{G}^{(n)})$ sampled from a grapheur M back to M, which are edge-based analogs of the second sampling lemma and Szemerédi's regularity lemma for graphons [146, Lemma 10.16, Lemma 9.15]. Here and throughout the chapter, a superscript $G^{(n)}$ denotes a graph on n edges (and hence at most $2n$ vertices) while a subscript G_n denotes a graph on n vertices.

Theorem 5.1.7. *Fix any grapheur $\mathsf{M} \in \mathcal{M}$.*

(Edge sampling lemma) *For each $n \in \mathbb{N}$, let $\mathsf{G}^{(n)}$ be the random graph containing n edges sampled from M as in Section 5.4. With probability at least $1 - e^{-2\epsilon^2}$, we have that $W_\square(\mathsf{M}_{\mathsf{G}^{(n)}}, \mathsf{M}) \leq \frac{174+\epsilon}{\sqrt{n}}$.*

(Edge-based Szemerédi regularity) *For any $\epsilon > 0$, there exists a graph $G^{(k(\epsilon))}$ on $k(\epsilon) = \lceil (174/\epsilon)^2 \rceil$ edges satisfying $W_\square(\mathsf{M}, \mathsf{M}_{G^{(k(\epsilon))}}) \leq \epsilon$.*

Theorem 5.1.7 allows us to test properties of large graphs by sampling a few of their edges at random. We formally define the graph parameters that can be tested using our framework in Section 5.4. These consist of graph parameters that are continuous with respect to quotient convergence, and include all polynomial graph parameters that do not change when appending isolated vertices to their inputs (Proposition 5.4.9). In particular, the graph parameter $f(G) = \|G\mathbb{1}\|_2^2 + \|G^\top\mathbb{1}\|_2^2$ is quotient-testable in our framework, and can be used to test for the presence of hubs in a growing sequence of graphs. Specifically, if (G_n) is a quotient-convergent sequence, we show in Example 5.4.11 that $f(G_n) \to 0$ precisely when (G_n) does not have hubs asymptotically, i.e., it converges to a limiting grapheur (5.4) with $E = 0$ and $\sigma = \varsigma = 0$. We then describe an edge-based sampling procedure for testing this graph parameter. In contrast, this parameter cannot be tested in the graphons framework; see Section 5.4.

Equipartition-Consistent Random Graph Models (Section 5.5)

As our final contribution, we observe that random quotients $(\mathsf{G}_k[\mathsf{M}])_k$ of any grapheur M are related to each other by equipartitions of their vertex sets. Specifically, we observe that $\rho(d_{k,nk})\mathsf{G}_{nk}[\mathsf{M}] \overset{d}{=} \mathsf{G}_k[\mathsf{M}]$ for any equipartition $d_{k,nk}\colon [nk] \to [k]$, i.e., a map with equal-sized fibers. We call such sequences of random graphs *equipartition-consistent random graph models*, which play a role in our framework analogous to that played by consistent random graph models in graphon theory [148].

Definition 5.1.8 (Equipartition-consistent models)**.** *A sequence of probability measures $(\mu_k \in \mathcal{P}(\Delta^{k\times k}))_{k\in\mathbb{N}}$ is an* equipartition-consistent random graph model *if $\rho(d_{k,nk})\mathsf{G}_{nk} \overset{d}{=} \mathsf{G}_k$ whenever $\mathsf{G}_{nk} \sim \mu_{nk}$ and $\mathsf{G}_k \sim \mu_k$ and for any equipartition $d_{k,nk}\colon [nk] \to [n]$.*

Further, we show that any equipartition-consistent random graph model is derived from a mixture of grapheurs.

Theorem 5.1.9. *For any equipartition-consistent random graph model $(\mu_k \in \mathcal{P}(\Delta^{k\times k}))_k$, there is a unique random exchangeable measure M on $[0,1]^2$ satisfying $\mu_k = \mathrm{Law}(\mathsf{G}_k[\mathsf{M}])$ for all $k \in \mathbb{N}$.*

Recall that grapheurs are extremal random exchangeable measures by Definition 5.1.4. By Choquet's theorem, any random exchangeable measure is thus a mixture of grapheurs. Finally, we show in Proposition 5.5.1 that the sequence $(\mathsf{G}_k[\mathsf{M}])$ of quotients of a grapheur M converges back to M in a certain sense.

Related Work

There are several well-studied notions of graph limits and random graph models in the literature. We review this body of work and discuss similarities and differences with the present chapter.

Dense and degree-bounded graph limits. The earliest two types of graph limits pertain to sequences of simple graphs that are dense, whose limits are given by graphons [20, 147], and degree-bounded, whose limits are given by graphings [6, 15, 67]. We refer the reader to [146] for a survey of both limits. Both of these notions of graph limits are incomparable to ours, as both dense and degree-bounded sequences of simple graphs converge to the same limit in our framework, see Corollary 5.3.7. Further, dense graph limits can be characterized in terms of quotients [146, §12], called "right convergence" of graphons. However, these quotients are differently normalized compared to ours, and it is the set of quotients that converges in the graphon framework rather than their distribution as in our framework. Despite these distinctions, we highlight parallels between our development and the theory of graphons throughout this chapter; in particular, grapheurs can be viewed as dual to graphons in a precise sense, see Section 5.3.

Edge-exchangeable random graphs. A class of random graph models closely related to the ones we obtain by sampling edges from a grapheur is called edge-exchangeable [33, 34, 53, 54, 110]. These random graphs are obtained by adding random edges to a fixed infinite vertex set (often with repetition, yielding multigraphs), and can yield both sparse and dense random graphs [110]. As explained in [108, Rmk. 4.4], every edge-exchangeable multigraph model can be obtained by sampling edges from a random exchangeable measure on $[0, 1]^2$. In fact, the random graphs we obtain by sampling edges in Section 5.4 are precisely these edge-exchangeable multigraphs, normalized to have unit edge weight, see Remark 5.4.2. From this perspective, convergence of grapheurs can be interpreted as convergence of the edge-exchangeable multigraph models associated to them, an interpretation we make precise in Proposition 5.4.1. Nevertheless, to our knowledge this notion of convergence of edge-exchangeable models has not been previously studied in the literature. In particular, the only metric between edge-exchangeable models we are aware of is an ℓ_1 distance between edge probabilities [54, Eq. (8)]. Convergence in this metric is too restrictive for our purposes, see Proposition 5.3.8. Further, our development of convergence via random quotients, which are not edge-exchangeable, allows us to take limits of general weighted graphs, to relate it to convergence of homomorphism densities and other graph parameters, and to test their values.

Graphex processes and limits. The closest line of work to ours is the study of graphex processes and their associated graph limits and random graph models [22–25, 37, 109, 217]. Graphexes form a subset of random exchangeable measures on $\mathbb{R}^2_{\geq 0}$, and admit a characterization due to Kallenberg similarly to our grapheurs [113]. Graphexes give rise to random graph models obtained by sampling edges from them [24, 217, 218], analogously to our edge sampling procedure in Section 5.4. There are also several notions of limits for simple graphs and multigraphs, whereby such graphs converge to a limiting graphex [23, 25]. These notions of limits are conjectured to be inequivalent [109], and the space of graphexes with either topology is not compact. Some of these limits can be characterized by sampling random subgraphs (whose number of vertices is also random) [24], while others can be characterized by convergence of differently-normalized homomorphism numbers [25, §2.5] (the authors of [25] normalize $\hom(H; G)$ by $(2|E(G)|)^{|V(H)|/2}$, whereas we normalize by $(2|E(G)|)^{|E(H)|}$). Graphex-based graph limits are again incomparable to ours. On the one hand, for a sequence of simple graphs our grapheur-based limits can only distinguish stars from a uniform component (Corollary 5.3.7), whereas graphex-based limits can capture more general limiting structures (see [109, §9] or [23]). On the other hand, grapheur-based graph limits are defined for general sequences of nonnegatively-weighted graphs, for which graphex-based limits have not been defined, allowing us to test properties pertaining to the distribution of edge weight in the graph (see Section 5.4). In fact, grapheurs are particularly well-suited for property testing, since the space of grapheurs we study here is compact, in contrast to the space of graphexes; to our knowledge, there is no characterization of properties that can be tested using graphex-based limits. Another consequence of the compactness of the space of grapheurs is that our notion of graph limits can be characterized in several equivalent ways: by convergence of random quotients, convergence of random edge-sampled graphs, and convergence of homomorphism numbers and related graph parameters.

Notation and Preliminaries

We denote by $\mathbb{N} = \{0, 1, 2, \ldots\}$ the set of nonnegative integers. If $n \in \mathbb{N}$, then $[n] = \{1, \ldots, n\}$. If $a \leq b$ are real numbers then $[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}$ and $[a, b) = [a, b] \setminus \{b\}$. We denote by λ the uniform measure on $[0, 1]$, by λ^2 the uniform measure on $[0, 1]^2$, and by λ_D the uniform measure on the diagonal $D = \{(x, x) : x \in [0, 1]\}$. We denote by $\mathbb{1}_n$ the vector of all-1's of length n, and we omit the subscript n when the size (possibly infinite) of the all-1's vector is clear from context. We denote by I_n the $n \times n$ identity matrix and by $\mathbb{1}_{n\times n} = \mathbb{1}_n \mathbb{1}_n^\top$ the $n \times n$ all-1's matrix.

We identify a (directed, weighted) graph with its adjacency matrix $G \in \mathbb{R}^{n\times n}_{\geq 0}$, where $G_{i,j}$ is the weight of the edge $i \to j$. The collection of all adjacency matrices of graphs with unit total edge weight is denoted by $\Delta^{n\times n}$. In this chapter, all edge weights are nonnegative and sum to one unless stated otherwise. A graph is undirected if its adjacency matrix is symmetric. We denote the space of $n \times n$ symmetric matrices by $\mathbb{S}^n$, and the set of such matrices with nonnegative entries, which is the set of undirected graphs, by $\mathbb{S}^n_{\geq 0}$. An undirected graph G is simple if $\mathrm{diag}(G) = 0$ and $G_{i,j} \in \{0, 1\}$ for all i, j, and we denote the set of simple graphs by $\mathbb{S}^n_{\mathrm{sim}}$. A graph H is called a *multigraph* if it has integer edge weights $H \in \mathbb{N}^{k\times k}$, where we view $H_{i,j}$ as representing the number of parallel edges from i to j. Note that a simple graph is a multigraph. We denote by $|E(G)| = \mathbb{1}^\top G \mathbb{1}/2$ the number of edges in a *simple* graph G. If H is a multigraph, we write $\mathbb{1}^\top H \mathbb{1} = \|H\|_1$ to be the number of (directed) edges in H. We assume that multigraphs have no isolated vertices, defined as vertices with no self-loops and no edges incident on them. In other words, if $H \in \mathbb{N}^{k\times k}$ is a multigraph, then for every $i \in [k]$ there is a $j \in [k]$ such that $H_{i,j} > 0$ or $H_{j,i} > 0$. If $G \in \mathbb{R}^{n\times n}_{\geq 0}$ and $H \in \mathbb{N}^{k\times k}$ with $k \leq n$, we denote by $G^H = \prod_{i,j=1}^k G_{i,j}^{H_{i,j}}$ the monomial over $\mathbb{R}[G_{i,j}]$ defined by H. In this notation, we have $t_Q(H; G) = \mathbb{E}(\rho(F_{k,n})G)^H$ for a multigraph $H \in \mathbb{N}^{k\times k}$ and graph $G \in \Delta^{n\times n}$.

We denote by S_n the group of permutations of $[n]$, and by S_∞ the group of permutations of $\mathbb{N}$ that fixes all but finitely-many integers. If $\pi \in \mathsf{S}_n$ and $G \in \mathbb{R}^{n\times n}_{\geq 0}$ is a graph, we denote by $\pi G \in \mathbb{R}^{n\times n}_{\geq 0}$ the permuted graph with entries $(\pi G)_{i,j} = G_{\pi^{-1}(i),\pi^{-1}(j)}$. This corresponds to relabeling the vertices of G according to π. Similarly, if $x \in \mathbb{R}^n$ is a vector we denote by $\pi x \in \mathbb{R}^n$ the permuted vector with entries $(\pi x)_i = x_{\pi^{-1}(i)}$. If $G \in \mathbb{R}^{\mathbb{N}\times\mathbb{N}}_{\geq 9}$ and $x \in \mathbb{R}^{\mathbb{N}}$ are infinite matrices and vectors, and $\pi \in \mathsf{S}_\infty$, we define πG and πx analogously.

If $\mu \in \mathcal{P}(S)$ is a probability measure on a compact Polish space S, we write $X \sim \mu$ and $\mathrm{Law}(X) = \mu$ to denote a random variable distributed according to μ. If X, Y are two random variables on S, we write $X \overset{d}{=} Y$ to denote equality in distribution, i.e., the equality $\mathrm{Law}(X) = \mathrm{Law}(Y)$. If f is a function, we denote $\mathbb{E}_\mu f = \mathbb{E}_{X\sim\mu} f(X)$ its expectation with respect to μ. If (μ_n) is a sequence of measures in $\mathcal{P}(S)$, we say (μ_n) converges weakly to μ and write $\mu_n \to \mu$ if $\mathbb{E}_{\mu_n} f \to \mathbb{E}_\mu f$ for all continuous f. All measures in this chapter are supported in a single ambient compact set. In particular, all sequences of measures are tight.

A coupling of random variables $X_1, \dots, X_k$ taking values in Polish spaces $S_1, \dots, S_k$, respectively, is a probability distribution $\Gamma \in \mathcal{P}(S_1 \times \cdots \times S_k)$ such that if $(X'_1, \dots, X'_k) \sim \Gamma$ then $X'_i \overset{d}{=} X_i$ for all $i \in [k]$. In this chapter we refer to a coupling of $X_1, \dots, X_k$ simply by the tuple of jointly-distributed random variables $(X'_1, \dots, X'_k)$. If X_1, X_2 are random graphs taking values in $\mathbb{R}^{k \times k}$, we define the Wasserstein-1 distance between them by

$$W_1(X_1, X_2) = \inf_{\substack{\text{random } (X'_1, X'_2) \\ X'_i \overset{d}{=} X_i}} \mathbb{E}\|X'_1 - X'_2\|_1 = \sup_{\substack{f \colon \mathbb{R}^{k \times k} \to \mathbb{R} \\ \text{1-Lipschitz}}} |\mathbb{E} f(X_1) - \mathbb{E} f(X_2)|,$$

where the infimum is over couplings (X'_1, X'_2) of X_1 and X_2, the norm $\|\cdot\|_1$ is the entrywise ℓ_1 norm, and the supremum is over 1-Lipschitz functions with respect to this norm. Note that a sequence of random graphs $(X_n) \subseteq \Delta^{k \times k}$ converges weakly to the random graph $X \in \Delta^{k \times k}$ if and only if $\lim_n W_1(X_n, X) = 0$ by [220, Thm. 6.9].

For a compact Polish space S, the space $\mathcal{P}(S)$ endowed with the weak topology is itself compact and Polish. If M is a random measure, that is, a random variable taking values in the Polish space $\mathcal{P}(S)$, we write $\mathbb{E}_{\mathsf{M}} f$ for the scalar random variable obtained by integrating a deterministic function f against the random measure M. In particular, we write $\mathsf{M}(B) = \mathbb{E}_{\mathsf{M}} \mathbb{1}_B$ to denote the scalar random variable giving the (random) measure of a measurable set B, where $\mathbb{1}_B(x) = 1$ if $x \in B$ and zero otherwise. A sequence of random measures (M_n) converges weakly to M if $\mathrm{Law}(\mathsf{M}_n) \to \mathrm{Law}(\mathsf{M})$ weakly, where these laws are probability measures over the space of probability measures. Equivalently, the sequence (M_n) converges weakly to M if the sequence of scalar random variables $\mathbb{E}_{\mathsf{M}_n} f$ converges weakly to $\mathbb{E}_{\mathsf{M}} f$ for all continuous functions f [115, Thm. A2.3].

5.2 Quotient Convergence: Global and Local Views

In this section we prove Theorem 5.1.2 relating quotient convergence to convergence of homomorphism numbers and quotient densities. Before presenting this proof, we give a few examples illustrating sequences of graphs that are quotient-convergent. As our first example, we show that the sequence of complete graphs is quotient-convergent.

Example 5.2.1 (Convergence of complete graphs)**.** *To show that the sequence $(\mathbb{1}_{n \times n})_{n \in \mathbb{N}}$ of complete graphs converges, consider the associated normalized graphs $(G_n = \mathbb{1}_{n \times n}/n^2 \in \Delta^{n \times n})$ and observe that the random quotients are given as*

$\rho(F_{k,n})G_n = \frac{1}{n^2}NN^\top$ for a multinomial random variable $N = (N_1, \ldots, N_k) \sim \mathrm{Multinom}(n, k, \mathbb{1}_k/k)$. Since $\mathbb{E}\rho(F_{k,n})G_n = \frac{1}{k^2}\mathbb{1}_{k\times k}+O(1/n)$ and $\mathrm{Var}[\rho(F_{k,n})G_n]_{i,j} = O(1/n)$ for all $i, j \in [k]$, we conclude that $\rho(F_{k,n})G_n$ converges weakly to the deterministic matrix $\frac{1}{k^2}\mathbb{1}_{k\times k}$ as $n \to \infty$. Since this holds for all $k \in \mathbb{N}$, we conclude that (G_n) is quotient-convergent. Note that the quotient densities and homomorphism numbers of this sequence likewise converge. Indeed, for any $H \in \mathbb{N}^{k\times k}$ we have $t_Q(H; G_n) = \mathbb{E}(\rho(F_{k,n})G_n)^H \xrightarrow{n\to\infty} (\frac{1}{k^2}\mathbb{1}_k\mathbb{1}_k^\top)^H = \frac{1}{k^{2\|H\|_1}}$. Similarly, $\mathrm{hom}(H; G_n) = \frac{n^k}{n^{2\|H\|_1}}$, which converges to 1 if H is a disjoint union of edges and to 0 otherwise. An analogous set of arguments show that the sequence (I_n) of n self-loops is quotient-convergent.

As our next example, we consider sequences of (directed) star graphs and we show that these are quotient-convergent as well. Note that stars are sparse graphs, as their edge density converges to zero. Nevertheless, we will see in Section 5.3 that growing-sized stars have nontrivial limits in our framework, whereas the limits of stars in the context of graphons yield trivial limiting objects. Similarly, stars are not degree-bounded, as the degree of the star center increases with the size of the star, hence they do not have a limit in the framework of graphings. Stars do have limits in the graphex sense, though they differ from ours, see Section 5.1.

Example 5.2.2 (Convergence of stars)**.** *The sequence of directed star graphs all of whose edges point away from the center may be expressed as $(e_1\mathbb{1}_n^\top)$ for $e_1 = (1, 0, \ldots, 0)^\top$. This sequence quotient-converges, since random quotients $\rho(F_{k,n})G_n$ of the associated normalized graphs $G_n = \frac{1}{n}e_1\mathbb{1}_n^\top$ are given by $\rho(F_{k,n})G_n = \frac{1}{n}e_I N^\top$ for a multinomial vector $N \sim \mathrm{Multinom}(n, k, \mathbb{1}_k/k)$ and a uniformly random index $I \sim \mathrm{Unif}([k])$. Since N/n weakly converges to the deterministic vector $\mathbb{1}_k/k$, we conclude that $\rho(F_{k,n})G_n$ converges weakly to the randomly-labeled star $\frac{1}{k}e_I\mathbb{1}_k^\top$, and hence that (G_n) is quotient-convergent. Similarly, the quotient densities and homomorphism numbers of this sequence of stars converge, sinces*

$$\lim_{n\to\infty} t_Q(H; G_n) = \mathbb{E}\left(\frac{1}{k}e_I\mathbb{1}_k^\top\right)^H = \begin{cases} \frac{1}{k^{1+\sum_j m_j}} & \text{if } H = e_i m^\top \text{ for } m \in \mathbb{N}^k \text{ and } i \in [k], \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\lim_{n\to\infty} \hom(H;G_n) = \begin{cases} \lim_{n\to\infty} \frac{n^{k-1}}{n^{\sum_j m_j}} & \text{if } H = e_i m^\top, \\ 0 & \text{otherwise}, \end{cases}$$
$$= \begin{cases} 1 & \text{if } H = e_i m^\top \text{ with } \sum_j m_j = k-1, \\ 0 & \text{otherwise}. \end{cases}$$

Analogous arguments show that the sequence of directed star graphs all of whose edges point towards the center quotient-converges, and so does the sequence of undirected stars.

As our final example of quotient-convergence, we consider an infinite weighted graph with edge weights summing to one and we show that any sequence of finite subgraphs obtained from this infinite graph is quotient-convergent.

Example 5.2.3 (Subgraphs of an infinite weighted graph)**.** *Suppose $G_\infty = (G_{i,j})_{i,j\in\mathbb{N}}$ is an infinite two-dimensional array with $G_{i,j} \geq 0$ and $\sum_{i,j} G_{i,j} = 1$, and consider the sequence of normalized truncations $G_n = \frac{1}{\sum_{i,j=1}^n G_{i,j}}(G_{i,j})_{i,j\in[n]}$. This sequence of subgraphs (G_n) is quotient-convergent. Indeed, let $F_k\colon \mathbb{N} \to [k]$ be a uniformly random map (whose images $F_k(i)$ are iid uniformly distributed over $[k]$), and consider the random graph $L_k = \rho(F_k)G_\infty$ whose entries are given by (1.10). We claim that the sequence $(\rho(F_{k,n})G_n)_n$ converges weakly to L_k. To this end, view G_n as an infinite array by zero-padding it, and note that $(\rho(F_k)G_n, \rho(F_k)G_\infty)$ is a coupling of $(\rho(F_{k,n})G_n, L_k)$. Using this coupling, we obtain the bound*

$$W_1(\rho(F_{k,n})G_n, L_k) \leq \mathbb{E}\|\rho(F_k)(G_n - G_\infty)\|_1 \leq \|G_n - G_\infty\|_1 \xrightarrow{n\to\infty} 0,$$

where the second inequality follows from (1.10). Thus, we conclude that (G_n) is quotient-convergent, as claimed. Quotient densities and homomorphism numbers of this sequence again converge, with limits $\lim_n t_Q(H;G_n) = \mathbb{E}L_k^H$ and $\lim_n \hom(H;G_n) = \sum_{f\colon [k]\to\mathbb{N}} \prod_{i,j=1}^k G_{f(i),f(j)}^{H_{i,j}}$, which is finite since $(G_{i,j})$ is summable.

Finally, we remark that convex combinations of quotient-convergent sequences also converge.

Example 5.2.4 (Combinations of graphs)**.** *Suppose (G_n) and (G_n') are quotient-convergent sequences of graphs. Then for any $\theta \in [0,1]$ the convex combination $(\theta G_n + (1-\theta)G_n')$ of these graphs is quotient-convergent as well. Indeed, for each $k \in \mathbb{N}$, let L_k and L_k' be the weak limits of $\rho(F_{k,n})G_n$ and*

$\rho(F_{k,n})G'_n$, respectively. For each $n \in \mathbb{N}$, let $(\widetilde{L}_k, \widetilde{F}_{k,n})$ and $(\widetilde{L}'_k, \widetilde{F}'_{k,n})$ be couplings of $(L_k, F_{k,n})$ and $(L'_k, F_{k,n})$, respectively, attaining the Wasserstein-1 distances $W_1(L_k, \rho(F_{k,n})G_n)$ and $W_1(L'_k, \rho(F_{k,n})G_n)$. We define a coupling of $\theta L_k + (1-\theta)L'_k$ and $\rho(F_{k,n})(\theta G_n + (1-\theta)G'_n)$ as follows. Draw $F_{k,n}$ uniformly at random, and draw $\widetilde{L}_k$ from the coupling $(\widetilde{L}_k, \widetilde{F}_{k,n})$ conditioned on $\widetilde{F}_{k,n} = F_{k,n}$. Draw $\widetilde{L}'_k$ similarly and independently. Then

$$\begin{aligned}
&W_1(\theta L_k + (1-\theta)L'_k, \rho(F_{k,n})(\theta G_n + (1-\theta)G'_n)) \\
&\leq \mathbb{E}\|\theta\widetilde{L}_k + (1-\theta)\widetilde{L}'_k - \rho(F_{k,n})(\theta G_n + (1-\theta G'_n))\|_1 \\
&\leq \theta W_1(L_k, \rho(F_{k,n})G_n) + (1-\theta)W_1(L'_k, \rho(F_{k,n})G'_n) \xrightarrow{n\to\infty} 0.
\end{aligned}$$

This proves the claim. An analogous argument shows that a convex combination of any finitely-many quotient-convergent sequences of graphs is also quotient-convergent.

Having seen a few examples of quotient convergence, we proceed to prove Theorem 5.1.2. We begin by showing the first part of that theorem on the equivalence between quotient convergence and convergence of quotient densities.

Proposition 5.2.5. *A sequence of graphs $(G_n \in \Delta^{n\times n})$ is quotient-convergent if and only if $(t_Q(H; G_n))_n$ converges for each $H \in \mathbb{N}^{k\times k}$ and $k \in \mathbb{N}$.*

Proof. By definition, the sequence $(G_n \in \Delta^{n\times n})$ is quotient-convergent when the sequence$(\rho(F_{k,n})G_n \in \Delta^{k\times k})_n$ converges weakly for each $k \in \mathbb{N}$. Since $\Delta^{k\times k}$ is compact, the latter sequence converges weakly if and only if all its associated sequences of moments converge. Now observe that the monomials in $\mathbb{R}[X_{i,j}]_{i,j\in[k]}$ all have the form $X^H = \prod_{i,j} X_{i,j}^{H_{i,j}}$ for some $H \in \mathbb{N}^{k\times k}$, and that the corresponding moment of $\rho(F_{k,n})G_n$ is precisely $\mathbb{E}(\rho(F_{k,n})G_n)^H = t_Q(H; G_n)$ by definition of t_Q in (5.2). This yields the desired conclusion. □

This proves the first part of Theorem 5.1.2. To show the second part relating quotient convergence with convergence of homomorphism numbers, we express homomorphism numbers in terms of quotient densities and vice-versa. To this end, we begin by expressing quotient densities in terms of injective homomorphism numbers using [138, Prop. 6.8]. This expression is stated in terms of a partial order on multigraphs induced by quotients. Specifically, for two multigraphs $K \in \mathbb{N}^{n\times n}$, $H \in \mathbb{N}^{m\times m}$ we say that K is a *refinement* of H, denoted $K \leq_R H$, if H is a quotient of K, i.e., if there exists a surjective map $f\colon [n] \to [m]$ such that $H = \rho(f)K$. We further let $R_{K,H}$ be the number of such maps f. Note that $R_{H,H} = |\mathrm{Aut}(H)|$ for any multigraph H. The poset of all multigraphs with the refinement order is a disjoint union of posets, one for each number of edges. The poset of all multigraphs with n edges has a minimal element, namely, the disjoint union of n edges, and a maximal element, namely, a single vertex with n self-loops. We illustrate the poset of all multigraphs with two edges in Figure 5.1.

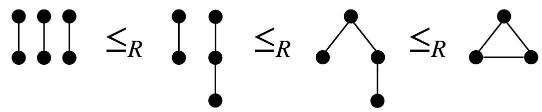


Figure 5.1: Hasse diagram for the refinement poset consisting of multigraphs with two edges. Here there is an upward path from K to H precisely when $K \leq_R H$.

The quotient density $t_Q(H; G)$ can be expressed in terms of a linear combination of injective homomorphism numbers over all refinements of H [138, Prop. 6.8]:

$$t_Q(H; G) = \sum_{K \leq_R H} |V(H)|^{-|V(K)|} \frac{H!}{K!} \frac{R_{K,H}}{R_{K,K} R_{H,H}} \mathrm{inj}(K; G), \tag{5.9}$$

where $H! = \prod_{i,j=1}^{k} (H_{i,j}!)$ and similarly for $K!$. Here $\mathrm{inj}(K; G)$ is defined similarly to $\hom(K; G)$ in (5.1) but with a sum only over injective maps. The expression (5.9) states that H is a subgraph of some quotient of G, i.e., $t_Q(H; G) > 0$, if and only if some refinement of H is a subgraph of G itself, i.e., $\mathrm{inj}(K; G) > 0$.

We are now ready to prove the second part of Theorem 5.1.2. The following argument is a simple modification of the one from [138, Prop. 6.8].

Proposition 5.2.6. *A sequence of graphs $(G_n \in \Delta^{n \times n})$ is quotient-convergent if and only if $(\mathrm{inj}(H; G_n))$ converges for each $H \in \mathbb{N}^{k \times k}$ and $k \in \mathbb{N}$, which is equivalent to $(\hom(H; G_n))$ converging for each $H \in \mathbb{N}^{k \times k}$ and $k \in \mathbb{N}$.*

Proof. For each $m \in \mathbb{N}$, define the matrix M indexed by all multigraphs containing m edges whose entries are given by $M_{H,K} = |V(H)|^{-|V(K)|} \frac{H!}{K!} \frac{R_{K,H}}{R_{K,K} R_{H,H}}$ if $K \leq_R H$ and $M_{H,K} = 0$ otherwise. This is a lower-triangular matrix with respect to the refinement partial order, and has positive diagonal entries $M_{H,H} = (|V(H)|^{|V(H)|} |\mathrm{Aut}(H)|)^{-1}$, hence it is invertible. Thus, we can write

$$\mathrm{inj}(H; G) = \sum_{H \leq_R K} (M^{-1})_{H,K} t_Q(K; G). \tag{5.10}$$

Based on (5.10) and (5.9), we conclude that $(t_Q(H; G_n))$ converges for all multigraphs H if and only if $(\mathrm{inj}(H; G_n))$ converges for all multigraphs H. Injective and non-injective homomorphism numbers can be similarly related, with one of the quantities being expressed in terms of linear combinations of the other as in [146, §5.2.3]. For example, we have [138, Eq. (73)] that

$$\hom(H;G) = \sum_{H \le_R K} \frac{R_{H,K}}{R_{K,K}} \operatorname{inj}(K;G), \tag{5.11}$$

where the sum is over multigraphs that H refines. This expression can be inverted as above to obtain injective homomorphism numbers in terms of non-injective homomorphism numbers. This allows us to conclude that convergence of $(\operatorname{inj}(H;G_n))$ for all multigraphs H is equivalent to convergence of $(\hom(H;G_n))$ for all multigraphs H. □

Combined together, Propositions 5.2.5 and 5.2.6 complete the proof of Theorem 5.1.2. We illustrate the utility of these results next by expressing quotient convergence of simple graphs in terms of appropriately-normalized homomorphism numbers. The resulting expression allows us to conclude that all dense simple graphs are quotient-convergent; in fact, we prove that all the homomorphism numbers of dense simple graphs converge to the same limit (see also Corollary 5.3.7 below).

Corollary 5.2.7. *Let $(G_n \in \mathbb{S}^n_{\mathrm{sim}})$ be a sequence of simple graphs.*

(Convergence of simple graphs) *The sequence (G_n) is quotient-convergent if and only if $\left(\frac{\hom(H;G_n)}{|E(G_n)|^{|E(H)|}}\right)$ converges for all simple and connected graphs $H \in \mathbb{S}^k_{\mathrm{sim}}$.*

(Convergence of dense graphs) *Suppose $\liminf_{n\to\infty} \frac{|E(G_n)|}{n^2} > 0$. Then (G_n) is quotient-convergent. In fact, for a simple and connected $H \in \mathbb{S}^k_{\mathrm{sim}}$, we have*

$$\lim_{n\to\infty} \frac{\hom(H;G_n)}{|E(G_n)|^{|E(H)|}} = \begin{cases} 1 & \text{if } H = K_2, \\ 0 & \text{otherwise} \end{cases}$$

where K_2 denotes a graph with a single edge.

Proof. For the first claim, if $H \in \mathbb{N}^{k\times k}$ we define a strictly upper-triangular $\widetilde{H} \in \{0,1\}^{n\times n}$ by $\widetilde{H}_{i,j} = 1$ if $i < j$ and either $H_{i,j} > 0$ or $H_{j,i} > 0$, and $\widetilde{H}_{i,j} = 0$ otherwise. Then $\hom(H;G_n) = \hom(\widetilde{H}, G_n)$ for all simple graphs G_n of all sizes. Therefore, we have

$$\hom(H; \tfrac{G_n}{\mathbb{1}^\top G_n \mathbb{1}}) = \hom(\widetilde{H}; \tfrac{G_n}{\mathbb{1}^\top G_n \mathbb{1}}) \cdot (2|E(G_n)|)^{\mathbb{1}^\top(\widetilde{H}-H)\mathbb{1}}.$$

Since $\mathbb{1}^\top \widetilde{H} \mathbb{1} \leq \mathbb{1}^\top H \mathbb{1}$, we conclude that $\hom(H; \frac{G_n}{\mathbb{1}^\top G_n \mathbb{1}})$ converges if $\hom(\widetilde{H}; \frac{G_n}{\mathbb{1}^\top G_n \mathbb{1}})$ converges, since either $|E(G_n)| \to \infty$ and $\hom(H; \frac{G_n}{\mathbb{1}^\top G_n \mathbb{1}}) \to 0$ or $|E(G_n)|$ is constant for all large n. In turn, defining the simple graph $\widetilde{H}_{\mathrm{sym}} = (\widetilde{H} + \widetilde{H}^\top)/2$, we have

$$\hom\left(\widetilde{H}; \frac{G_n}{\mathbb{1}^\top G_n \mathbb{1}}\right) = \frac{\hom(\widetilde{H}_{\mathrm{sym}}; G_n)}{(2|E(G_n)|)^{|E(\widetilde{H}_{\mathrm{sym}})|}}.$$

Thus, the sequence (G_n) quotient-converges if and only if $(\frac{\hom(H;G_n)}{|E(G_n)|^{|E(H)|}})_n$ converges for all simple H. To see that it suffices to consider simple and connected graphs, note that if $H = H_1 \sqcup \ldots \sqcup H_\ell$ is a disjoint union of connected graphs, then $\frac{\hom(H;G_n)}{|E(G_n)|^{|E(H)|}} = \prod_{i=1}^{\ell} \frac{\hom(H_i;G_n)}{|E(G_n)|^{|E(H_i)|}}$. This proves the first claim.

For the second claim, if $H = K_2$ then $\hom(H; G_n)/|E(G_n)| = 1$ by definition. Otherwise, suppose $\frac{|E(G_n)|}{n^2} \geq c > 0$ for all n. Note that $\hom(H; G_n) \leq 2|E(G_n)| \cdot n^{|V(H)|-2}$, because the number of graph homomorphisms $V(H) \to V(G_n)$ mapping all edges of H to edges of G_n is upper-bounded by the number of maps $V(H) \to V(G_n)$ that send one particular edge in H to an edge in G_n. Since H is connected and contains more than one edge, we have $|V(H)| < 2|E(H)|$, in which case

$$\frac{\hom(H; G_n)}{|E(G_n)|^{|E(H)|}} \leq \frac{2n^{|V(H)|-2}}{|E(G_n)|^{|E(H)|-1}} \leq \frac{2}{c^{|E(H)|-1}} \cdot n^{|V(H)|-2|E(H)|} \xrightarrow{n\to\infty} 0.$$

This proves the quotient-convergence of (G_n) by the first part of this corollary. $\square$

We have now seen several examples and characterizations of quotient-convergent sequences. However, so far it is not clear to what limits such sequences converge. In the next section, we derive limits for quotient-convergent graphs in the form of random exchangeable measures on $[0, 1]^2$.

5.3 Limits of Quotient-Convergent Graph Sequences

This section is devoted to characterizing limits of quotient-convergent graph sequences and understanding their properties. In Section 5.3, we prove Theorem 5.1.5 characterizing limits of quotient-convergent graph sequences via grapheurs. In Section 5.3 we interpret the different components of limiting grapheurs (5.4) in terms of combinatorial aspects of the sequence of graphs converging to them. In Section 5.3 we present a (random) duality pairing between grapheurs and graphons, and prove the formal version of Theorem 5.1.6 given in Theorem 5.3.10. Finally, Section 5.3 contains some tedious but straightforward proofs omitted from Section 5.3 for clarity of exposition.

Characterizing Limits via Grapheurs

To prove Theorem 5.1.5, we show that the random quotients of size k of a finite graph G are equal in distribution to the random graph $\mathsf{G}_k[\mathsf{M}_G]$ of size k obtained in (5.7) from the grapheur M_G associated in (5.3) to G.

Proposition 5.3.1. *For any $n, k \in \mathbb{N}$ and $G \in \Delta^{n\times n}$, we have $\rho(F_{k,n})G \overset{d}{=} \mathsf{G}_k[\mathsf{M}_G]$.*

Proof. Recall from (5.3) that $\mathsf{M}_G = \sum_{i,j=1}^n G_{i,j}\delta_{(T_i,T_j)}$ where $T_1, \ldots, T_n$ are iid uniform on $[0,1]$. With probability one, for each $i \in [n]$ there is a unique $j_i \in [k]$ such that $T_i \in I_{j_i}^{(k)}$. Moreover, the index j_i is uniformly distributed in $[k]$ by uniformity of T_i, and the indices $j_1, \ldots, j_n$ are independent of each other since the T_i are independent. Define the random map $F\colon [n] \to [k]$ sending $F(i) = j_i$ and observe that $F \overset{d}{=} F_{k,n}$ is uniformly random, since each $i \in [n]$ is mapped to an independent and uniformly random element of $[k]$. Further, note that

$$(\mathsf{G}_k[\mathsf{M}_G])_{i,j} = \mathsf{M}_G(I_i^{(k)}\times I_j^{(k)}) = \sum_{u,v=1}^n G_{u,v}\mathbb{1}[T_u \in I_i^{(k)}, T_v \in I_j^{(k)}] = \sum_{\substack{u\in F^{-1}(i)\\ v\in F^{-1}(j)}} G_{u,v} = (\rho(F)G)_{i,j}.$$

Hence, we conclude that $\mathsf{G}_k[\mathsf{M}_G] = \rho(F)G \overset{d}{=} \rho(F_{k,n})G$, as claimed. □

Consequently, we also have $t_Q(H;G) = t_Q(H;\mathsf{M}_G)$, where the latter is given by (5.8). Proposition 5.3.1 implies that a sequence (G_n) is quotient-convergent if and only if the sequence of random graphs $(\mathsf{G}_k[\mathsf{M}_{G_n}])_n$ constructed from the associated grapheurs (M_{G_n}) converge weakly for each k. Since $\Delta^{k\times k}$ is compact, this is equivalent to convergence of $(\mathsf{G}_k[\mathsf{M}_{G_n}])_n$ in the Wasserstein-1 metric W_1 with respect to the entrywise ℓ_1 norm on $\mathbb{R}^{k\times k}$ for each k (see Section 5.1). Motivated by this observation, we relate our $W_\square$-metric (5.6) to this W_1-metric.

Proposition 5.3.2. *For any two grapheurs $\mathsf{M}, \mathsf{N} \in \mathcal{M}$ and any $k \in \mathbb{N}$, we have*

$$W_1(\mathsf{G}_k[\mathsf{M}], \mathsf{G}_k[\mathsf{N}]) \leq k^2 W_\square(\mathsf{M},\mathsf{N}) \quad \textit{and} \quad W_\square(\mathsf{M},\mathsf{N}) \leq W_1(\mathsf{G}_k[\mathsf{M}], \mathsf{G}_k[\mathsf{N}]) + \frac{4}{k}.$$

The proof is straightforward but somewhat technical, as it is based on approximating any rectangle appearing in our expression (5.6) for $W_\square$ by squares in a uniform grid. We defer the proof to Section 5.3. As an immediate consequence, we conclude that a sequence of graphs (G_n) is quotient-convergent if and only if the associated sequence of grapheurs (M_{G_n}) is Cauchy in the $W_\square$-metric (5.6). We now observe that the metric space $(\mathcal{M}, W_\square)$ is compact, because the metric $W_\square$ metrizes weak convergence in $\mathcal{M}$ and $\mathcal{M}$ is weakly compact; see Section 5.1 for a review of weak convergence.

Proposition 5.3.3. *The space $\mathcal{M}$ of grapheurs (5.5) is weakly compact. In particular, any Cauchy sequence $(\mathsf{M}_n) \subseteq \mathcal{M}$ has a limit $\mathsf{M} \in \mathcal{M}$. Moreover, the following statements are equivalent for $(\mathsf{M}_n) \subseteq \mathcal{M}$ and $\mathsf{M} \in \mathcal{M}$.*

1. *We have $\lim_n \mathsf{M}_n = \mathsf{M}$ weakly.*
2. *We have that $\lim_n W_\square(\mathsf{M}_n, \mathsf{M}) = 0$.*
3. *For each $k \in \mathbb{N}$, we have that $\lim_n \mathsf{G}_k[\mathsf{M}_n] = \mathsf{G}_k[\mathsf{M}]$ weakly.*
4. *For each multigraph $H \in \mathbb{N}^{k\times k}$, we have $\lim_n t_Q(H; \mathsf{M}_n) = t_Q(H; \mathsf{M})$.*

In particular, we have $\lim_n W_\square(\mathsf{M}_{G_n}, \mathsf{M}) = 0$ for a sequence $(G_n \in \Delta^{n\times n})$ if and only if $\lim_n t_Q(H; G_n) = t_Q(H; \mathsf{M})$ for all multigraphs $H \in \mathbb{N}^{k\times k}$, as can be seen by combining Proposition 5.3.1 and Proposition 5.3.3(4). Once again, the proof is straightforward but involved, and we defer it to Section 5.3. Combining the preceding propositions, we are ready to prove Theorem 5.1.5.

Proof (Theorem 5.1.5). To prove the first statement, note that the sequence (G_n) is quotient-convergent if and only if $(\mathsf{G}_k[\mathsf{M}_{G_n}])$ is Cauchy in W_1 for each $k \in \mathbb{N}$ by Definition 5.1.1. By Proposition 5.3.2, this is equivalent to (M_{G_n}) being Cauchy in the $W_\square$-metric, as we remark previously. By Proposition 5.3.3, this is further equivalent to the existence of a limit $\mathsf{M} \in \mathcal{M}$ for (M_{G_n}) in this metric, thus proving the first statement.

The second statement follows from Proposition 5.3.3.

Finally, we prove the third statement by explicitly constructing a convergent sequence (G_n) of graphs for each grapheur $\mathsf{M} \in \mathcal{M}$. Suppose M is given by (5.4) with $E, \sigma, \varsigma, \theta, \vartheta \geq 0$ and iid uniform $(T_i)_{i\in\mathbb{N}}$. Define the sequence of graphs $(G_n \in \Delta^{n\times n})$ with entries

$$(G_n)_{i,j} = \frac{1}{s_n}\left(E_{i,j} + \tfrac{1}{n}\sigma_i + \tfrac{1}{n}\varsigma_j + \tfrac{1}{n^2}\theta + \tfrac{1}{n}\vartheta\mathbb{1}_{i=j}\right), \qquad s_n = \sum_{i,j=1}^{n} E_{i,j} + \sum_{i=1}^{n}(\sigma_i + \varsigma_i) + \theta + \vartheta, \tag{5.12}$$

where we note that $s_n \nearrow 1$ and we may assume $s_n > 0$ for all n by reordering the rows and columns of E, σ, ς if needed. Each summand in (5.12) quotient-converges to the corresponding summand in (5.4) as shown by combining Proposition 5.3.3 with the examples of Section 5.2. Moreover, by Example 5.2.4 the convex combination of these summands in (5.12) converges to the convex combination of the summands in (5.4), which is precisely the grapheur M. Thus, the third statement in Theorem 5.1.5 is proved. □

In particular, the sequence of limits of random quotients of a convergent graph sequence fully determines its limit.

Corollary 5.3.4. *Two quotient-convergent sequences* $(G_n^{(1)})$ *and* $(G_n^{(2)})$ *have the same limit if and only if* $\lim_n \rho(F_{k,n})G_n^{(1)} \overset{d}{=} \lim_n \rho(F_{k,n})G_n^{(2)}$ *for all* $k \in \mathbb{N}$.

Proof. From Proposition 5.3.3, the two sequences have the same limit M if and only if $\lim_n \rho(F_{k,n})G_n^{(i)} \overset{d}{=} \mathsf{G}_k[\mathsf{M}]$ for both $i = 1, 2$. □

We end this section by noting that valid polynomial inequalities in quotient densities precisely correspond to valid inequalities over grapheurs. We hope that grapheurs can be used to study such inequalities in future work, dually to the use of graphons in the study of inequalities in homomorphism densities [146, §16].

Corollary 5.3.5 (Inequalities in quotient densities)**.** *Suppose* $H_1, \ldots, H_k$ *are multigraphs and* $f : \mathbb{R}^k \to \mathbb{R}$ *is a continuous function. We have* $f(t_Q(H_1; G), \ldots, t_Q(H_k; G)) \geq 0$ *for all* $G \in \Delta^{n\times n}$ *and all* $n \in \mathbb{N}$ *if and only if* $f(t_Q(H_1; \mathsf{M}), \ldots, t_Q(H_k; \mathsf{M})) \geq 0$ *for all grapheurs* $\mathsf{M} \in \mathcal{M}$.

Proof. Since $t_Q(H; G) = t_Q(H; \mathsf{M}_G)$, if $f(t_Q(H_1; \mathsf{M}), \ldots, t_Q(H_k; \mathsf{M})) \geq 0$ for all grapheurs $\mathsf{M} \in \mathcal{M}$ then setting $\mathsf{M} = \mathsf{M}_G$ for $G \in \Delta^{n\times n}$ yields $f(t_Q(H_1; G), \ldots, t_Q(H_k; G)) \geq 0$. Conversely, suppose $f(t_Q(H_1; G), \ldots, t_Q(H_k; G)) \geq 0$ for all $G \in \Delta^{n\times n}$ and all $n \in \mathbb{N}$. For any $\mathsf{M} \in \mathcal{M}$, let (G_n) be a sequence of finite graphs that quotient-converges to M. Then $t_Q(H; G_n) \to t_Q(H; \mathsf{M})$ for all multigraphs H, hence by continuity of f we have $f(t_Q(H_1; \mathsf{M}), \ldots, t_Q(H_k; \mathsf{M})) = \lim_n f(t_Q(H_1; G_n), \ldots, t_Q(H_k; G_n)) \geq 0$. □

Structure of Grapheurs

Theorem 5.1.5 shows that limits of quotient-convergent sequences correspond to grapheurs in $\mathcal{M}$. Each such grapheur is described by a tuple $(E, \sigma, \varsigma, \theta, \vartheta)$ as in (5.4), and the goal of this section is to interpret each element of the tuple in terms of the combinatorics of the convergent sequence of graphs. Informally, the five parameters describing quotient limits correspond to the following properties.

- The array $E \in \mathbb{R}_{\geq 0}^{\mathbb{N}\times\mathbb{N}}$ describes a discrete collection of hubs and the fraction of edge weight between them. Specifically, each entry corresponds to an edge in a sequence of graphs that carries an asymptotically positive fraction of the total edge weight.
- The arrays $\sigma, \varsigma \in \mathbb{R}_{\geq 0}^{\mathbb{N}}$ specify directed stars. Specifically, each entry corresponds to the fraction of in-degree or out-degree, respectively, of a vertex that comes from all edges with asymptotically vanishing edge weights.
- The parameters θ and ϑ capture the remaining edge weight in the graph, coming from all edges between vertices with asymptotically vanishing degrees and their self-loops, respectively.

These properties can be seen in all the examples of Section 5.2. The following proposition proves these properties in general.

Proposition 5.3.6. *Suppose* $(G_n \subseteq \Delta^{n\times n})$ *quotient-converges to a grapheur* M *that is described by the tuple* $(E, \sigma, \varsigma, \theta, \vartheta)$ *as in* (5.4). *View each* G_n *as an element of* $\mathbb{R}^{\mathbb{N}\times\mathbb{N}}$ *by zero-padding it with infinitely-many rows and columns. Then there exist permutations* $\pi_n \in \mathsf{S}_\infty$ *of* $\mathbb{N}$ *fixing all but finitely-many integers such that* $\lim_n \|E - \pi_n G_n\|_\infty = 0$*,* $\lim_n \|E\mathbb{1} + \sigma - \pi_n G_n \mathbb{1}\|_\infty = 0$*, and* $\lim_n \|E^\top \mathbb{1} + \varsigma - \pi_n G_n^\top \mathbb{1}\|_\infty = 0$*. Furthermore, we have* $\vartheta = \lim_n \mathrm{Tr}(G_n) - \mathrm{Tr}(E)$.

Proposition 5.3.6 justifies the intuition outlined above: Up to a relabelling of the vertices of G_n, the entries of E are entrywise limits of the edge weights of G_n, the entries of σ, ς are limits of the remaining degrees of vertices in (G_n) after removing edges with positive limiting weights, and the parameters ϑ, θ capture the remaining edge weights in self loops and the rest of the edges, respectively. We prove Proposition 5.3.6 in Section 5.4 using an edge-sampling perspective on quotient convergence and an edge-based Szemerédi regularity lemma.

An immediate consequence of Proposition 5.3.6 is that all dense and degree-bounded sequences of simple graphs have the same limit.

Corollary 5.3.7. *Suppose* $(G_n \in \mathbb{S}_{\geq 0}^n)$ *is a sequence of undirected graphs such that the sequence* $(G_n/\mathbb{1}^\top G_n \mathbb{1})$ *quotient-converges to a grapheur described by* $(E, \sigma, \varsigma, \theta, \vartheta)$.

1. *We have $\sigma = \varsigma$ and $E^\top = E$.*

2. *If each G_n has no self-loops, then* $\operatorname{diag}(E) = \vartheta = 0$.

3. *If $G_n \in \mathbb{S}^n_{\mathrm{sim}}$ is a sequence of simple graphs with $|E(G_n)| \to \infty$, then $E = 0$. If (G_n) is either dense or degree-bounded, then the sequence $(\frac{G_n}{2|E(G_n)|})$ quotient converges to the grapheur λ^2.*

Proof. The first two claims follow directly from Proposition 5.3.6, since for any permutation $\pi_n \in \mathsf{S}_\infty$ the matrix $\pi_n G_n$ is, respectively, symmetric and has zero diagonal if G_n is symmetric and has no self-loops. For the third claim, since each G_n is simple, we have $G_n/\mathbb{1}^\top G_n \mathbb{1} \in \{0, 1/n\}$ which converges entrywise to zero. Thus $E = 0$ by Proposition 5.3.6. Since G_n is symmetric, we have $G_n^\top \mathbb{1} = G_n \mathbb{1} = \deg(G_n)$ is the vector of degrees of the vertices in G_n. If (G_n) is dense, then $\frac{\max_i \deg(G_n)_i}{2|E(G_n)|} \leq \frac{c}{n} \xrightarrow{n\to\infty} 0$ for some constant $c > 0$. If (G_n) is degree-bounded, with all degrees bounded by k, then $\frac{\max_i \deg(G_n)_i}{2|E(G_n)|} \leq \frac{k}{2|E(G_n)|} \xrightarrow{n\to\infty} 0$. In either case, Proposition 5.3.6 shows that $\lim_n \frac{G_n}{2|E(G_n)|} = \lambda^2$ since $\operatorname{diag}(G_n) = 0$ for all n. $\square$

We end this section by giving a sufficient condition for the limit of a quotient-convergent sequence (G_n) to consist entirely of discrete hubs, i.e., such that $\sigma = \varsigma = 0$ and $\theta = \vartheta = 0$ in (5.4). In particular, this sufficient condition implies that the random quotients $(\rho(F_{k,n})G_n)$ and the grapheurs (M_{G_n}) converge not just weakly but also in total variation. Such total variation convergence was shown for graphs sampled both from graphons [146, Cor. 10.25] and from graphexes [24, §5], and was the primary mode of convergence studied for edge-exchangeable random graph models [54]. However, this condition is too restrictive for our purposes in this chapter; for example, it does not apply to the sequence of stars in Example 5.2.2. In the following result, if μ, ν are two discrete measures on a space S then their total variation distance is $\operatorname{dist}_{\mathrm{TV}}(\mu, \nu) = \frac{1}{2}\sum_{x\in S} |\mu(\{x\}) - \nu(\{x\})|$.

Proposition 5.3.8. *Suppose $(G_n \in \Delta^{n\times n})$ and $E \in \mathbb{R}^{\mathbb{N}\times\mathbb{N}}_{\geq 0}$ satisfy $\lim_n \|\pi_n G_n - E\|_1 = 0$ for some permutations $\pi_n \in \mathsf{S}_n$. Then (G_n) is quotient-convergent to $\mathsf{M} = \sum_{i,j} E_{i,j}\delta_{(T_i,T_j)}$.*

Proof. Draw iid uniform (T_i) and consider the coupling $\mathsf{M}'_{G_n} = \sum_{i,j} (\pi_n G_n)_{i,j}\delta_{(T_i,T_j)}$, $\mathsf{M}' = \sum_{i,j} E_{i,j}\delta_{(T_i,T_j)}$, using the same locations for both. Then with probability 1 we have $\operatorname{dist}_{\mathrm{TV}}(\mathsf{M}', \mathsf{M}'_{G_n}) \leq \|\pi_n G_n - E\|_1/2 \to 0$ since the atoms (T_i, T_j) are distinct almost surely. By (5.7), we then have $W_1(\mathsf{G}_k[\mathsf{M}], \mathsf{G}_k[\mathsf{M}_{G_n}]) \leq \mathbb{E}\|\mathsf{G}_k[\mathsf{M}'] - \mathsf{G}_k[\mathsf{M}'_{G_n}]\|_1 \leq k^2 \mathbb{E}\mathrm{dist}_{\mathrm{TV}}(\mathsf{M}', \mathsf{M}'_{G_n}) \to 0$ for each k, proving quotient convergence. □

Duality between Grapheurs and Graphons

In this section, we characterize convergence of grapheurs in terms of graphons and vice-versa, exhibiting a facet of the duality between the two graph limit theories. Since graphon theory is most simply expressed in the undirected case (see [60, §8] for the directed case), we restrict ourselves to this setting in the present section.

We begin by recalling a few basic notions from graphon theory. Let $\widetilde{\mathcal{W}} = \{W : [0,1]^2 \to [0,1] \text{ measurable, symmetric}\}$ be the space of $[0,1]$-valued kernels. We define two kernels W and U to be equivalent, denoted $W \sim U$, if there exist (Lebesgue-)measure-preserving maps $\varphi, \psi : [0,1] \to [0,1]$ satisfying $W \circ (\varphi, \varphi) = U \circ (\psi, \psi)$. This is the weak isomorphism between kernels commonly defined in the dense graph limits literature [146, Cor. 10.35]. We denote by $\mathcal{W} = \widetilde{\mathcal{W}}/\sim$ the quotient space under this equivalence relation. Elements of $\mathcal{W}$ are called graphons [146, §7.1]. A graphon $W \in \mathcal{W}$ defines random simple graphs of each size by setting $\mathbb{G}_k[W] = (\mathsf{B}^{(W)}_{T_i,T_j})_{i,j\in[k]}$ where $\mathsf{B}^{(W)}_{T_i,T_j} \sim \mathrm{Ber}(W(T_i,T_j))$ are drawn independently for $i < j$ while $\mathsf{B}^{(W)}_{T_j,T_i} = \mathsf{B}^{(W)}_{T_i,T_j}$ for $i > j$ and $\mathsf{B}^{(W)}_{T_i,T_i} = 0$. Here $T_1, \ldots, T_k \overset{iid}{\sim} \mathrm{Unif}([0,1])$ as before. A sequence $(W_n) \subseteq \mathcal{W}$ converges if for each k, the sequence of random graphs $(\mathbb{G}_k[W_n])_k$ converges weakly for each k.[3] This is equivalent to convergence of homomorphism densities and to convergence in cut metric [146, §10]. Finally, a finite graph $G \in \mathbb{S}^n_{\mathrm{sim}}$ defines a step graphon $W_G(x,y) = G_{i,j}$ if $x \in I^{(n)}_i \times I^{(n)}_j$ where $I^{(n)}_i = [(i-1)/n, i/n)$.

In our context, let $\mathcal{M}_{\mathrm{sym}} = \{\mathsf{M} \in \mathcal{M} : E = E^\top, \sigma = \varsigma, \mathrm{diag}(E) = 0, \vartheta = 0, \theta \in \mathbb{R}_+ \text{ in (5.4)}\}$ be the collection of grapheurs corresponding to limits of undirected graphs with unit edge weight and no self-loops. For any $W \in \widetilde{\mathcal{W}}$ and $\mathsf{M} \in \mathcal{M}_{\mathrm{sym}}$, define the random variable

$$\langle \mathsf{M}, W\rangle = \sum_{i,j\in\mathbb{N}} E_{i,j}\mathsf{B}^{(W)}_{T_i,T_j} + 2\sum_{i\in\mathbb{N}} \sigma_i \int_0^1 W(T_i, y)\, dy + \theta \int_{[0,1]^2} W(x,y)\, dx\, dy. \tag{5.13}$$

Note that this random variable exists and is bounded by $|\langle \mathsf{M}, W\rangle| \leq 1$ because both $(E_{i,j})$ and (σ_i) are summable. Also note that $\langle \mathsf{M}, W_G\rangle = \mathbb{E}_{\mathsf{M}} W_G = \langle G, \mathsf{G}_k[\mathsf{M}]\rangle$ if

[3] Graphon convergence is usually defined by total variation convergence of these random graphs, which is equivalent to their weak convergence since they are supported on the finite set $\mathbb{S}^k_{\mathrm{sim}}$.

$G \in \mathbb{S}^k_{\text{sim}}$, since W_G takes values in $\{0,1\}$ so $\mathsf{B}^{(W_G)}_{T_i,T_j} = W_G(T_i,T_j)$. Moreover, for finite graphs $G_1 \in \Delta^{n\times n}$, $G_2 \in \mathbb{S}^m_{\text{sim}}$, we have

$$\langle \mathsf{M}_{G_1}, W_{G_2}\rangle = \langle \rho(F_{m,n})G_1, G_2\rangle = \langle G_1, \rho(F_{m,n})^\star G_2\rangle,$$

which can equivalently be viewed as comparing a random quotient of G_1 to G_2 or comparing a random subgraph of G_2 to G_1 (see Section 3.1). We now prove that the random variable (5.13) only depends on the equivalence class of a graphon W.

Proposition 5.3.9. *For any $W, U \in \widetilde{\mathcal{W}}$, if $W \sim U$ then $\langle \mathsf{M}, W\rangle \overset{d}{=} \langle \mathsf{M}, U\rangle$ for any $\mathsf{M} \in \mathcal{M}_{\text{sym}}$.*

Proof. Suppose $\mathsf{M} \in \mathcal{M}_{\text{sym}}$ is defined by $(E, \sigma = \varsigma, \theta, \vartheta = 0)$ as in (5.4). Since $W \sim U$, there are measure-preserving maps $\varphi, \psi\colon [0,1] \to [0,1]$ satisfying $W \circ (\varphi,\varphi) = U \circ (\psi,\psi)$. We now define a coupling of $\langle \mathsf{M}, W\rangle$ and $\langle \mathsf{M}, U\rangle$ under which the two are equal almost surely. Sample locations (T_i) iid uniformly on $[0,1]$, and note that both $(\varphi(T_i)), (\psi(T_i)) \overset{iid}{\sim} \mathrm{Unif}([0,1])$ because φ and ψ are measure-preserving. Note that, conditioned on the locations (T_i) above, we have $\mathsf{B}^{(W)}_{\varphi(T_i),\varphi(T_j)} \overset{d}{=} \mathsf{B}^{(U)}_{\psi(T_i),\psi(T_j)}$, since they are both Bernoulli random variables with success probability $W(\varphi(T_i),\varphi(T_j)) = U(\psi(T_i),\psi(T_j))$. Letting $\mathsf{B}_{i,j}$ be a sample from the above distribution independently for each $i < j$, setting $\mathsf{B}_{i,j} = \mathsf{B}_{j,i}$ for $i > j$, and $\mathsf{B}_{i,i} = 0$ for all i, we define the coupling of $\langle \mathsf{M}, W\rangle$ and $\langle \mathsf{M}, U\rangle$ by using (5.13) with these common samples $(\mathsf{B}_{i,j})$ for both, and note that under this coupling they are equal almost surely. Thus, we have $\langle \mathsf{M}, W\rangle \overset{d}{=} \langle \mathsf{M}, U\rangle$ as desired. $\square$

Thus, the pairing (5.13) defines a pairing from $\mathcal{M}_{\text{sym}} \times \mathcal{W}$ to bounded scalar random variables. Importantly, we can characterize convergence in $\mathcal{M}_{\text{sym}}$ in terms of graphons and characterize graphon convergence in $\mathcal{W}$ in terms of grapheurs. The following is the formal version of Theorem 5.1.6 from Section 3.1.

Theorem 5.3.10. *We have $\lim_n \mathsf{M}_n = \mathsf{M}$ in $\mathcal{M}_{\text{sym}}$ if and only if $\langle \mathsf{M}_n, W_G\rangle \to \langle \mathsf{M}, W_G\rangle$ weakly for all simple graphs G. We also have $\lim_n W_n \to W$ in $\mathcal{W}$ if and only if $\langle \mathsf{M}_G, W_n\rangle \to \langle \mathsf{M}_G, W\rangle$ weakly for all undirected graphs G with unit edge weight and no self-loops.*

Proof. For the first claim, suppose $\langle \mathsf{M}_n, W_G\rangle \to \langle \mathsf{M}, W_G\rangle$ for all $G \in \mathbb{S}^k_{\text{sim}}$ and all $k \in \mathbb{N}$. Therefore, $\mathbb{E}_{\mathsf{M}_n} W_G \to \mathbb{E}_{\mathsf{M}} W_G$ for any such G. Taking linear combinations of such W_G, we get $\mathbb{E}_{\mathsf{M}_n} W \to \mathbb{E}_{\mathsf{M}} W$ for any $W\colon [0,1]^2 \to \mathbb{R}$ that is piecewise-constant on the $k \times k$ uniform grid and equals zero on the diagonal for some $k \in \mathbb{N}$. For any continuous and symmetric $f\colon [0,1]^2 \to \mathbb{R}$, define $f_k = \sum_{i,j=1}^{k} f_{i,j}^{(k)} \mathbb{1}_{I_i^{(k)} \times I_j^{(k)}}$ where if $i \neq j$ we set $f_{i,j}^{(k)} = f(\frac{i}{k} - \frac{1}{2k}, \frac{j}{k} - \frac{1}{2k})$ to the value of f at the center of the rectangle $I_i^{(k)} \times I_j^{(k)}$, and if $i = j$ we set $f_{i,i}^{(k)} = 0$. Since f is continuous on a compact set, it is uniformly continuous, so for any $\epsilon > 0$ we have $|f(x) - f_k(x)| \leq \epsilon$ for all $x \in [0,1]^2 \setminus \bigcup_{i=1}^{k} (I_i^{(k)} \times I_i^{(k)})$. Therefore, for any grapheur $\mathsf{M}' \in \mathcal{M}_{\mathrm{sym}}$ we have

$$\mathbb{E}_{\mathsf{M}'}|f - f_k| \leq \epsilon + \mathbb{E}\mathsf{M}'\left(\bigcup_{i=1}^{k} I_i^{(k)} \times I_i^{(k)}\right) = \epsilon + \frac{1}{k},$$

for all large k since $\mathbb{E}\mathsf{M}' = \lambda^2$ for $\mathsf{M}' \in \mathcal{M}_{\mathrm{sym}}$. Thus, we have

$$W_1(\mathbb{E}_{\mathsf{M}_n} f, \mathbb{E}_{\mathsf{M}} f) \leq 2(\epsilon + k^{-1}) + W_1(\mathbb{E}_{\mathsf{M}_n} f_k, \mathbb{E}_{\mathsf{M}} f_k) \xrightarrow{n \to \infty} 2(\epsilon + k^{-1}),$$

since f_k is piecewise-constant on the $k \times k$ uniform grid and has zero values on its diagonal. We conclude by taking $k \to \infty$ that $\mathbb{E}_{\mathsf{M}_n} f \to \mathbb{E}_{\mathsf{M}} f$ weakly for any continuous $f\colon [0,1]^2 \to \mathbb{R}$, and hence that $\mathsf{M}_n \to \mathsf{M}$ weakly. Conversely, suppose $\mathsf{M}_n \to \mathsf{M}$ weakly, and let $W = W_G$ for $G \in \mathbb{S}_{\mathrm{sim}}^{k}$. Then

$$\langle \mathsf{M}_n, W_G \rangle = \mathbb{E}_{\mathsf{M}_n} W_G = \langle G, \mathsf{G}_k[\mathsf{M}_n] \rangle \to \langle G, \mathsf{G}_k[\mathsf{M}] \rangle = \mathbb{E}_{\mathsf{M}} W_G = \langle \mathsf{M}, W_G \rangle,$$

weakly, since $\mathsf{G}_k[\mathsf{M}_n] \to \mathsf{G}_k[\mathsf{M}]$ weakly for each $k \in \mathbb{N}$. This proves the first claim.

For the second claim, suppose $\langle \mathsf{M}_G, W_n \rangle \to \langle \mathsf{M}_G, W \rangle$ weakly for all graphs $G \in \mathbb{S}^k \cap \Delta^{k \times k}$ with $\mathrm{diag}(G) = 0$, i.e., undirected weighted graphs with no self-loops and total edge weight of 1. We then have

$$\langle G, \mathbb{G}_k[W_n] \rangle = \langle \mathsf{M}_G, W_n \rangle \to \langle \mathsf{M}_G, W \rangle = \langle G, \mathbb{G}_k[W] \rangle,$$

weakly. Therefore, we get $\langle G, \mathbb{G}_k[W_n] \rangle \to \langle G, \mathbb{G}_k[W] \rangle$ for all $G \in \mathbb{S}^k$ by linearity and the fact that $\mathbb{G}_k[W_n]$ all have zero diagonal. This implies $\mathbb{G}_k[W_n] \to \mathbb{G}_k[W]$ weakly by Cramér–Wold [114, Cor. 6.5]. Since this holds for all $k \in \mathbb{N}$, we conclude that $W_n \to W$. Conversely, suppose $W_n \to W$. Then $\langle \mathsf{M}_G, W_n \rangle = \langle G, \mathsf{G}_k[W_n] \rangle \to \langle G, \mathsf{G}_k[W] \rangle = \langle \mathsf{M}_G, W \rangle$ for any $G \in \mathbb{S}^k \cap \Delta^{k \times k}$. □

We conjecture that the above duality pairing is, in fact, bi-continuous.

Conjecture 5.3.11. *If* $(\mathsf{M}_n) \subseteq \mathcal{M}_{\mathrm{sym}}$ *and* $(W_n) \subseteq \mathcal{W}$ *converge to* M *and* W*, respectively, then* $\langle \mathsf{M}_n, W_n \rangle \to \langle \mathsf{M}, W \rangle$ *weakly.*

In particular, this conjecture implies that $\mathsf{M}_n \to \mathsf{M}$ if and only if $\langle \mathsf{M}_n, W\rangle \to \langle \mathsf{M}, W\rangle$ for all graphons $W \in \mathcal{W}$, and $W_n \to W$ if and only if $\langle \mathsf{M}, W_n\rangle \to \langle \mathsf{M}, W\rangle$ for all grapheurs $\mathsf{M} \in \mathcal{M}_{\mathrm{sym}}$, a strengthening of Theorem 5.3.10.

Theorem 5.3.10 exhibits a concrete, formal duality between quotient convergence and graphon convergence. This duality can be extended to general grapheurs and more general kernels on $[0,1]^2$, but the extension of graphon convergence to this setting is more complicated [60, §8]. We therefore leave this and other extensions of this duality for future work.

Missing Proofs from Section 5.3

In this section, we give all the proofs deferred from Section 5.3 and Section 5.3, namely, Propositions 5.3.2 and 5.3.3.

Proof of Proposition 5.3.2

Many of the proofs in this section are based on standard approximations of rectangles, and continuous functions by step functions on a uniform grid and vice-versa. Such approximations allow us to compare the measures of these objects under our random exchangeable measures using the following lemma.

Lemma 5.3.12. *Fix any grapheur* $\mathsf{M} \in \mathcal{M}$.

1. *There is* $\theta \in [0,1]$ *satisfying* $\mathbb{E}\mathsf{M} = \theta\lambda^2 + (1-\theta)\lambda_D$.
2. *For any Borel set* $B \subseteq [0,1]^2$*, we have* $\mathbb{E}\mathsf{M}(B) \leq \max\{\lambda^2(B), \lambda_D(B)\}$*. In particular, if* $\lambda^2(B) = \lambda_D(B) = 0$*, then* $\mathsf{M}(B) = 0$ *almost surely.*

Proof. For the first claim, since M is exchangeable we conclude that $\mu = \mathbb{E}\mathsf{M}$ is a (deterministic) exchangeable measure on $[0,1]^2$, meaning that $\mu \circ (\sigma, \sigma) = \mu$ for all (Lebesgue)-measure preserving bijections $\sigma\colon [0,1] \to [0,1]$. The space of such measures is two-dimensional and is spanned by λ^2, λ_D by [115, Thm. 9.12], so $\mu = \theta\lambda + \vartheta\lambda_D$ for some $\theta, \vartheta \in \mathbb{R}$. Since M is a probability measure almost surely, its mean μ is a probability measure as well, necessitating $\theta + \vartheta = 1$ and $\theta, \vartheta \geq 0$. This proves the first claim.

The second claim immediately follows from the first, noting in particular that if $\lambda^2(B) = \lambda_D(B) = 0$ then the scalar random variable $\mathsf{M}(B)$ is nonnegative with a vanishing mean. $\square$

To prove Proposition 5.3.2, we also need to relate couplings of random measures to couplings of random graphs derived from them.

Lemma 5.3.13. *For any* $\mathsf{M}_1, \mathsf{M}_2 \in \mathcal{M}$ *and coupling* $(\mathsf{G}_1, \mathsf{G}_2)$ *of* $(\mathsf{G}_k[\mathsf{M}_1], \mathsf{G}_k[\mathsf{M}_2])$ *for some* $k \in \mathbb{N}$*, there is a coupling* $(\mathsf{M}_1', \mathsf{M}_2')$ *of* $(\mathsf{M}_1, \mathsf{M}_2)$ *such that* $(\mathsf{G}_k[\mathsf{M}_1'], \mathsf{G}_k[\mathsf{M}_2']) \overset{d}{=} (\mathsf{G}_1, \mathsf{G}_2)$.

Proof. We apply the disintegration theorem [178, Thm. A] to the continuous map $\pi_k\colon \mathcal{P}([0,1]^2) \to \Delta^{k\times k}$ between compact metric spaces sending $\mu \mapsto (\mu(I_i^{(k)} \times I_j^{(k)}))_{i,j\in[k]}$. The disintegration theorem yields for each $\mathsf{M} \in \mathcal{M}$ a Borel collection of probability measures $(\lambda_G^{(\mathsf{M})})_{G\in\Delta^{k\times k}}$ for each $\mathsf{M} \in \mathcal{M}$ such that $\operatorname{supp}(\lambda_G^{(\mathsf{M})}) \subseteq \pi_k^{-1}(G)$ and such that $\operatorname{Law}(\mathsf{M}) = \mu_k^{(\mathsf{M})} \otimes \lambda_G^{(\mathsf{M})}$ in the sense that

$$\mathbb{P}[\mathsf{M} \in B] = \int_{\Delta^{k\times k}} \int_{\pi_k^{-1}(G)} \mathbb{1}_B(\mu)\, d\lambda_G^{(\mathsf{M})}(\mu)\, d\mu_k^{(\mathsf{M})}(s),$$

for any Borel $B \subseteq \mathcal{P}([0,1]^2)$. The measure $\lambda_G^{(\mathsf{M})}$ is the distribution of M conditioned on $\pi_k(\mathsf{M}) = G$.

Let $\gamma = \operatorname{Law}((G_1, G_2))$. We construct the coupling $\Gamma = \operatorname{Law}((\mathsf{M}_1', \mathsf{M}_2'))$ explicitly by setting

$$\Gamma(B) = \int_{\Delta^{k\times k}\times\Delta^{k\times k}} \int_{\pi_k^{-1}(G_2)} \int_{\pi_k^{-1}(G_1)} \mathbb{1}_B(\mu, \nu)\, d\lambda_{G_1}^{(\mathsf{M})}(\mu)\, d\lambda_{G_2}^{(\mathsf{N})}(\nu)\, d\gamma(G_1, G_2).$$

for each Borel $B \subseteq \mathcal{P}([0,1]^2) \times \mathcal{P}([0,1]^2)$. This distribution corresponds to first sampling $(\mathsf{G}_1, \mathsf{G}_2) \sim \gamma$ and then sampling M_1' and M_2' conditioned on $\mathsf{G}_k[\mathsf{M}_1'] = \mathsf{G}_1$ and $\mathsf{G}_k[\mathsf{M}_2'] = \mathsf{G}_2$. It is routine to check that this is a coupling of $(\mathsf{M}_1, \mathsf{M}_2)$ and that its pushforward under (π_k, π_k) is the given coupling γ of $(\mathsf{G}_1, \mathsf{G}_2)$, which is the claim we sought to prove. □

We can now prove Proposition 5.3.2.

Proof (Proposition 5.3.2). Define the intervals $I_i^{(n)} = [(i-1)/n, i/n)$ for $i \in [n]$ and let $(\mathsf{M}', \mathsf{N}')$ be a coupling of (M, N) attaining $W_\square(\mathsf{M}, \mathsf{N})$ in (5.6). Such a coupling exists since the infimum in (5.6) consists of minimizing a lower-semicontinuous function of the coupling over the compact space of all possible couplings. For each $k \in \mathbb{N}$, we have

$$\begin{aligned} W_1(\mathsf{G}_k[\mathsf{M}], \mathsf{G}_k[\mathsf{N}]) &\le \sum_{i,j=1}^{k} \mathbb{E}|\mathsf{M}'(I_i^{(k)} \times I_j^{(k)}) - \mathsf{N}'(I_i^{(k)} \times I_j^{(k)})| \\ &\le k^2 W_\square(\mathsf{M}, \mathsf{N}), \end{aligned}$$

giving the first claimed inequality.

For the second inequality, consider arbitrary intervals $S = [a_1, a_2]$ and $T = [a_3, a_4]$ for $a, b, c, d \in [0, 1]$. For each $k \in \mathbb{N}$ and $j \in [4]$, let $i_j \in [k]$ be the unique index such that $a_j \in I_{i_j}^{(k)}$ (if $a_j = 1$ set $i_j = k$). Note that $S \subseteq \bigcup_{i=i_1}^{i_2} I_i^{(k)}$ and $|\bigcup_{i=i_1}^{i_2} I_i^{(k)} \setminus S| \leq 2/k$, and similarly for T. Therefore, we have

$$\lambda^2\left(\bigcup_{i=i_1}^{i_2}\bigcup_{j=i_3}^{i_4} I_i^{(k)} \times I_j^{(k)} \setminus S \times T\right) \leq \frac{4}{k}, \quad \lambda_D\left(\bigcup_{i=i_1}^{i_2}\bigcup_{j=i_3}^{i_4} I_i^{(k)} \times I_j^{(k)} \setminus S \times T\right) \leq \frac{2\sqrt{2}}{k},$$

hence by Lemma 5.3.12 we have

$$\mathbb{E}|\mathsf{M}'(S \times T) - \mathsf{N}'(S \times T)| \leq \mathbb{E}\|\mathsf{G}_k[\mathsf{M}'] - \mathsf{G}_k[\mathsf{N}']\|_1 + \frac{4}{k}.$$

Taking the supremum over S, T, we get

$$W_\square(\mathsf{M}', \mathsf{N}') \leq \frac{4}{k} + \inf_{\substack{\text{couplings}\\(\mathsf{M}',\mathsf{N}')}} \mathbb{E}\|\mathsf{G}_k[\mathsf{M}'] - \mathsf{G}_k[\mathsf{N}']\|_1 = \frac{4}{k} + W_1(\mathsf{G}_k[\mathsf{M}], \mathsf{G}_k[\mathsf{N}]),$$

where the last equality follows from Lemma 5.3.13. This is the second claimed inequality. $\square$

Proof of Proposition 5.3.3

Next, we turn to proving Proposition 5.3.3, which states that $\mathcal{M}$ is weakly compact and that the metric $W_\square$ metrizes this weak topology. To prove that $\mathcal{M}$ is weakly compact, we rely on the following lemma and the characterization of $\mathcal{M}$ from [115, §9] as the space of ergodic random exhangeable measures.

Lemma 5.3.14. *If (μ_n) is a weakly convergent sequence of probability distributions on a Polish space S with limit μ, and if $B \subseteq S$ is a measurable subset satisfying $\mu_n(B) = 1$ for all n, then $\mu(B) = 1$.*

Proof. By Urysohn's lemma, for any $\epsilon > 0$ we can find a subset $B_\epsilon \supseteq B$ with $\mu(B_\epsilon \setminus B) \leq \epsilon$ and a continuous function $f \colon S \to [0, 1]$ satisfying $f|_B = 1$ and $f|_{S \setminus B_\epsilon} = 0$. Since f is continuous, we have $\mathbb{E}_\mu f = \lim_n \mathbb{E}_{\mu_n} f \geq \lim_n \mu_n(S) = 1$. On the other hand, since $f \leq \mathbb{1}_{B_\epsilon \setminus B} + \mathbb{1}_B$ pointwise, we have $\mathbb{E}_\mu f \leq \mu(B) + \epsilon$. We conclude that $\mu(B) \geq 1$, and since μ is a probability measure, that $\mu(B) = 1$, as desired. $\square$

We are now to prove Proposition 5.3.3.

Proof (Proposition 5.3.3). We break down the proof into several steps.

Compactness of $\mathcal{M}$: The space $\mathcal{P}([0,1]^2)$ of probability measures on $[0,1]^2$ is compact in the weak topology. Therefore, the space $\mathcal{P}(\mathcal{P}([0,1]^2))$ of random such measures is again compact in its own weak topology (see Section 5.1). Thus, it suffices to prove that $\mathcal{M} \subseteq \mathcal{P}(\mathcal{P}([0,1]^2))$ is closed in this weak topology.

By [115, Prop. 9.1], the space of random exchangeable measures is precisely the collection of random measures on $[0,1]^2$ invariant under the countable group $\widetilde{\mathsf{S}}_\infty$ consisting of permutations of the intervals $(I_i^{(n)})_{i\in[n]}$ for all $n \in \mathbb{N}$. By [115, Lemma A1.2], we can characterize its extreme points by

$$\mathcal{M} = \left\{ \mathsf{M} \text{ random exchangeable} : \mathbb{P}[\mathsf{M} \in B] \in \{0,1\} \text{ whenever } B \subseteq \mathcal{P}([0,1]^2) \text{ is } \widetilde{\mathsf{S}}_\infty\text{-invariant} \right\}.$$

If $\mathsf{M}_n \to \mathsf{M}$ and $\mathrm{Law}(\mathsf{M}_n)(B) \in \{0,1\}$ for all n, then passing to a subsequence we may assume $\mathrm{Law}(\mathsf{M}_n)(B) = 1$ for all n or $\mathrm{Law}(\mathsf{M}_n)(B) = 0$ for all n. Applying Lemma 5.3.14 to either B or its complement, respectively, yields $\mathrm{Law}(\mathsf{M})(B) \in \{0,1\}$. We thus conclude that $\mathcal{M}$ is weakly closed, and hence compact.

Equivalence of statements 2 and 3: The equivalence of statements 2 and 3 follows from Proposition 5.3.2, since $\lim_n \mathsf{G}_k[\mathsf{M}_n] = \mathsf{G}_k[\mathsf{M}]$ weakly for a given $k \in \mathbb{N}$ if and only if $\lim_n W_1(\mathsf{G}_k[\mathsf{M}_n], \mathsf{G}_k[\mathsf{M}]) = 0$.

Equivalence of statements 3 and 4: The equivalence of these two statements follows by the proof of Proposition 5.2.5. Specifically, the sequence $(\mathsf{G}_k[\mathsf{M}_n])$ converges weakly to $\mathsf{G}_k[\mathsf{M}]$ if and only if the moments of $(\mathsf{G}_k[\mathsf{M}_n])$ converge to those of $\mathsf{G}_k[\mathsf{M}]$, and these moments are precisely the quotient densities of grapheurs defined in (5.8).

Equivalence of statements 1 and 2: Endow $\mathbb{R}^2$ with the ℓ_∞ norm, endow $\mathcal{P}([0,1]^2)$ with the Wasserstein-1 metric with respect to this norm, and endow $\mathcal{M}$ with the Wasserstein-1 metric with respect to the Wasserstein-1 metric on $\mathcal{P}([0,1]^2)$, which metrizes weak convergence of random measures [220, Thm. 6.9]. Explicitly, this metric is given by

$$\mathrm{dist}_{W_1}(\mathsf{M}_1, \mathsf{M}_2) = \min_{\substack{\text{random } (\mathsf{M}_1', \mathsf{M}_2') \\ \mathsf{M}_i' \overset{d}{=} \mathsf{M}_i \text{ for } i=1,2}} \mathbb{E}_{(\mathsf{M}_1', \mathsf{M}_2')} \sup_{\substack{f\colon [0,1]^2 \to \mathbb{R} \\ f(0)=0 \\ \text{1-Lipschitz in } \ell_\infty}} |\mathbb{E}_{\mathsf{M}_1'} f - \mathbb{E}_{\mathsf{M}_2'} f|,$$

where the min over couplings is attained [220, Thm. 4.1]. It therefore suffices to prove that $\lim_n W_\square(\mathsf{M}_n, \mathsf{M}) = 0$ if and only if $\lim_n \mathrm{dist}_{W_1}(\mathsf{M}_n, \mathsf{M}) = 0$. We proceed to prove this by standard approximations of continuous functions by step functions and vice-versa.

Suppose that $\lim_n \mathrm{dist}_{W_1}(\mathsf{M}_n, \mathsf{M}) = 0$ and fix any $\epsilon > 0$. Let $(\mathsf{M}'_n, \mathsf{M}')$ be a coupling of $(\mathsf{M}_n, \mathsf{M})$ attaining dist_{W_1}. For any interval $S \subseteq [0,1]$, define $S_\epsilon = (S + [-\epsilon, \epsilon]) \cap [0,1]$. Note that

$$\lambda^2(S_\epsilon \times T_\epsilon \setminus S \times T) \leq 4\epsilon, \qquad \lambda_D(S_\epsilon \times T_\epsilon \setminus S \times T) \leq 2\sqrt{2}\epsilon.$$

We can find an ϵ^{-1}-Lipschitz $f_\epsilon \colon [0,1]^2 \to [0,1]$ satisfying $f|_{S\times T} = 1$ and $f|_{[0,1]^2 \setminus S_\epsilon \times T_\epsilon} = 0$. For example, set $f_\epsilon(x) = \frac{\mathrm{dist}(x, S_\epsilon \times T_\epsilon)}{\mathrm{dist}(x, S\times T) + \mathrm{dist}(x, S_\epsilon \times T_\epsilon)}$ where dist is with respect to the ℓ_∞ norm. For any $\mathsf{M} \in \mathcal{M}$, we have

$$\mathbb{E}|\mathsf{M}(S\times T) - \mathbb{E}_{\mathsf{M}} f_\epsilon| \leq \mathbb{E}\mathsf{M}(S_\epsilon \times T_\epsilon \setminus S \times T) \leq 4\epsilon,$$

by Lemma 5.3.12, hence

$$|\mathbb{E}\mathsf{M}'_n(S\times T) - \mathbb{E}\mathsf{M}'(S\times T)| \leq 8\epsilon + |\mathbb{E}_{\mathsf{M}'_n} f_\epsilon - \mathbb{E}_{\mathsf{M}'} f_\epsilon| \leq 8\epsilon + \frac{1}{\epsilon}\mathrm{dist}_{W_1}(\mathsf{M}_n, \mathsf{M}).$$

We conclude that $W_\square(\mathsf{M}_n, \mathsf{M}) \leq 8\epsilon + \epsilon^{-1}\mathrm{dist}_{W_1}(\mathsf{M}_n, \mathsf{M})$, so taking $n \to \infty$ and then $\epsilon \to 0$ we conclude that $\lim_n W_\square(\mathsf{M}_n, \mathsf{M}) = 0$.

Conversely, suppose $\lim_n W_\square(\mathsf{M}_n, \mathsf{M}) = 0$. Once again, we note that the infimum over couplings in (5.6) is attained, since it consists of minimizing a lower-semicontinuous function of the coupling over the compact space of all possible couplings. For each n, let $(\mathsf{M}', \mathsf{M}'_n)$ be a coupling attaining $W_\square(\mathsf{M}_n, \mathsf{M})$. For each $N \in \mathbb{N}$, define the intervals $I_i^{(N)} = [(i-1)/N, i/N)$ for $i \in \{1, \ldots, N-1\}$ and $I_N^{(N)} = [1 - 1/N, 1]$. For any 1-Lipschitz $f \colon [0,1] \to \mathbb{R}$ with $f(0) = 0$, let

$$f_N(x) = \sum_{i,j=1}^{N} f_{i,j}^{(N)} \mathbb{1}_{I_i^{(N)} \times I_j^{(N)}}(x),$$

where $f_{i,j}^{(N)} = f(\frac{i}{N} - \frac{1}{2N}, \frac{j}{N} - \frac{1}{2N})$ is the value of f on the center of the squares $I_i^{(N)} \times I_j^{(N)}$. Since f is 1-Lipschitz in ℓ_∞ and the diameter of these squares is $1/2N$ in ℓ_∞, we have $|f(x) - f_N(x)| \leq \frac{1}{2N}$ for all $x \in [0,1]^2$. Furthermore, since $f(0) = 0$ we have $|f_{i,j}^{(N)}| \leq 1$ for all i, j. Therefore, for any $\mathsf{M} \in \mathcal{M}$ we have

$$|\mathbb{E}_{\mathsf{M}} f - \mathbb{E}_{\mathsf{M}} f_N| \leq \frac{1}{2N},$$

and for our particular coupling $(\mathsf{M}', \mathsf{M}'_n)$, we have

$$|\mathbb{E}_{\mathsf{M}'} f - \mathbb{E}_{\mathsf{M}'_n} f| \leq \frac{1}{N} + |\mathbb{E}_{\mathsf{M}'} f_N - \mathbb{E}_{\mathsf{M}'_n} f_N| \leq \frac{1}{N} + \sum_{i,j=1}^{N} |\mathsf{M}'(I_i^{(N)} \times I_j^{(N)}) - \mathsf{M}'_n(I_i^{(N)} \times I_j^{(N)})|.$$

Since this inequality holds for any such function f, we conclude that

$$\mathrm{dist}_{W_1}(\mathsf{M}_n, \mathsf{M}) \leq \frac{1}{N} + \sum_{i,j=1}^{N} \mathbb{E}|\mathsf{M}'(I_i^{(N)} \times I_j^{(N)}) - \mathsf{M}'_n(I_i^{(N)} \times I_j^{(N)})| \leq \frac{1}{N} + N^2 W_\square(\mathsf{M}_n, \mathsf{M}).$$

Letting $n \to \infty$ and then $N \to \infty$, we conclude that $\lim_n \mathrm{dist}_{W_1}(\mathsf{M}_n, \mathsf{M}) = 0$. □

5.4 Edge Sampling and Property Testing

In this section, we consider the problem of testing properties of large graphs by randomly sampling small summaries using our framework. For properties specified by graph parameters that are continuous with respect to quotient convergence, it would seem that random quotients are natural candidates for obtaining small summaries of large graphs. However, random quotients require access to the entire graph and are therefore impractical for large graphs. Moreover, we will see in Section 5.5 that random quotients do not concentrate very well, further obstructing property testing based on these summaries.

To remedy this situation, we present in this section an edge-based sampling procedure that yields a practical method for obtaining small summaries approximating large graphs of any size in the $W_\square$-metric with universal rates. This sampling procedure is based on an equivalent view of quotient convergence in terms of convergence of edge-exchangeable random graph models, the equivalence being furnished by the classical de Finetti's theorem. We describe this perspective on quotient convergence in Section 5.4 and an edge-based analog of the Szemerédi regularity lemma in Section 5.4. We leverage these results to present a practical approach for property testing, which we formally define and study in Section 5.4. We conclude in Section 5.4 by presenting proofs from Section 5.4 that are omitted from notational clarity as well as the proof of Proposition 5.3.6 from Section 5.3.

Edge Sampling View of Quotient Convergence

We begin by describing a procedure to obtain finite random graphs from a grapheur by sampling edges with replacement. We then reinterpret quotient convergence in terms of convergence of the resulting random edge-exchangeable graphs.

Given a grapheur M, we can obtain a sequence of weighted random graphs $(\mathsf{G}^{(n)}[\mathsf{M}])_n$ where $\mathsf{G}^{(n)}[\mathsf{M}]$ has n edges as follows. We sample a measure $\mu \sim \mathsf{M}$ on $[0,1]^2$, followed by sampling n edges $(x_1, y_1), \ldots, (x_n, y_n) \overset{iid}{\sim} \mu$. Next we construct the weighted graph $\mathsf{G}^{(n)}[\mathsf{M}]$ on $N = |\{x_i\} \cup \{y_i\}|$ vertices by arbitrarily enumerating

$\{t_1, \ldots, t_N\} = \{x_i\} \cup \{y_i\}$ and setting the edge weights to:

$$(\mathsf{G}^{(n)}[\mathsf{M}])_{i,j} = |\{k \in [n] : (x_k, y_k) = (t_i, t_j)\}|/n$$

for $1 \le i, j \le N$. Note that $|V(\mathsf{G}^{(n)}[\mathsf{M}])| \le 2n$.

If the grapheur M is the limit of undirected graphs, we can further replace $\mathsf{G}^{(n)}[\mathsf{M}]$ by its symmetrization $\frac{\mathsf{G}^{(n)}[\mathsf{M}]+(\mathsf{G}^{(n)}[\mathsf{M}])^\top}{2}$ to obtain an undirected random graph; see Remark 5.4.6 below. When $\mathsf{M} = \mathsf{M}_G$ is the grapheur corresponding to a finite graph G as in (5.3), the above sampling procedure amounts to sampling with replacement the edges of G proportionally to their weight, and randomly labeling the resulting graph.

For a general grapheur M described by parameters $(E, \sigma, \varsigma, \theta, \vartheta)$, edge sampling from M as above corresponds to first sampling random locations $(T_i)_i$ and then sampling edges as follows.

- With probability $E_{i,j}$, sample the edge (T_i, T_j).
- With probability σ_i, we sample a new vertex $T' \sim \mathrm{Unif}([0,1])$ and sample the edge (T_i, T'). With probability ς_i, we analogously sample (T', T_i).
- With probability θ we sample two new vertices $T_1', T_2' \sim \mathrm{Unif}([0,1])$ and add the edge (T_1', T_2'). Similarly, with probability ϑ we add the edge (T'', T'') for a new location $T'' \sim \mathrm{Unif}([0,1])$.

After repeating the above procedure n times, we get a random exchangeable multi-graph containing n edges (possibly with repetition), and normalize the resulting graph to have total edge weight one.

We now observe that convergence of edge-sampled graphs is equivalent to quotient convergence.

Proposition 5.4.1. *We have* $\lim_n \mathsf{M}_n = \mathsf{M}$ *in* $\mathcal{M}$ *if and only if* $(\mathsf{G}^{(k)}[\mathsf{M}_n])_n \subseteq \Delta^{2k\times 2k}$ *converges weakly to* $\mathsf{G}^{(k)}[\mathsf{M}]$ *for all* $k \in \mathbb{N}$.

Proof. The grapheurs (M_n) converge weakly if and only if the random exchangeable arrays

$$\begin{bmatrix} x_1^{(n)} & \cdots & x_k^{(n)} \\ y_1^{(n)} & \cdots & y_k^{(n)} \end{bmatrix} \in [0,1]^{2\times k}, \tag{5.14}$$

obtained by first sampling $\mu_n \sim \mathsf{M}_n$ and then sampling $(x_i, y_i) \overset{iid}{\sim} \mu_n$ converge weakly by [112, Thm. 1.1], since the distributions of (5.14) are precisely the finite-dimensional distributions of M_n. The exchangeable array (5.14) is precisely the list of edges of the multigraph $k\mathsf{G}^{(k)}[\mathsf{M}_n]$ listed in arbitrary order after assigning its vertices iid labels in $\mathrm{Unif}([0,1])$. Formally, we have $\mathsf{M}_{\mathsf{G}^{(k)}[\mathsf{M}_n]} \overset{d}{=} \frac{1}{k}\sum_{i=1}^{k} \delta_{(x_i^{(n)}, y_i^{(n)})}$ for all $n, k \in \mathbb{N}$. Thus, the sequence (5.14) converges weakly for each $k \in \mathbb{N}$ precisely when the sequence of random graphs $(\mathsf{G}^{(k)}[\mathsf{M}_n])_n$ converges weakly. □

In particular, a sequence of graphs $(G_n \in \Delta^{n\times n})$ quotient-converges precisely when the sequence of random graphs $(\mathsf{G}^{(k)}[\mathsf{M}_{G_n}])$—obtained by independently sampling k edges from each G_n proportionally to their edge weights and randomly labeling the vertices—converges weakly for each $k \in \mathbb{N}$.

Remark 5.4.2 (Edge-exchangeable random graphs)**.** *The random graphs we obtain by sampling edges from a grapheur appeared previously in the literature as edge-exchangeable random multigraph models [110]. As explained in [110, Rmk. 4.4], every edge-exchangeable random multigraph model is obtained from a random exchangeable measure on* $[0,1]^2$ *as above after labelling the vertices by iid labels from* $\mathrm{Unif}([0,1])$. *Nevertheless, to our knowledge graph limits based on these edge-exchangeable random graphs as in Proposition 5.4.1 have not been previously studied in the literature, see also Section 5.1.*

As noted in [110], if we repeat the above edge sampling process infinitely-many times, each edge (T_i, T_j) corresponding to a positive edge weight $E_{i,j} > 0$ will be repeated infinitely-many times. In contrast, all the other edges sampled above will only appear once. In particular, the vertices T', T_1', T_2' sampled in this process will only appear in one edge. Such vertices are called "blips" in [24, 110], and the isolated edges and self-loops are called "dust".

Edge-Based Szemerédi Regularity

We now provide bounds on the rate of convergence for $W_\square(\mathsf{M}_{\mathsf{G}^{(n)}[\mathsf{M}]}, \mathsf{M})$ to zero as n grows large, both in expectation and with high probability. In bounding the quantity $W_\square(\mathsf{M}_{\mathsf{G}^{(n)}[\mathsf{M}]}, \mathsf{M})$, it is important to distinguish between the randomness in the grapheur $\mathsf{M}_{\mathsf{G}^{(n)}[\mathsf{M}]}$ coming from the random atoms T_i and randomness coming from $\mathsf{G}^{(n)}[\mathsf{M}]$. In our next result, we fix a realization of $\mathsf{G}^{(n)}[\mathsf{M}]$, compute the distance $W_\square(\mathsf{M}_{\mathsf{G}^{(n)}[\mathsf{M}]}, \mathsf{M})$ depending on this realization (and hence random), and seek to bound both its expectation and its tails with respect to the randomness in $\mathsf{G}^{(n)}[\mathsf{M}]$.

In fact, our results provide convergence rates for the larger metric $\overline{W_\square}(\mathsf{M},\mathsf{N})$ defined by

$$W_\square(\mathsf{M},\mathsf{N}) \le \overline{W_\square}(\mathsf{M},\mathsf{N}) = \inf_{\substack{\text{random } (\mathsf{M}'_1,\mathsf{M}'_2) \\ \mathsf{M}'_i \overset{d}{=} \mathsf{M}_i \text{ for } i=1,2}} \mathbb{E}\left[\sup_{\substack{S,T\subseteq[0,1] \\ \text{intervals}}} |\mathsf{M}'_1(S\times T) - \mathsf{M}'_2(S\times T)|\right],$$

where the expectation is taken with respect to the coupling $(\mathsf{M}'_1,\mathsf{M}'_2)$ of M_1 and M_2. Note that $\overline{W_\square}$ differs from $W_\square$ in (5.6) by moving the supremum over axis-aligned rectangles into the expectation.

Theorem 5.4.3. *For any grapheur $\mathsf{M} \in \mathcal{M}$ and each $n \in \mathbb{N}$, let $\mathsf{G}^{(n)}[\mathsf{M}]$ be a random graph obtained by sampling n edges from M. Then*

$$\mathbb{E}_{\mathsf{G}^{(n)}[\mathsf{M}]}\left[W_\square(\mathsf{M},\mathsf{M}_{\mathsf{G}^{(n)}[\mathsf{M}]})\right] \le \mathbb{E}_{\mathsf{G}^{(n)}[\mathsf{M}]}\left[\overline{W_\square}(\mathsf{M},\mathsf{M}_{\mathsf{G}^{(n)}[\mathsf{M}]})\right] \le \frac{174}{\sqrt{n}}.$$

Moreover, $W_\square(\mathsf{M},\mathsf{M}_{\mathsf{G}^{(n)}[\mathsf{M}]}) \le \overline{W_\square}(\mathsf{M},\mathsf{M}_{\mathsf{G}^{(n)}[\mathsf{M}]}) \le \frac{174+\epsilon}{\sqrt{n}}$ with probability at least $1-\exp(-2\epsilon^2)$ over the random graph $\mathsf{G}^{(n)}[\mathsf{M}]$.

The proof of Theorem 5.4.3 relies on two key lemmas. The first involves the construction of a particular coupling of M, $\mathsf{M}_{\mathsf{G}^{(n)}[\mathsf{M}]}$, and $\mathsf{G}^{(n)}[\mathsf{M}]$.

Lemma 5.4.4. *There is a coupling $(\mathsf{M}',\mathsf{M}'_{\mathsf{G}^{(n)}},\mathsf{G}^{(n)})$ of M, $\mathsf{M}_{\mathsf{G}^{(n)}[\mathsf{M}]}$, and $\mathsf{G}^{(n)}[\mathsf{M}]$ satisfying the following two properties. First, in this coupling $\mathsf{M}'_{\mathsf{G}^{(n)}}$ is obtained by sampling $\mu \sim \mathsf{M}'$, then sampling n iid points from μ, and finally forming the associated empirical measure. Second, conditioned on $\mathsf{G}^{(n)} = G$, the pair $(\mathsf{M}',\mathsf{M}'_{\mathsf{G}^{(n)}})$ is a coupling of M and M_G.*

The second step in the proof of Theorem 5.4.3 involves using Vapnik-Chervonenkis theory to uniformly bound the discrepancy between a measure and its empirical sample on axis-aligned rectangles.

Lemma 5.4.5. *For any measure $\mu \in \mathcal{P}([0,1]^2)$, let $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ be the random empirical measure obtained by sampling $X_1,\ldots,X_n \overset{iid}{\sim} \mu$. Then*

$$\mathbb{E}_{\mu_n} \sup_{\substack{S,T\subseteq[0,1] \\ \textit{intervals}}} |\mu(S\times T) - \mu_n(S\times T)| \le \frac{174}{\sqrt{n}}.$$

The proofs of these two lemmas are straightforward but technical, and we defer them to Section 5.4 for clarity of exposition. Using the lemmas, we can now prove Theorem 5.4.3.

Proof (Theorem 5.4.3). Let $\mathcal{AAR} = \{[a,b]\times[c,d] : a,b,c,d \in [0,1]\}$ be the class of axis-aligned rectangles contained in $[0,1]^2$, and let $(\mathsf{M}', \mathsf{M}'_{\mathsf{G}^{(n)}}, \mathsf{G}^{(n)})$ be the coupling from Lemma 5.4.4. Define

$$f(\mathsf{G}^{(n)}[\mathsf{M}]) = f((x_1,y_1),\ldots,(x_n,y_n)) = \overline{W_\square}(\mathsf{M}_{\mathsf{G}^{(n)}[\mathsf{M}]}, \mathsf{M}),$$

which is a function of the edges $(x_1,y_1),\ldots,(x_n,y_n)$ sampled iid from $\mu \sim \mathsf{M}$ to create $\mathsf{G}^{(n)}[\mathsf{M}]$. First note that

$$\mathbb{E}_{\mathsf{G}^{(n)}[\mathsf{M}]} f(\mathsf{G}^{(n)}[\mathsf{M}]) = \mathbb{E}_{\mathsf{G}^{(n)}} f(\mathsf{G}^{(n)}) \le \mathbb{E}_{(\mathsf{M}',\mathsf{M}'_{\mathsf{G}^{(n)}},\mathsf{G}^{(n)})} \sup_{R\in\mathcal{AAR}} |\mathsf{M}'(R) - \mathsf{M}'_{\mathsf{G}^{(n)}}(R)| \le \frac{174}{\sqrt{n}},$$

where the equality follows from $\mathsf{G}^{(n)}[\mathsf{M}] \overset{d}{=} \mathsf{G}^{(n)}$, the first inequality follows by considering the coupling $(\mathsf{M}', \mathsf{M}'_{\mathsf{G}^{(n)}})|\mathsf{G}^{(n)}$ of M and $\mathsf{M}_{\mathsf{G}^{(n)}}$, and the last inequality follows by Lemma 5.4.5 and the fact that $\mathsf{M}'_{\mathsf{G}^{(n)}}$ is an empirical measure obtained from n iid samples from M'.

Second, note that f is $1/n$-Lipschitz in each (x_i,y_i). Indeed, if $\mathsf{G}^{(n,i)}[\mathsf{M}]$ differs from $\mathsf{G}^{(n)}[\mathsf{M}]$ only in the ith location (x_i,y_i), and we glue the two couplings $(\mathsf{M}'', \mathsf{M}''_{\mathsf{G}^{(n)}[\mathsf{M}]})$ and $(\mathsf{M}'', \mathsf{M}''_{\mathsf{G}^{(n,i)}[\mathsf{M}]})$ attaining the corresponding $\overline{W_\square}$ distances as in [220, §1], then

$$|f(\mathsf{G}^{(n,i)}[\mathsf{M}]) - f(\mathsf{G}^{(n)}[\mathsf{M}])| \le \mathbb{E}_{(\mathsf{M}''_{\mathsf{G}^{(n)}[\mathsf{M}]},\mathsf{M}''_{\mathsf{G}^{(n,i)}[\mathsf{M}]})} \sup_{R\in\mathcal{AAR}} |\mathsf{M}''_{\mathsf{G}^{(n)}[\mathsf{M}]}(R) - \mathsf{M}''_{\mathsf{G}^{(n,i)}[\mathsf{M}]}(R)| \le \frac{1}{n},$$

since the only difference between $\mathsf{M}''_{\mathsf{G}^{(n,i)}[\mathsf{M}]}$ and $\mathsf{M}''_{\mathsf{G}^{(n)}[\mathsf{M}]}$ is a weight of $1/n$ placed at one of their atoms. Since $(x_1,y_1),\ldots,(x_n,y_n)$ are sampled iid from a fixed measure μ (after it was sampled once from M), the bounded-difference concentration inequality [223, Cor. 2.21] implies that $f(\mathsf{G}^{(n)}[\mathsf{M}]) \le \mathbb{E} f(\mathsf{G}^{(n)}[\mathsf{M}]) + \epsilon/\sqrt{n}$ with probability at least $1-\exp(-2\epsilon^2)$, proving the theorem. □

Using Theorem 5.4.3, we can now prove Theorem 5.1.7 from Section 3.1.

Proof (Theorem 5.1.7). The first claim follows from Theorem 5.4.3. The second claim follows from the fact that $\mathbb{E}W_\square(\mathsf{M}, \mathsf{M}_{\mathsf{G}^{(n)}[\mathsf{M}]}) \le \frac{174}{\sqrt{n}}$, which implies that some realization $G^{(n)}$ of $\mathsf{G}^{(n)}[\mathsf{M}]$, satisfies $W_\square(\mathsf{M}, \mathsf{M}_{G^{(n)}}) \le \frac{174}{\sqrt{n}}$. Setting $n = k(\epsilon) = \lceil (174/\epsilon)^2 \rceil$, we obtain the second claim. □

We remark that $\overline{W_\square}$ also metrizes weak converge in $\mathcal{M}$, as can be shown by combining Theorem 5.4.3 with Proposition 5.4.1. We opted to use the $W_\square$ metric in the rest of the chapter since it is easier to relate to random quotients as in Proposition 5.3.2. We conclude this section by remarking that the above discussion applies to undirected random graphs sampled from symmetric grapheurs.

Remark 5.4.6 (Undirected graphs)**.** *If* M *is the limit of undirected graphs, then by Corollary 5.3.7 it is symmetric, i.e., we have* $\mathsf{M}(S\times T) = \mathsf{M}(T\times S)$ *for any measurable* $S, T \subseteq [0,1]$*. In that case, we can sample undirected graphs from* M *by first sampling* $\mathsf{G}^{(n)}[\mathsf{M}]$ *as above and symmetrizing it to get* $\mathsf{G}^{(n)}_{\mathrm{sym}}[\mathsf{M}] = \frac{\mathsf{G}^{(n)}[\mathsf{M}]+(\mathsf{G}^{(n)}[\mathsf{M}])^\top}{2}$*. For any* $S, T \subseteq [0,1]$*, we have*

$$|\mathsf{M}(S\times T) - \mathsf{M}_{\mathsf{G}^{(n)}_{\mathrm{sym}}[\mathsf{M}]}(S\times T)| = \frac{1}{2}|\mathsf{M}(S\times T) + \mathsf{M}(T\times S) - \mathsf{M}_{\mathsf{G}^{(n)}[\mathsf{M}]}(S\times T) - \mathsf{M}_{\mathsf{G}^{(n)}[\mathsf{M}]}(T\times S)|$$
$$\leq \max\{|\mathsf{M}(S\times T) - \mathsf{M}_{\mathsf{G}^{(n)}[\mathsf{M}]}(S\times T)|, |\mathsf{M}(T\times S) - \mathsf{M}_{\mathsf{G}^{(n)}[\mathsf{M}]}(T\times S)|\},$$

hence $W_\square(\mathsf{M}, \mathsf{M}_{\mathsf{G}^{(n)}_{\mathrm{sym}}[\mathsf{M}]}) \leq W_\square(\mathsf{M}, \mathsf{M}_{\mathsf{G}^{(n)}[\mathsf{M}]})$ *and similarly for* $\overline{W_\square}$*. Thus, all the results of this section apply to the undirected random graph* $\mathsf{G}^{(n)}_{\mathrm{sym}}[\mathsf{M}]$*.*

Property Testing

We apply the preceding edge-based sampling procedure to test properties of large graphs based on small summaries. We begin by formally defining the class of graph parameters we can test using our framework. Here a *graph parameter* is a function $f\colon \bigsqcup_{n\in\mathbb{N}} \Delta^{n\times n} \to \mathbb{R}$ on graphs of all sizes that is permutation-invariant, in the sense that $f(\pi_n G_n) = f(G_n)$ for any permutation $\pi_n \in \mathsf{S}_n$ of the vertex labels.

Definition 5.4.7 (Quotient-testable graph parameters)**.** *A* quotient-testable graph parameter *is a graph parameter* $f\colon \bigsqcup_{n\in\mathbb{N}} \Delta^{n\times n} \to \mathbb{R}$ *satisfying the following properties:*

(Isolate-indifference) $f(G_N) = f(G_n)$ *whenever* $G_N \in \Delta^{N\times N}$ *differs from* $G_n \in \Delta^{n\times n}$ *by isolated vertices;*

(Continuity) *The sequence* $(f(G_n))$ *converges whenever* $(G_n \in \Delta^{n\times n})$ *is quotient-convergent.*

We proceed to give several equivalent characterizations of quotient-testable parameters.

Proposition 5.4.8 (Testability as continuity)**.** *The following are equivalent for a graph parameter* f*.*

1. *The parameter f is quotient-testable.*

2. *There is a function $\bar{f}\colon \mathcal{M} \to \mathbb{R}$ satisfying $\bar{f}(\mathsf{M}_G) = f(G)$ for all $G \in \bigsqcup_n \Delta^{n\times n}$. Such $\bar{f}$ is then unique and continuous.*

3. *For any $\epsilon > 0$ there is a $\delta = \delta(\epsilon) > 0$ such that if $W_\square(\mathsf{M}_{G_n}, \mathsf{M}_{G_m}) \leq \delta$ then $|f(G_n) - f(G_m)| \leq \epsilon$ for any two graphs G_n, G_m of any two sizes.*

Proof. We prove these statements imply each other in order, with the last statement implying the first.

Suppose f is quotient-testable. Define $\mathsf{M}_{<\infty} = \{\mathsf{M}_G : G \in \Delta^{n\times n} \text{ for some } n\} \to \mathbb{R}$ to be the grapheurs in M arising from finite graphs and define $\bar{f}\colon \mathsf{M}_{<\infty} \to \mathbb{R}$ by $\bar{f}(\mathsf{M}_G) = f(G)$. This is well-defined because f is permutation-invariant and isolate-indifferent, as $\mathsf{M}_G = \mathsf{M}_{G'}$ if and only if G and G' differ by isolated vertices and node relabeling [113, Prop. 3]. Further, $\bar{f}$ is continuous by continuity of f, hence it extends uniquely to a continuous function on the closure $\mathcal{M}$ of $\mathcal{M}_{<\infty}$. This proves statement 2.

Suppose f extends to a continuous $\bar{f}\colon \mathcal{M} \to \mathbb{R}$. Since $\mathcal{M}$ is compact and this compact topology is metrized by the $W_\square$ metric by Proposition 5.3.3, the function $\bar{f}$ is uniformly continuous in the $W_\square$ metric. This proves statement 3.

Finally, assume statement 3 holds. Since $W_\square(\mathsf{M}_{G_N}, \mathsf{M}_{G_n}) = 0$ if G_n and G_N differ by isolated vertices, we conclude that f is isolate-indifferent. Since quotient-convergent sequences $(G_n)_n$ are Cauchy in the $W_\square$-metric by Theorem 5.1.5, we conclude that $(f(G_n))_n \subseteq \mathbb{R}$ is Cauchy and hence converges. □

As a first set of examples, quotient densities and homomorphism numbers are quotient-testable. More generally, we show next that a large class of polynomial graph parameters are testable, because all such polynomial parameters are functions of fixed-dimensional random quotients of their input graphs.

Proposition 5.4.9 (Polynomial parameters)**.** *Suppose f is an isolate-indifferent graph parameter such that $f|_{\Delta^{n\times n}}$ is given by a polynomial function for each n, and that all these polynomials have degree $d \in \mathbb{N}$. Then f is quotient-testable.*

Further, a graph parameter f is quotient-testable if and only if for any $\epsilon > 0$ there exists a polynomial quotient-testable parameter p satisfying $|f(G) - p(G)| \leq \epsilon$ for all graphs G of all sizes.

Proof. For the first claim, it was shown in [138, Thm. 5.2] that there exists a polynomial q_k over $\mathbb{R}^{k\times k}$ satisfying $f(G_n) = \mathbb{E}q_k(\rho(F_{k,n})G_n)$ for all $G_n \in \Delta^{n\times n}$ and all $n \in \mathbb{N}$. Since q_k is L-Lipschitz on $\Delta^{k\times k}$ for some $L > 0$, we have by Proposition 5.3.2 that

$$\begin{aligned}|f(G_n) - f(G_m)| = |\mathbb{E}q_k(\rho(F_{k,n}G_n)) - \mathbb{E}q_k(\rho(F_{k,m})G_m)| &\leq LW_1(\rho(F_{k,n})G_n, \rho(F_{k,m})G_m)\\ &\leq Lk^2 W_\square(\mathsf{M}_{G_n}, \mathsf{M}_{G_m}),\end{aligned}$$

for any two graphs G_n, G_m of any two sizes. Thus, the function f is testable by Proposition 5.4.8(3), as can be seen by setting $\delta(\epsilon) = \frac{\epsilon}{Lk^2}$.

For the second claim, if for any $\epsilon > 0$ there is a testable parameter p satisfying $|f(G) - p(G)| \leq \epsilon$ for all G, then f is a uniform limit of $W_\square$-continuous functions and hence is $W_\square$-continuous itself. By Proposition 5.4.8, the parameter f is therefore testable. Conversely, if f is testable then it extends to a continuous $\bar{f}\colon \mathcal{M} \to \mathbb{R}$ by Proposition 5.4.8 again. Note that the collection of polynomial testable parameters is an algebra, contains constant functions, and separates points by Corollary 5.3.4. Since $\mathcal{M}$ is compact, Stone–Weierestrass implies that $\bar{f}$ is the uniform limit of polynomial testable parameters. $\square$

We combine the insights of Sections 5.4-5.4 along with the above characterization of quotient-testable parameters to show that these can indeed be tested by edge sampling.

Theorem 5.4.10. *Consider a graph parameter $f\colon \bigsqcup_n \Delta^{n\times n} \to \mathbb{R}$ whose restrictions $f|_{\Delta^{n\times n}}$ are continuous (in the usual topology on $\Delta^{n\times n}$) for each n. Then f is quotient-testable if and only if for every $\epsilon > 0$, there exists $k \in \mathbb{N}$ such that for any finite graph G of any size we have $|f(G) - f(\mathsf{G}^{(k)}[\mathsf{M}_G])| \leq \epsilon$ with probability at least $1 - \epsilon$. In this case, we can take $k(\epsilon) = \left\lceil \left(\frac{174+\sqrt{\log(\epsilon^{-1/2})}}{\delta(\epsilon)}\right)^2 \right\rceil$ where $\delta(\epsilon) > 0$ is as in Proposition 5.4.8(3).*

Proof. Suppose f is quotient-testable. For any $\epsilon > 0$, let $\delta(\epsilon)$ be as in Proposition 5.4.8(3) and set $k = k(\epsilon)$ as in the theorem statement. Then, noting that $\epsilon = e^{-2\tilde{\epsilon}^2}$ where $\tilde{\epsilon} = \sqrt{\log(\epsilon^{-1/2})}$, with probability at least $1-\epsilon$ we have $W_\square(\mathsf{M}_G, \mathsf{M}_{\mathsf{G}^{(k)}}) \leq \frac{174+\tilde{\epsilon}}{\sqrt{k}} \leq \delta(\epsilon)$ and hence $|f(G) - f(\mathsf{G}^{(k)}[\mathsf{M}_G])| \leq \epsilon$ by Theorem 5.4.3.

We turn to proving the converse. If G_1 and G_2 differ by isolated vertices, then $\mathsf{M}_{G_1} = \mathsf{M}_{G_2}$ and hence $\mathsf{G}^{(k)}[\mathsf{M}_{G_1}] \stackrel{d}{=} \mathsf{G}^{(k)}[\mathsf{M}_{G_2}]$. Therefore, for any $\epsilon > 0$ let $k \in \mathbb{N}$ be as in the theorem statement and let $\mathsf{G}^{(k)}$ be a sample from this common distribution. Then with probability $1-2\epsilon$ we have $|f(G_1)-f(\mathsf{G}^{(k)})| \le \epsilon$ and $|f(\mathsf{G}^{(k)})-f(G_2)| \le \epsilon$, so in particular there is a realization $G^{(k)}$ of $\mathsf{G}^{(k)}$ satisfying both inequalities. We then have $|f(G_1)-f(G_2)| \le |f(G_1)-f(G^{(k)})| + |f(G^{(k)})-f(G_2)| \le 2\epsilon$. Since $\epsilon > 0$ was arbitrary, we conclude that f is isolate-indifferent. A similar argument shows that $|f|$ is bounded. Indeed, since $|f|$ restricted to the compact set $\Delta^{2k\times 2k}$ is continuous, it is bounded by some $B_{2k} > 0$. For any other graph G of any size, some realization $G^{(k)}$ of $\mathsf{G}^{(k)}[\mathsf{M}_G]$ satisfies $|f(G) - f(G^{(k)})| \le \epsilon$, and hence $|f(G)| \le B_{2k} + \epsilon$ for all graphs G. Denoting $B = \sup_G |f(G)|$, we then have $|f(G) - \mathbb{E} f(\mathsf{G}^{(k)}[\mathsf{M}_G])| \le \mathbb{E}|f(G) - f(\mathsf{G}^{(k)}[\mathsf{M}_G])| \le \epsilon(1-\epsilon) + 2B\epsilon$. Finally, if (G_n) is quotient-convergent, then $(\mathsf{G}^{(k)}[\mathsf{M}_{G_n}])_n$ is weakly-convergent, and by continuity of $f|_{\Delta^{2k\times 2k}}$, the sequence $(\mathbb{E} f(\mathsf{G}^{(k)}[\mathsf{M}_{G_n}]))_n$ is convergent. Therefore, we have

$$|f(G_n)-f(G_m)| \le 2\epsilon(1-\epsilon+2B)+|\mathbb{E} f(\mathsf{G}^{(k)}[\mathsf{M}_{G_n}])-\mathbb{E} f(\mathsf{G}^{(k)}[\mathsf{M}_{G_m}])| \le 3\epsilon(1-\epsilon+2B)$$

for all large n, m, showing that $(f(G_n))$ is convergent. □

We remark that the graphon analog of the characterization of quotient-testable parameters in Theorem 5.4.10 was used as the definition of testable parameters in [20, Def. 2.11].

The upshot of this result is that a continuous graph parameter being quotient-testable is equivalent to the parameter being testable using edge sampling. In particular, Theorem 5.4.10 gives a simple recipe for estimating the value of a testable parameter $f(G)$ on an arbitrarily-large graph G: sample k edges iid from G proportionally to their edge weights, and compute the parameter $f(G^{(k)})$ of the resulting small graph $G^{(k)}$. The number of edges needed is given explicitly as $k = \left\lceil \left(\frac{174+\sqrt{\log(\epsilon^{-1/2})}}{\delta(\epsilon)}\right)^2 \right\rceil$ in terms of the desired accuracy ϵ and the modulus of continuity $\delta(\epsilon)$ of f in the $W_\square$-metric. When f is L-Lipschitz continuous in the $W_\square$-metric, we can set $\delta(\epsilon) = \epsilon/L$.

Many graph parameters of interest are indeed Lipschitz-continuous, including all polynomial parameters as the proof of Proposition 5.4.9 demonstrates. Thus, the key quantity that is needed to bound the required number of edges to test the values of such parameters is their Lipschitz constants. We can explicitly bound the Lipschitz constants of quotient densities. Specifically, for a multigraph $H \in \mathbb{N}^{k\times k}$ we have

$$|t_Q(H; \mathsf{M}_1) - t_Q(H; \mathsf{M}_2)| = |\mathbb{E}\mathsf{G}_k[\mathsf{M}_1]^H - \mathbb{E}\mathsf{G}_k[\mathsf{M}_2]^H| \le k^2 \|H\|_\infty W_\square(\mathsf{M}_1, \mathsf{M}_2),$$

since the polynomial $G \mapsto G^H$ is $\|H\|_\infty$-Lipschitz with respect to the ℓ_1 norm on $\Delta^{k\times k}$. By using (5.9) and (5.11), we can similarly derive Lipschitz constants for homomorphism numbers. In the next example, we derive the Lipschitz constant of a particular graph parameter that can be used to test the presence of hubs.

Example 5.4.11 (Testing for hubs)**.** *Consider the graph parameter*

$$f(G) = \hom(S_2^{(\mathrm{out})}; G) + \hom(S_2^{(\mathrm{in})}; G) = \|G\mathbb{1}\|_2^2 + \|G^\top \mathbb{1}\|_2^2,$$

where $S_2^{(\mathrm{out})}, S_2^{(\mathrm{in})} \in \mathbb{N}^{3\times 3}$ are outward- and inward-pointing stars with two leaves, respectively. This is an extension to weighted and directed graphs of the so-called first Zagreb index, originally introduced for chemical networks in [94]. Define the polynomial q_2 on 2×2 matrices by

$$q_2(G) = (G_{1,1} + G_{1,2} - G_{2,1} - G_{2,2})^2 + (G_{1,1} + G_{2,1} - G_{1,2} - G_{2,2})^2,$$

and observe that $f(G_n) = \mathbb{E} q_2(\rho(F_{2,n})G_n)$ for any $G_n \in \Delta^{n\times n}$ and any $n \in \mathbb{N}$. Indeed, more generally for any grapheur M with random locations (T_i), let $R_i \overset{iid}{\sim}$ $\mathrm{Unif}(\{\pm 1\})$ be iid Rademacher random variables and note that

$$\begin{aligned}
\mathbb{E} q_2(\mathsf{G}_2[\mathsf{M}]) &= \mathbb{E}\left[\sum_{i\in\mathbb{N}} (E\mathbb{1} + \sigma)_i \left(\mathbb{1}_{T_i<1/2} - \mathbb{1}_{T_i>1/2}\right)\right]^2 + \mathbb{E}\left[\sum_{i\in\mathbb{N}} (E^\top\mathbb{1} + \varsigma)_i \left(\mathbb{1}_{T_i<1/2} - \mathbb{1}_{T_i>1/2}\right)\right]^2 \\
&= \mathbb{E}\left[\sum_{i\in\mathbb{N}} (E\mathbb{1} + \sigma)_i R_i\right]^2 + \mathbb{E}\left[\sum_{i\in\mathbb{N}} (E^\top\mathbb{1} + \varsigma)_i R_i\right]^2 = \|E\mathbb{1} + \sigma\|_2^2 + \|E^\top\mathbb{1} + \varsigma\|_2^2,
\end{aligned}$$

where $\mathbb{1}_{T_i<1/2} = 1$ if $T_i < 1/2$ and zero otherwise and similarly for $\mathbb{1}_{T_i>1/2}$, and we note that $(\mathbb{1}_{T_i<1/2} - \mathbb{1}_{T_i>1/2})_{i\in\mathbb{N}} \overset{d}{=} (R_i)_{i\in\mathbb{N}}$. We conclude that f can be extended to all grapheurs M by $\bar{f}(\mathsf{M}) = \mathbb{E} q_2(\mathsf{G}_2[\mathsf{M}])$. This extension is Lipschitz-continuous. Explicitly, since $|\partial_{G_{i,j}} q_2(G)| = 2|G_{1,1} - G_{2,2}|$ if $i = j$ and $2|G_{i,j} - G_{j,i}|$ if $i \neq j$, we conclude that q_2 is 2-Lipschitz on $\Delta^{2\times 2}$ with respect to $\|\cdot\|_1$, and hence that

$$|\bar{f}(\mathsf{M}_1) - \bar{f}(\mathsf{M}_2)| = |\mathbb{E} q_2(\mathsf{G}_2[\mathsf{M}_1]) - \mathbb{E} q_2(\mathsf{G}_2[\mathsf{M}_2])| \leq 2W_1(\mathsf{G}_2[\mathsf{M}_1], \mathsf{G}_2[\mathsf{M}_2]) \leq 8W_\square(\mathsf{M}_1, \mathsf{M}_2),$$

for any two grapheurs $\mathsf{M}_1, \mathsf{M}_2 \in \mathcal{M}$ by Proposition 5.3.2. Finally, observe that $\bar{f}(\mathsf{M}) = 0$ if and only if $E = 0$ and $\sigma = \varsigma = 0$, hence if and only if M has no hubs. In particular, $W_\square(\mathsf{M}_G, \mathsf{M}) \geq f(G)/8$ for any grapheur M with no hubs.

We can now use the above parameter to test for the existence of significant hubs in a graph. Specifically, suppose we are given an arbitrarily-large graph G, and we sample $k = \left\lceil 64\left(\frac{174+\sqrt{\log(\epsilon^{-1/2})}}{\epsilon}\right)^2\right\rceil$ edges from G to obtain a graph $G^{(k)}$. Then with *probability* $1-\epsilon$*, we have* $W_\square(\mathsf{M}_G, \mathsf{M}) \geq \frac{f(G^{(k)})-\epsilon}{8}$ *for any grapheur* M *with no hubs. In this sense, large values of* $f(G^{(k)})$ *indicate the likely presence of significant hubs in* G*.*

We observe that testing for hubs also has an appealing interpretation in terms of quotients. A grapheur M has no hubs if and only if there is a $\theta \in [0,1]$ satisfying $\mathsf{G}_m[\mathsf{M}] = \theta\frac{1}{m^2}\mathbb{1}\mathbb{1}^\top + (1-\theta)\frac{1}{m}I_m$ (a deterministic matrix) for all $m \in \mathbb{N}$. Thus, testing for hubs also amounts to testing whether the random quotients of a graph G deviate from a uniform distribution of edge weights, i.e., quotients of the above form. See Figures 1.2-1.3 for example. The parameter of Example 5.4.11 yields a lower bound on this deviation, as by Proposition 5.3.2 we have

$$\min_{\theta\in[0,1]} \mathbb{E}\left\|\theta\frac{1}{m^2}\mathbb{1}\mathbb{1}^\top + (1-\theta)\frac{1}{m}I_m - \mathsf{G}_m[\mathsf{M}_G]\right\|_1 \geq \frac{f(G)}{8} - \frac{4}{m}.$$

For example, the values of the parameter f on the full graphs in Figures 1.2-1.3 are 0.0073 and 0.010, respectively. Estimating this parameter by sampling $k = 10^3$ edges, we obtain values of 0.0084 and 0.012, respectively. For comparison, the value of this parameter on an Erdős–Rényi random graph with edge probability $1/2$ on 1,500 vertices is 0.001.

We end this section by giving two further examples of quotient-testable parameters, each quantifying different structural properties.

Example 5.4.12 (Global Clustering Coefficient)**.** *One extension of the global clustering coefficient to weighted* undirected *graphs, a global analog of the coefficient proposed in [234, §7] for gene co-expression networks, is*

$$f(G) = \frac{\operatorname{inj}(K_3; G)}{\inf(P_2; G)\max_{i,j} G_{i,j}}, \tag{5.15}$$

where $\operatorname{inj}(H;G)$ *denotes the injective homomorphism number and is defined as in* (5.1) *but with the sum restricted to injective maps,* K_3 *is an arbitrarily-directed triangle graph, and* P_2 *is the directed path on three vertices. We proceed to show that* (5.15) *is quotient-testable. Proposition 5.4.9 shows that* $\operatorname{inj}(K_3;G)$ *and* $\operatorname{inj}(P_2;G)$ *extend to continuous functions on all of* $\mathcal{M}$*. Similarly, we note that* $G \mapsto \max_{i,j} G_{i,j}$ *extends to such a continuous function, since if* (G_n) *quotient-converges to a grapheur described by* $(E, \sigma, \varsigma, \theta, \vartheta)$*, then* $\max_{i,j}(G_n)_{i,j} \to \max_{i,j} E_{i,j}$ *by Proposition 5.3.6. Finally, note that* $f(G) \leq 1$ *for all graphs* G *of any size, proving that* f *extends to a continuous function on all of* $\mathcal{M}$ *and hence is quotient-testable by Proposition 5.4.8.*

Example 5.4.13 (Katz Centrality). *Another measure of importance, or influence, of a node, originally introduced in [118] for social networks, is its Katz centrality. Specifically, for a parameter* $\alpha \in (0,1)$*, we consider the sum of the Katz centralities of all the nodes*

$$f(G) = \mathbb{1}^\top \Big[(I - \alpha G)^{-1} - I \Big] \mathbb{1} = \sum_{k \geq 1} \alpha^k (\mathbb{1}^\top G^k \mathbb{1}),$$

which is another parameter quantifying the presence of hubs in a network. Once again, this parameter is quotient-testable. Indeed, if (G_n) *quotient-converges, then* $\mathbb{1}^\top G_n^k \mathbb{1}$ *converges for each* k *by Proposition 5.4.9, hence* $f(G_n)$ *converges as well for any fixed* $\alpha \in (0,1)$.

Missing Proofs from Section 5.4

In this section we give all the proofs missing from Section 5.4.

Proofs from Section 5.4

In this section, we prove the two lemmas used in Section 5.4 to prove Theorem 5.4.3. We begin by proving Lemma 5.4.4 by constructing an explicit coupling between M, $\mathsf{M}_{\mathsf{G}^{(n)}[\mathsf{M}]}$, and $\mathsf{G}^{(n)}[\mathsf{M}]$ with the claimed properties.

Proof (Lemma 5.4.4). Sample $\mu \sim \mathsf{M}$ and set $\mathsf{M}' = \mu$. Next, sample $(x_i, y_i) \overset{iid}{\sim} \mu$, let $\{t_1, \dots, t_N\} = \{x_i\} \cup \{y_i\}$, form $\mathsf{G}^{(n)}$ as in Section 5.4, and set

$$\mathsf{M}'_{\mathsf{G}^{(n)}} = \frac{1}{n} \sum_{i=1}^{n} \delta_{(x_i, y_i)} = \sum_{i,j=1}^{N} (\mathsf{G}^{(n)})_{i,j} \delta_{(t_i, t_j)}. \tag{5.16}$$

We proceed to argue that this coupling satisfies all the claimed properties.

By construction, we have $\mathsf{M}' \overset{d}{=} \mathsf{M}$ and $\mathsf{G}^{(n)} \overset{d}{=} \mathsf{G}^{(n)}[\mathsf{M}]$. Next, because M is exchangeable, we claim that $t_1, \dots, t_N \overset{iid}{\sim} \mathrm{Unif}([0,1])$. Indeed, $(x_i, y_i) \overset{d}{=} (\sigma(x_i), \sigma(y_i))$ for all $\sigma \in S_{[0,1]}$, hence $(x_i, y_i) \sim \theta\lambda^2 + (1-\theta)\lambda_D$. Therefore, for each i either $x_i = y_i \sim \mathrm{Unif}([0,1])$ or $x_i, y_i \overset{iid}{\sim} \lambda^2$. Combining this with the fact that the (x_i, y_i) are independent for different i, we obtain the claim. Therefore, $\mathsf{M}_{\mathsf{G}^{(n)}[\mathsf{M}]} \overset{d}{=} \frac{1}{n} \sum_{i=1}^{n} \delta_{(x_i, y_i)} = \mathsf{M}'_{\mathsf{G}^{(n)}}$, the latter being an empirical measure drawn from $\mu \sim \mathsf{M}$. Finally, we have $\mathsf{M}' | \mathsf{G}^{(n)} \overset{d}{=} \mathsf{M}$ and $\mathsf{M}'_{\mathsf{G}^{(n)}} | \mathsf{G}^{(n)} \overset{d}{=} \mathsf{M}_{\mathsf{G}^{(n)}}$ as can be seen from (5.16), as $\mathsf{G}^{(n)}$ only depends on the multiplicities of $\{(t_i, t_j)\}$ rather than their locations, and the multiplicities and locations of a random exchangeable measure are independent of each other [113, Thm. 2]. $\square$

We now turn to proving Lemma 5.4.5, combining standard results from VC theory.

Proof (Lemma 5.4.5). Fix any $\mu \in \mathcal{P}([0,1]^2)$. By the proof of [223, Thm. 4.10], we have

$$\mathop{\mathbb{E}}_{\substack{\mu^{(n)} \\ \text{emp. meas.}\sim\mu}} \sup_{R \in \mathcal{AAR}} |\mu_n(R) - \mu(R)| \leq 2R_n(\mathcal{F}_{\mathcal{AAR}}),$$

where $R_n(\mathcal{F}) = \mathop{\mathbb{E}}_{\substack{X_i \overset{iid}{\sim} \mu \\ \varepsilon_i \overset{iid}{\sim} \mathrm{Unif}(\{\pm 1\})}} \sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(X_i) \right|$ is the Rademacher complexity of a class of functions $\mathcal{F}$ and we set $\mathcal{F}_{\mathcal{AAR}} = \{\mathbb{1}_R : R \in \mathcal{AAR}\}$. We further have by [223, Ex. 5.24] that

$$R_n(\mathcal{F}) \leq \frac{24}{\sqrt{n}} \int_0^{\mathrm{diam}(\mathcal{F})} \mathbb{E}\sqrt{\log N(t; \mathcal{F}; \|\cdot\|_{\mu_n})}\, dt,$$

where $N(t; \mathcal{F}; \|\cdot\|_{\mu_n})$ is the t-covering number of $\mathcal{F}$ in the (random) metric $\|f - g\|_{\mu_n} = \sqrt{\mathbb{E}_{\mu_n}(f-g)^2}$ defined by the empirical measure μ_n. If $\mathcal{F}$ is class of $\{0,1\}$-valued functions, then $\mathrm{diam}(\mathcal{F}) \leq 1$. Finally, we have by [98, Cor. 1] that

$$N(t; \mathcal{F}_{\mathcal{AAR}}; \|\cdot\|_{\mu}) \leq e(V+1)\left(\frac{2e}{t}\right)^V,$$

for any measure μ, where $V = \mathrm{VC}(\mathcal{AAR}) = 4$ is the VC-dimension of the class of axis-aligned rectangles. Combining the above results, we conclude that

$$\mathop{\mathbb{E}}_{\substack{\mu^{(n)} \\ \text{emp. meas.}\sim\mu}} \sup_{R \in \mathcal{AAR}} |\mu_n(R) - \mu(R)| \leq \frac{48}{\sqrt{n}} \int_0^1 \sqrt{\log(5e) + 4\log(2e/t)}\, dt \leq \frac{174}{\sqrt{n}},$$

where we numeriucally evaluated the last integral to 12 digits of accuracy to get $3.624000076562 \leq 3.625$. ◻

Proof of Proposition 5.3.6 from Section 5.3

Using Theorem 5.4.3 and its proof, we can now prove Proposition 5.3.6 by comparing both the quotient-convergent sequence (G_n) and its limit M to their edge-sampled random graphs.

Proof (Proposition 5.3.6). Since (G_n) quotient-converges to M, we have $W_1(\mathsf{G}^{(k)}[\mathsf{M}_{G_n}], \mathsf{G}^{(k)}[\mathsf{M}]) \to 0$ as $n \to \infty$ for each $k \in \mathbb{N}$. Pick $k \geq (174/\epsilon)^2$, sample $(G_n^{(k)}, G_{\mathsf{M}}^{(k)})$

from a coupling of $\mathsf{G}^{(k)}[\mathsf{M}_{G_n}]$ and $\mathsf{G}^{(k)}[\mathsf{M}]$ attaining $W_1(\mathsf{G}^{(k)}[\mathsf{M}_{G_n}], \mathsf{G}^{(k)}[\mathsf{M}])$, and note that with probability $1 - 4e^{-2\epsilon^2} - \frac{W_1(\mathsf{G}^{(k)}[\mathsf{M}_{G_n}],\mathsf{G}^{(k)}[\mathsf{M}])}{\epsilon}$ we have

$$\sup_{R\in\mathcal{ARR}} |\mu_{G_n^{(k)}}(R) - \mu_{G_n}(R)| \le \epsilon, \quad \sup_{R\in\mathcal{ARR}} |\mu_{G_{\mathsf{M}}^{(k)}}(R) - \mu_{\mathsf{M}}(R)| \le \epsilon, \quad \text{and} \quad \|G_n^{(k)} - G_{\mathsf{M}}^{(k)}\|_1 \le \epsilon, \tag{5.17}$$

for $(\mu_{G_{\mathsf{M}}^{(k)}}, \mu_{\mathsf{M}}) \sim (\mathsf{M}'_{G_{\mathsf{M}}^{(k)}}, \mathsf{M}')$ and $(\mu_{G_n^{(k)}}, \mu_{G_n}) \sim (\mathsf{M}'_{G_n^{(k)}}, \mathsf{M}'_{G_n})$ sampled from the coupling of Lemma 5.4.4 conditioned on $G_{\mathsf{M}}^{(k)}$ and $G_n^{(k)}$, respectively. Picking n large enough, the above occurs with positive probability, so we can find deterministic $G_n^{(k)}$ and $G_{\mathsf{M}}^{(k)}$ satisfying (5.17) for all large n. Explicitly, by Lemma 5.4.4 there are locations $t_i \in [0,1]$ and a permutation $\pi_n \in \mathsf{S}_n$ satisfying

$$\mu_{G_n} = \sum_{i,j=1}^{n} (G_n)_{i,j} \delta_{(t_i,t_j)}, \quad \mu_{G_n^{(k)}} = \sum_{i,j=1}^{n} (\pi_n G_n^{(k)})_{i,j} \delta_{(t_i,t_j)}.$$

By considering rectangles in (5.17) centered at each of the atoms of μ_{G_n} and $\mu_{G_n^{(k)}}$, and around rows and columns of these two measures, we conclude that

$$\|G_n - \pi_n G_n^{(k)}\|_\infty \le \epsilon, \quad \|G_n \mathbb{1} - \pi_n G_n^{(k)} \mathbb{1}\|_\infty \le \epsilon, \quad \text{and} \quad \|G_n^\top \mathbb{1} - \pi_n (G_n^{(k)})^\top \mathbb{1}\|_\infty \le \epsilon. \tag{5.18}$$

Similarly, we can find a permutation $\widetilde{\pi}_M \in \mathsf{S}_M$ for large enough $M \in \mathbb{N}$ and atoms $(t_i)_{i\in\mathbb{N}}$ and $(t'_i)_{i=1}^{M-m}$ for some $m \le M$ satisfying

$$\mu_{\mathsf{M}} = \sum_{i,j\in\mathbb{N}} E_{i,j} \delta_{(t_i,t_j)} + \sum_{i\in\mathbb{N}} \sigma_i (\delta_{t_i} \otimes \lambda) + \sum_{i\in\mathbb{N}} \varsigma_i (\lambda \otimes \delta_{t_i}) + \theta\lambda^2 + \vartheta\lambda_D,$$

$$\mu_{G_{\mathsf{M}}^{(k)}} = \sum_{i,j=1}^{m} (\widetilde{\pi}_M G_{\mathsf{M}}^{(k)})_{i,j} \delta_{(t_i,t_j)} + \sum_{i=1}^{m} \sum_{j=m+1}^{M} (\widetilde{\pi}_M G_{\mathsf{M}}^{(k)})_{i,j} \delta_{(t_i,t'_j)} + \sum_{i=m+1}^{M} \sum_{j=1}^{m} (\widetilde{\pi}_M G_{\mathsf{M}}^{(k)})_{i,j} \delta_{(t'_i,t_j)}$$
$$+ \sum_{i,j=m+1}^{M} (\widetilde{\pi}_M G_{\mathsf{M}}^{(k)})_{i,j} \delta_{(t'_i,t'_j)}.$$

Here we have separated the edges sampled from each summand in M. Moreover, with probability 1 we have $(\widetilde{\pi}_M G_{\mathsf{M}}^{(k)})_{i,j} \in \{0, 1/k\}$ if $i > m$ or $j > m$. Note that $k^{-1} \le (\epsilon/174)^2 \le \epsilon$ for all small $\epsilon > 0$. By considering similar rectangles in (5.17), we therefore

$$\|E - \widetilde{\pi}_M G_{\mathsf{M}}^{(k)}\|_\infty \le \epsilon, \quad \|E\mathbb{1} + \sigma - \widetilde{\pi}_M G_{\mathsf{M}}^{(k)} \mathbb{1}\|_\infty \le \epsilon, \quad \text{and} \quad \|E^\top \mathsf{1} + \varsigma - (\widetilde{\pi}_M G_{\mathsf{M}}^{(k)})^\top \mathbb{1}\|_\infty \le \epsilon. \tag{5.19}$$

Finally, since $\|G_n^{(k)} - G_{\mathsf{M}}^{(k)}\|_1 \le \epsilon$, we conclude that

$$\|G_n^{(k)} - G_{\mathbb{M}}^{(k)}\|_\infty \le \epsilon, \quad \|G_n^{(k)} \mathsf{1} - G_{\mathbb{M}}^{(k)} \mathsf{1}\|_\infty \le \epsilon, \quad \text{and} \quad \|(G_n^{(k)})^\top \mathbb{1} - (G_{\mathsf{M}}^{(k)})^\top \mathbb{1}\|_\infty \le \epsilon. \tag{5.20}$$

Combining (5.18), (5.19), and (5.20), we obtain

$$\|\widetilde{\pi}_M\pi_n^{-1}G_n - E\|_\infty \leq 3\epsilon, \quad \|\widetilde{\pi}_M\pi_n^{-1}G_n\mathbb{1} - (E\mathbb{1}+\sigma)\|_\infty \leq 3\epsilon, \quad \text{and}$$
$$\|\widetilde{\pi}_M\pi_n^{-1}G_n^\top\mathbb{1} - (E^\top\mathbb{1}+\varsigma)\|_\infty \leq 3\epsilon.$$

Since $\epsilon > 0$ was arbitrary, our first three claims are proved. For the last claim, observe that $\lim_n \mathrm{Tr}(G_n) = \lim_n \hom([1], G_n)$ exists as it is the limit of the homomorphism numbers of a single self-loop. By Corollary 5.3.4 and the proof of Proposition 5.2.6, the limit of the homomorphism numbers is the same for any sequence of graphs quotient-converging to M. By considering the sequence of graphs (5.12) derived from M, we conclude that $\lim_n \mathrm{Tr}(G_n) = \mathrm{Tr}(E) + \vartheta$, proving the last claim. □

5.5 Equipartition-Consistent Random Graph Models

Each grapheur M yields a sequence of random graphs $(\mathsf{G}_k[\mathsf{M}])$ via the construction (5.7). In Section 5.5, we show that these graphs are related to each other via equipartitions, and hence form an equipartition-consistent random graph model (Definition 5.1.8). We then prove that realizations of these random graphs $(\mathsf{G}_k[\mathsf{M}])$ converge back to the grapheur M in a suitable fashion. Finally, in Section 5.5 we prove Theorem 5.1.9, stating that any equipartition-consistent random graph model is a mixture of ones obtained from grapheurs as above.

Equipartition Consistency

We begin by arguing that $(\mathsf{G}_k[\mathsf{M}])$ is equipartition-consistent. Note that $d_{k,nk} \circ F_{nk,m} \overset{d}{=} F_{k,m}$ for any $k, n, m \in \mathbb{N}$ by [138, Lemma 4.13], since each $i \in [m]$ is mapped independently by $F_{nk,m}$ to one of the fibers of $d_{k,nk}$ with the same probability. Therefore, for any $G \in \Delta^{n\times n}$ the sequence of distributions $(\mathrm{Law}(\rho(F_{k,n})G_n))_k$ is indeed an equipartition-consistent model. If (G_m) is a sequence of finite graphs converging to a grapheur M, then $\rho(F_{k,m})G_m \to \mathsf{G}_k[\mathsf{M}]$ and $\rho(F_{nk,m})G_m \to \mathsf{G}_{nk}[\mathsf{M}]$ weakly, hence

$$\mathrm{Law}(\rho(d_{k,nk})\mathsf{G}_{nk}[\mathsf{M}]) = \lim_m \mathrm{Law}(\rho(d_{k,nk}\circ F_{nk,m})G_m) = \lim_m \mathrm{Law}(\rho(F_{k,m})G_m) = \mathrm{Law}(\mathsf{G}_k[\mathsf{M}]),$$

showing that $(\mathrm{Law}(\mathsf{G}_k[\mathsf{M}]))_k$ is equipartition-consistent for any grapheur M.

We now ask whether samples from an equipartition-consistent model $(\mathsf{G}_k[\mathsf{M}])$ obtained from a grapheur $\mathsf{M} \in \mathcal{M}$ converge back to M. The answer depends on how this sampling is performed.

Proposition 5.5.1. *For any grapheur* $\mathsf{M} \in \mathcal{M}$*, denote by* $(\mu_k = \mathrm{Law}(\mathsf{G}_k[\mathsf{M}]))_k$ *its associated equipartition-consistent random graph model.*

1. *Fix a grapheur* $\mathsf{M} \in \mathcal{M}$*, draw* $\mu \sim \mathsf{M}$*, and let* $\mathsf{Q}_k = (\mu(I_i^{(n)} \times I_j^{(n)}))_{i,j\in[k]}$ *for each* $k \in \mathbb{N}$*. Then* $\mathrm{Law}(\mathsf{Q}_k) = \mu_k$ *for all* $k \in \mathbb{N}$*, and* (Q_k) *is quotient-convergent to* M *almost surely.*

2. *Suppose* $\mathsf{M} = \delta_{(T_1,T_2)}$ *is the grapheur associated to a single edge. Draw* $\mathsf{R}_k \sim \mu_k$ *independently for different* k*. Then* $\limsup_k W_\square(\mathsf{M}_{\mathsf{R}_k}, \mathsf{M}) \geq 1/4$ *almost surely, so* (R_k) *does not converge to* M*.*

Proof. For the first claim, fix any $\epsilon > 0$ and let $N \in \mathbb{N}$ be sufficiently large so that $\sum_{i>N \text{ or } j>N} E_{i,j} + \sum_{i>N}(\sigma_i + \varsigma_i) \leq \epsilon$. Drawing $\mu \sim \mathsf{M}$ amounts to sampling locations $T_i \overset{iid}{\sim} \mathrm{Unif}([0,1])$ in (5.4), and these samples are distinct almost surely. Therefore, there exists $K \geq N \in \mathbb{N}$ such that $T_1, \ldots, T_N$ lie in distinct intervals $(I_i^{(k)})_{i\in[k]}$ for all $k \geq K$. Therefore, for all $k \geq K$ we can find a permutation $\pi_k \in \mathsf{S}_k$ such that $\pi_k \mathsf{Q}_k = \rho(\iota_{k,N})(E_{i,j} + \sigma_i/k + \varsigma_j/k)_{i,j\in[N]} + N_k + \theta\frac{1}{k^2}\mathbb{1}_k\mathbb{1}_k^\top + \vartheta\frac{1}{k}I_k$ where $\|N_k\|_1 \leq \epsilon$. Using the coupling $(\mathsf{M}'_{\mathsf{Q}_k}, \mathsf{M}')$ obtained by drawing random locations (T_i), setting M' to (5.4), and setting $\mathsf{M}'_{\mathsf{Q}_k} = \sum_{i,j=1}^N (\pi_k\mathsf{Q}_k)_{i,j}\delta_{(T_i,T_j)}$, we conclude that $\limsup_{k\to\infty} W_\square(\mathsf{M}_{\mathsf{Q}_k}, \mathsf{M}) \leq \epsilon$. Since $\epsilon > 0$ was arbitrary, we obtain the claimed convergence.

For the second claim, note that with probability $1/k$ we have $T_1, T_2 \in I_i^{(k)}$ for the same $i \in [k]$, in which case $\mathsf{R}_k = \mathrm{diag}(0, \ldots, 1, \ldots, 0)$ and $W_\square(\mathsf{M}_{\mathsf{R}_k}, \mathsf{M}) \geq 1/4$ as can be seen by setting $S = [0, 1/2]$ and $T = [1/2, 1]$ in (5.6). Since these events are independent and the sum of their probabilities $\sum_k k^{-1}$ diverges, Borel–Cantelli implies that $\limsup_k W_\square(\mathsf{M}_{\mathsf{R}_k}, \mathsf{M}) \geq 1/4$ almost surely. □

In words, the quotients must be sampled in a manner that is related across different sizes for the samples to converge to the underlying grapheur; if we instead sample random quotients independently for each size, the resulting sequence almost surely does not converge to the underlying grapheur. In contrast, if we sample growing-sized random graphs from a graphon independently for different sizes, then the resulting sequence does converge back to the graphon almost surely as a consequence of Szemerédi's regularity lemma [146, Lemma 11.8]. Proposition 5.5.1(2) thus implies that the analog of Theorem 5.1.7 fails for random quotients. More precisely, for any rate of convergence $(R(k) \in \mathbb{R}_{\geq 0})_{n\in\mathbb{N}}$ with $R(k) \to 0$, we have

$\sum_{k\in\mathbb{N}} \mathbb{P}[W_\square(\mathsf{M}, \mathsf{G}_k[\mathsf{M}]) > R(k)] = \infty$ by Borel–Cantelli. We leave the precise analysis of convergence rates of random quotients for future work.

Proof of Theorem 5.1.9

Our goal in this section is to prove Theorem 5.1.9, showing that every equipartition-consistent random graph model arises from a unique random exchangeable measure. We deduce the existence of the desired random measure from results on inverse limits of random histograms [169], and proving that this random measure is exchangeable proceeds via standard approximation arguments. To apply the results of [169], we require the following lemma on the sequence of means of an equipartition-consistent model.

Lemma 5.5.2. *For any equipartition-consistent random graph model (μ_k), there exists $\theta \in [0,1]$ satisfying $\mathbb{E}_{X\sim\mu_k} X = \theta \frac{1}{k^2} \mathbb{1}_k \mathbb{1}_k^\top + (1-\theta)\frac{1}{k} I_k$ for all $k \in \mathbb{N}$.*

Proof. Note that $M_n = \mathbb{E}_{X\sim\mu_n} X \in \mathbb{S}^n$ is a deterministic matrix invariant under simultaneous permutations of its rows and columns, because μ_n is exchangeable. Any such matrix has the same on-diagonal and off-diagonal entries, hence there exist $\theta_n, \vartheta_n \in \mathbb{R}$ satisfying $M_n = \theta_n \frac{1}{n} \mathbb{1}_{n\times n} + \vartheta_n \frac{1}{n} I_n$. Since $\mathbb{1}^\top X \mathbb{1} = 1$ almost surely, we similarly have $\mathbb{1}^\top M_n \mathbb{1} = \theta_n + \vartheta_n = 1$. Also, since $X \geq 0$ almost surely, we must have $\theta_n \in [0,1]$. Finally, since $\rho(d_{n,N}) X_N \overset{d}{=} X_n$ whenever $n|N$, we must have

$$M_n = \rho(d_{n,N}) M_N = \theta_N \frac{1}{n^2} \mathbb{1}_{n\times n} + (1-\theta_N)\frac{1}{n} I_n.$$

Since $\frac{1}{n^2}\mathbb{1}_{n\times n}$ and $\frac{1}{n} I_n$ are linearly independent for all $n \geq 2$, we must have $\theta_N = \theta_n$ whenever $n|N$, and therefore for all $n, N \in \mathbb{N}$. This proves the claim. $\square$

Combining the above lemma with [169], we are ready to prove Theorem 5.1.9.

Proof (Theorem 5.1.9). Uniqueness is clear since squares of the form $I_i^{(k)} \times I_j^{(k)}$ generate the Borel σ-algebra on $[0,1]^2$. To prove existence we appeal to [169, Thm. 1.1], which states that there exists a random probability measure M on $[0,1]^2$ (not guaranteed to be exchangeable) satisfying $\mathsf{M}(I_i^{(k)} \times I_j^{(k)}) \sim \mu_k$ for all k if there is a determinisitic measure $\mu \in \mathcal{P}([0,1]^2)$ satisfying $(\mu(I_i^{(k)} \times I_j^{(k)}))_{i,j\in[k]} = \mathbb{E}_{X\sim\mu_k} X$ for all $k \in \mathbb{N}$. Such a deterministic measure exists by Lemma 5.5.2, since

$$((\theta\lambda^2 + (1-\theta)\lambda_D)(I_i^{(k)} \times I_j^{(k)}))_{i,j\in[k]} = \theta \frac{1}{k^2} \mathbb{1}_{k\times k} + (1-\theta)\frac{1}{k} I_k.$$

We proceed to prove that the above M is exchangeable. To do so, it suffices to prove that for any continuous function $f\colon [0,1)^2 \to \mathbb{R}$ and any measure-preserving bijection σ, we have $\mathbb{E}_{\mathsf{M}} f \overset{d}{=} \mathbb{E}_{\mathsf{M}} f \circ (\sigma, \sigma)$. By [115, Prop. 9.1], it further suffices to prove this equality for $\sigma \in \mathsf{S}_m$, a permutation of the m intervals $I_1^{(m)}, \ldots, I_m^{(m)}$ acting by $\sigma((i-1)/m + x) = (\sigma(i)-1)/m + x$ for $x \in [0, 1/m)$ for $i \in [m]$.

For any continuous function $f\colon [0,1]^2 \to \mathbb{R}$, let $f_n = \sum_{i,j=1}^n f_{i,j}^{(n)} \mathbb{1}_{I_i^{(n)} \times I_j^{(n)}}$ where $f_{i,j} = f(\frac{i}{n} - \frac{1}{2n}, \frac{j}{n} - \frac{1}{2n})$ is the value of f on the centers of the squares $I_i^{(n)} \times I_j^{(n)}$, whose diameter in ℓ_∞ is $1/2n$. Since f is continuous on a compact set, it is uniformly continuous, hence for any $\epsilon > 0$ we have $|f(x) - f_n(x)| \le \epsilon$ for all $x \in [0,1)^2$ and all large n. Note that

$$\mathbb{E}_{\mathsf{M}} f_n = \sum_{i,j=1}^n f_{i,j}^{(n)} \mathcal{M}(I_i^{(n)} \times I_j^{(n)}) = \sum_{i,j=1}^n f_{i,j}^{(n)} (\mathsf{G}_n[\mathsf{M}])_{i,j}.$$

Therefore, for any $m, k \in \mathbb{N}$ and $\sigma_m \in \mathsf{S}_m$, we view σ_m as an interval permutation and as a permutation on $[mk]$ letters permuting consecutive intervals of length k, in which case we get

$$\mathbb{E}_{\mathsf{M}} f_{mk} \circ (\sigma_m, \sigma_m) = \sum_{i,j=1}^m f_{i,j}^{(mk)} (G_{mk}^{(\mathsf{M})})_{\sigma^{-1}(i), \sigma^{-1}(j)} \overset{d}{=} \sum_{i,j=1}^m f_{i,j}^{(mk)} (G_{mk}^{(\mathsf{M})})_{i,j} = \mathbb{E}_{\mathsf{M}} f_{mk}.$$

Since this holds for all $k \in \mathbb{N}$, we have

$$\begin{aligned} W_1(\mathbb{E}_{\mathsf{M}} f, \mathbb{E}_{\mathsf{M}} f \circ (\sigma_m, \sigma_m)) \le & W_1(\mathbb{E}_{\mathsf{M}} f, \mathbb{E}_{\mathsf{M}} f_{mk}) + W_1(\mathbb{E}_{\mathsf{M}} f_{mk}, \mathbb{E}_{\mathsf{M}} f_{mk} \circ (\sigma_m, \sigma_m)) \\ & + W_1(\mathbb{E}_{\mathsf{M}} f_{mk} \circ (\sigma_m, \sigma_m), \mathbb{E}_{\mathsf{M}} f \circ (\sigma_m, \sigma_m)) \le 2\epsilon, \end{aligned}$$

for all large k, where the last inequality follows since M is supported on $[0,1)^2 = \bigcup_{i,j} I_i^{(n)} \times I_j^{(n)}$. Since $\epsilon > 0$ was arbitrary, this proves the exchangeability of M. $\square$

5.6 Conclusions and Future Directions

We have introduced a new notion of limits of growing graphs based on convergence of their fixed-size random quotients. We showed that this notion of convergence can be equivalently characterized in terms of convergence of graph parameters such as homomorphism numbers and quotient densities. We characterized limits of quotient-convergent graph sequences via grapheurs, which are certain random exchangeable measures on $[0,1]^2$. The limiting grapheur of a quotient-convergent sequence of graphs describes the asymptotic distribution of edge weight in the graphs. We then presented another equivalent view of quotient convergence via edge sampling, allowing us to prove an edge-based analog of Szemerédi's regularity lemma and to test properties of arbitrarily-large graphs by sampling edges from them. The number of edges needed is indpendent of the size of the graph, only depending on the desired accuracy and the Lipschitz constant of the graph parameter in question. Finally, we studied equipartition-consistent random graph models, showing that any such model corresponds to the sequence of random quotients associated to a mixture of grapheurs. We conclude by mentioning several questions suggested by our work.

(Optimal Szemerédi rate) Is the rate of $O(1/\sqrt{n})$ for approximating a grapheur by a graph on n edges optimal?

(Extending graphon duality) Is Conjecture 5.3.11 on the bi-continuity of our duality pairing true? Can the duality with graphons in Section 5.3 be extended to limits of non-simple graphs? Can we relate the cut metric and our $W_{\square}$-metric via this duality?

(Inequalities in quotient densities) Corollary 5.3.5 shows that valid inequalities in quotient densities (or equivalently, homomorphism numbers) that hold for graphs of all sizes correspond to inequalities over grapheurs. Can we exploit the structure of the space of grapheurs to prove such inequalities, in analogy with the use of graphon theory to prove homomorphism density inequalities?

Chapter 6

ANY-DIMENSIONAL LEARNING USING SAMPLING

This chapter is based on [137].

6.1 Introduction

We begin by formally stating the any-dimensional generalization and sketching questions.

Formalizing Any-Dimensional Learning

To frame our discussion concretely, let $(\mathbb{V}_n)$ be a sequence of vector spaces, with the index n specifying object size, e.g., vectors of length n, adjacency matrices of graphs on n vertices, and so on. Although our framework encompasses these and many other applications, we consider two running examples:

- Sequences of features or tokens, represented as n points in $\mathbb{R}^d$. Here we set $\mathbb{V}_n = \mathbb{R}^{d\times n}$ for a fixed $d \in \mathbb{N}$ and growing n.
- Weighted and directed graphs on n vertices, represented by their adjacency matrices. Here we set $\mathbb{V}_n = \mathbb{R}^{n\times n}$.

In many applications, it is often the case that the objects of interest belong to some proper subset $\Omega_n \subseteq \mathbb{V}_n$; for instance, unweighted graphs are precisely those whose adjacency matrix belongs to $\Omega_n = \{0, 1\}^{n\times n}$. In the sequel, we will see that various structural properties of this subset, and indeed of the collection $\bigsqcup_n \Omega_n \subseteq \bigsqcup_n \mathbb{V}_n$ of subsets containing inputs of all possible sizes, play a central role in our framework.

In order to formulate the any-dimensional generalization problem precisely, let $\Omega_{\leq n} = \bigsqcup_{i\leq n} \Omega_i$ be the set of objects of bounded size. For two functions $f, g\colon \bigsqcup_n \Omega_n \to \mathbb{R}$ defined on objects of all sizes, i.e., any-dimensional functions, define $\mathrm{e}_n(f, g) = \sup_{x\in\Omega_{\leq n}} |f(x) - g(x)|$ and $\mathrm{e}_\infty(f, g) = \sup_n \mathrm{e}_n(f, g)$. We then investigate the following question:

Any-dimensional generalization: Under what conditions on $\bigsqcup_n \Omega_n$ and on any-dimensional functions $f, g : \bigsqcup_n \Omega_n \to \mathbb{R}$ do we have that:

$$\lim_{n\to\infty} \mathrm{e}_n(f,g) = \mathrm{e}_\infty(f,g)?$$

At what rate does this limit converge? Is there a hypothesis class of any-dimensional functions over which the convergence rate is uniform?

If $\mathrm{e}_n(f,g)$ converges to $\mathrm{e}_\infty(f,g)$ as $n \to \infty$, then we can control the error $\mathrm{e}_\infty(f,\widehat{f})$ between a function f and its estimate $\widehat{f}$ on objects of arbitrarily-large size by the error $\mathrm{e}_n(f,\widehat{f})$ on objects of size at most n. We would thereby reduce the problem of approximating a target function on infinitely-many input sizes to approximating it on a fixed, sufficiently-large size, which can be done using classical approximation and learning theory. If we further have explicit and uniform rates of convergence over some hypothesis class, then we can obtain *a priori* bounds on the input size n needed to accurately approximate the target function, and precise bounds on $\mathrm{e}_\infty(f,\widehat{f})$ as a function of $\mathrm{e}_n(f,\widehat{f})$.

We have stated any-dimensional generalization in terms of uniform error over inputs of each size. If instead we seek to quantify generalization in terms of data distributions, we must further understand what natural any-dimensional data distributions are. Merely considering distributions μ on $\bigsqcup_n \Omega_n$ is not sufficient, since any such distribution will be concentrated on bounded-dimensional inputs because $\sum_n \mu(\Omega_n) = 1$.

The second question we consider involves approximating the value of an any-dimensional function on a large input by its value on a small sketch of this input.

Any-dimensional sketching: Under what conditions on $\bigsqcup_n \Omega_n$ and on an any-dimensional function $f : \bigsqcup_n \Omega_n \to \mathbb{R}$ can we find a (potentially random) map $\mathcal{S}_k : \bigsqcup_n \Omega_n \to \Omega_k$ satisfying

$$\lim_{k\to\infty} \mathrm{e}_\infty(f, f\circ\mathcal{S}_k) = 0 \quad \text{almost surely?}$$

At what rate does this limit converge? Is there a hypothesis class of any-dimensional functions over which the convergence rate is uniform?

Already at this stage, it is clear that any attempt to address the any-dimensional generalization question would require some notion of similarity between objects of different sizes in $\bigsqcup_n \Omega_n$ respected by the any-dimensional functions under consideration. Indeed, if a model performs well on large inputs of unseen size, then these inputs must in some sense be similar to smaller inputs from the training set. Similarly, if a function takes on similar values on a low-dimensional sketch of a high-dimensional object, then the object and its sketch must be similar to each other. More precisely, we need a distance between elements of $\bigsqcup_n \Omega_n$ that allows us to quantitatively compare objects of different sizes, and consider any-dimensional functions that are continuous in this distance. In particular, this distance will play a central role in addressing the above questions on any-dimensional generalization.

Motivated by the above discussion, we consider a metric $d\colon \bigsqcup_n \Omega_n \times \bigsqcup_n \Omega_n \to [0,\infty)$, with respect to which we have a distance between a point $x \in \Omega_\infty$ and a subset $S \subseteq \Omega_\infty$ given by $\mathrm{dist}(x,S) = \inf_{y\in S} d(x,y)$. We shall see that any-dimensional generalization and sketching are intimately related to the following question pertaining to this metric space.

Any-dimensional approximation: Under what conditions on $(\bigsqcup_n \Omega_n, d)$ do we have that:

$$\lim_{n\to\infty} \sup_{x\in\Omega_\infty} \mathrm{dist}(x, \Omega_{\leq n}) = 0? \tag{6.1}$$

At what rate does this limit approach zero?

The any-dimensional approximation property states that objects of any size can be approximated by bounded-size objects. In particular, the order of the limit and the supremum highlights the any-dimensional nature of the question, as approximability is required to hold uniformly over all objects of arbitrary size.

We formally relate the above three questions, and show that they are all equivalent to the compactness of $(\bigsqcup_n \Omega_n, d)$ in a precise sense, in Section 6.2.

A Sampling Based Approach for Comparing Objects of Different Sizes

To address the preceding questions in any-dimensional learning, we describe a sampling-based approach to compare objects of different sizes. Specifically, we fix random sampling maps $(\mathsf{S}_k\colon \bigsqcup_n \mathbb{V}_n \to \mathbb{V}_k)_{k\in\mathbb{N}}$ and compare any two inputs $x, y \in \bigsqcup_n \mathbb{V}_n$ by comparing the distributions of their samples $\mathrm{Law}(\mathsf{S}_k x)$ and $\mathrm{Law}(\mathsf{S}_k y)$ of each size k. In particular, we define the *sampling metric* between $x, y \in \bigsqcup_n \mathbb{V}_n$ by:

$$d_{\mathrm{samp}}(x,y) = \sum_{k\geq 1} 2^{-k} W_1(\mathrm{Law}(\mathsf{S}_k(x)), \mathrm{Law}(\mathsf{S}_k(y))), \tag{6.2}$$

where W_1 denotes the Wasserstein-1 distance between two probability measures with respect to some norm on $\mathbb{V}_k$ satisfying $\sup_k \mathbb{E}\|\mathsf{S}_k(x)\| < \infty$ for each x. A sequence of inputs $(x_n)_n$ converges with respect to the sampling metric (6.2) if the sequence of fixed-dimensional samples $(\mathsf{S}_k(x_n))_n$ converges weakly for each size k.

The above metric and corresponding notion of convergence has been previously used in conjungtion with specific sampling maps to compare graphs of different sizes and to take their limits, see [146]. In this chapter, we apply these notions to more general notions of sampling. We proceed to give a few examples of the sampling maps we consider.

Example 6.1.1 (Sampling columns)**.** *Suppose $\mathbb{V}_n = \mathbb{R}^{d\times n}$ for a fixed $d \in \mathbb{N}$. We then let the sampling map S_k sample k columns from $x = [x_1, \ldots, x_n] \in \mathbb{V}_n$ uniformly at random with replacement, i.e., we sample indices $J_1, \ldots, J_k$ iid uniformly from $[n]$ and set $\mathsf{S}_k x = [x_{J_1}, \ldots, x_{J_k}] \in \mathbb{V}_k$. We then consider the sampling map (6.2) using the ℓ_∞ norm to define the W_1 distances.*

In this case, we have $\mathrm{Law}(\mathsf{S}_k x) = \mathrm{Law}(\mathsf{S}_1 x)^{\otimes k}$ and if $\Omega_n = \Theta^n$ for some compact $\Theta \subseteq \mathbb{R}^d$ then a sequence $(x_n) \subseteq \bigsqcup_n \Omega_n$ converges in d_{samp} precisely when $\mathrm{Law}(\mathsf{S}_1 x_n)$ converges weakly.

Example 6.1.2 (Sampling vertices)**.** *Suppose $\mathbb{V}_n = \mathbb{R}^{n\times n}$, viewed as the space of weighted and directed graphs. Let S_k be the map that samples k vertices with replacement and extracts the corresponding induced subgraph, i.e., for $X \in \mathbb{V}_n$, we have $(\mathsf{S}_k X)_{i,j} = X_{J_i,J_j}$ for $J_1, \ldots, J_k \overset{iid}{\sim} \mathrm{Unif}([n])$. Once again, we consider the sampling map (6.2) with the entrywise ℓ_∞ norm defining the W_1 distances.*

When $\Omega_n \subseteq \Theta^{n\times n}$, convergence in sampling metric is precisely the dense graph limits studied in [1, 20, 60, 147].

Example 6.1.3 (Random binning, hashing)**.** *Suppose $\mathbb{V}_n = \mathbb{R}^{d\times n}$ and consider the sampling map S_k that randomly assigns the n columns of $x \in \mathbb{V}_n$ into k bins and sums the columns in each bin, i.e., we sample $J_1, \ldots, J_n$ iid uniformly from $[k]$ and set the ith column of $\mathsf{S}_k x \in \mathbb{V}_k$ to $(\mathsf{S}_k x)_i = \sum_{\ell : J_\ell = i} x_\ell$ for $i \in [k]$. We then consider the sampling metric (6.2) with the ℓ_2 norm.*

This random binning, also called hashing, has found uses in machine learning as a dimensionality reduction method [51, 203] and in information theory as part of source and channel coding protocols [232].

Example 6.1.4 (Species sampling, random partitions)**.** *Suppose* $\mathbb{V}_n = \mathbb{R}^n$ *and* $x \in \Delta^{n-1}$*. Such* x *defines a random partition of* $[k]$ *as follows. We view* x *as a distribution on indices in* $[n]$ *and sample* k *indices* $I_1, \dots, I_k \in [n]$ *iid from this distribution. These define a random partition of* $[k]$ *in which* $i, j \in [k]$ *belong to the same block if* $I_i = I_j$*, and we set the entries of* $\mathsf{S}_k(x) \in \Delta^{k-1}$ *to be the fraction of indices in* $[k]$ *belonging to each block of the partition. More explicitly, we denote the* $\ell \leq k$ *distinct indices sampled in this way by* $\{T_1, \dots, T_\ell\} = \{I_1, \dots, I_k\}$*, and set* $\mathsf{S}_k(x)_j = |\{i : I_i = T_j\}|/k$*. This sampling map extends to a general* $x \in \mathbb{R}^n$ *by homogeneity, see Example 6.2.7.*

This notion of sampling, and the resulting random partition models, has been studied extensively in [122, 123, 176], see [190] for example.

Although the sampling metric notion has appeared extensively in the graph limits literature [146] with the maps given by sampling vertices with replacement, we demonstrate in this chapter that the underlying idea itself is much more broadly relevant to comparing objects of different sizes in many other contexts. In particular, we go well beyond the preceding specific illustrations by describing next a combinatorial perspective on sampling based on random maps between finite sets. This viewpoint yields two broad classes of sampling maps that represent a significant generalization of Examples 6.1.1-6.1.4. As we shall see, the sampling perspective is useful both analytically and methodologically, as it yields explicit rates for any-dimensional learning and also provides an associated algorithmic framework.

Generalizing Sampling

We generalize the sampling maps used in Examples 6.1.1-6.1.4 and unify their analysis by using maps between finite sets. To motivate our generalization, we observe that any map $f\colon [k] \to [n]$ defines a linear map $\rho(f)\colon \mathbb{R}^n \to \mathbb{R}^k$ extracting entries specified by f. Formally for any $x \in \mathbb{R}^n$ the entries of $\rho(f)x \in \mathbb{R}^k$ are given by

$$(\rho(f)x)_i = x_{f(i)}. \tag{6.3}$$

Moreover, if $F_{n,k}\colon [k] \to [n]$ is a uniformly random such map, so that $F_{n,k}(1), \dots, F_{n,k}(k) \overset{iid}{\sim} \mathrm{Unif}([n])$, then $\rho(f_{n,k})x$ is precisely the vector obtained by sampling k coordinates from x uniformly at random with replacement as in Example 6.1.1. Likewise, any map $f\colon [k] \to [n]$ acts on tuples $\alpha(f)\colon [k]^2 \to [n]^2$ by $\alpha(f)(i,j) = (f(i), f(j))$, and hence acts linearly on matrices $x \in \mathbb{V}_n$ indexed by these tuples via

$(\rho(f)x)_{i,j} = x_{f(i),f(j)}$. Once again, applying a uniformly random map $\rho(F_{n,k})X$ recovers sampling vertices as in Example 6.1.2. More generally, we consider actions α of maps between finite sets on indices of growing-sized vectors, and define sampling with replacement R_k in this setting by applying a uniformly random map

$$\mathsf{R}_k x = \rho(F_{n,k})x \quad \text{if } x \in \mathbb{V}_n, \quad \text{where } F_{n,k} \colon [k] \to [n] \text{ is uniformly random.} \tag{6.4}$$

Observe that the distribution $\mathrm{Law}(\mathsf{R}_k x)$ is unchanged if we duplicate or permute the entries of x, appropriately defined. For example, if $x \in \mathbb{R}^n$ and we permute the entries of x or duplicate them k times (so $x \mapsto x \otimes \mathbb{1}_k$), or if $x \in \mathbb{R}^{n\times n}$ and we replace each entry of x by a $k \times k$ block (so $x \mapsto x \otimes \mathbb{1}_k \mathbb{1}_k^\top$). Thus, using sampling with replacement to compare inputs may be appropriate if the above permutation and duplication yields equivalent inputs in the context of the application at hand.

Our generalization of random binning is dual to sampling with replacement in the following precise sense. Any map $f \colon [n] \to [k]$ defines another linear map on vectors that bins their coordinates, given by $\beta(f) \colon \mathbb{R}^n \to \mathbb{R}^k$ where

$$(\beta(f)x)_i = \sum_{j \in f^{-1}(i)} x_j. \tag{6.5}$$

Note that $\beta(f) = \rho(f)^\top$, so the sampling action (6.3) and binning action (6.5) are dual to each other. By letting f act on tuples $\alpha(f) \colon [n]^2 \to [k]^2$ as before, we obtain a binning action on graphs that corresponds to taking quotient graphs, see [140]. More generally, we again consider actions α of maps between finite sets on indices of vectors, and define random binning B_k by applying a uniformly random map

$$\mathsf{B}_k x = \beta(F_{k,n})x \quad \text{if } x \in \mathbb{V}_n, \quad \text{where } F_{k,n} \colon [n] \to [k] \text{ is uniformly random.} \tag{6.6}$$

Observe that the distribution $\mathrm{Law}(\mathsf{B}_k x)$ is unchanged if we zero-pad or permute the entries of x, appropriately defined. For example, if $x \in \mathbb{R}^n$ and we append k zeros to x, or if $x \in \mathbb{R}^{n\times n}$ and we append k zero rows and columns to x. Thus, using random binning to compare inputs may be appropriate if the above permutation and zero-padding yields equivalent inputs in the context of the application at hand.

Rates for Sampling with Replacement

We derive rates for any-dimensional sketching and generalization for the sampling metric (6.2) defined with respect to general sampling with replacement maps

($\mathsf{R}_k\colon \bigsqcup_n \mathbb{V}_n \to \mathbb{V}_k$). To state our results, we fix a sequence of compact subsets ($\Omega_n \subseteq \mathbb{V}_n$) such that $\mathsf{R}_k \bigsqcup_n \Omega_n \subseteq \Omega_k$ almost surely for all k, i.e., the subsets Ω_n are closed under sampling with replacement. For example, the hypercubes $\Omega_n = [-1,1]^n \subseteq \mathbb{R}^n$ and the collection of unweighted graphs $\Omega_n = \{0,1\}^{n\times n} \subseteq \mathbb{R}^{n\times n}$ are closed under sampling with replacement.

Theorem 6.1.5 (Informal, see Theorem 6.3.5)**.** *We have the following rates for sampling with replacement* ($\mathsf{R}_k\colon \bigsqcup_n \Omega_n \to \Omega_k$) *acting on compact sets* (Ω_n) *closed under sampling.*

(Any-dimensional approximation) *For any object $x \in \bigsqcup_n \Omega_n$ of arbitrary size, we have that:*

$$\mathbb{E} d_{\mathrm{samp}}(x, \mathsf{R}_n x) \le c_1 \exp\left[-(c_2^{-1}\log n)^{\frac{1}{1+D}}\right],$$

where the expectation is with respect to the randomness in R_n. Moreover, the distance $d_{\mathrm{samp}}(x, \mathsf{S}_n x)$ is c_3/n-subgaussian.

(Any-dimensional sketching) *For any function $f\colon \bigsqcup_n \Omega_n \to \mathbb{R}$ that is L-Lipschitz in d_{samp}, and any $x \in \bigsqcup_n \Omega_n$ with probability at least $1 - e^{-2\epsilon^2}$ we have*

$$|f(x) - f(\mathsf{R}_n x)| \le Lc_1 \exp\left[-(c_2^{-1}\log n)^{\frac{1}{1+D}}\right] + \epsilon\sqrt{\frac{c_3}{n}}.$$

(Any-dimensional generalization) *For any other L-Lipschitz function $\widehat{f}$, we have that:*

$$\mathrm{e}_\infty(f, \widehat{f}) \le \mathrm{e}_n(f, \widehat{f}) + 2Lc_1 \exp\left[-(c_2^{-1}\log n)^{\frac{1}{1+D}}\right].$$

The constants $c_1, c_2, c_3 > 0$ are explicit and depend only on the collection (Ω_n), and the parameter D is a 'degree' that quantifies the complexity of the action underlying the sampling maps (B_k).

In words, the any-dimensional approximation result states that an object of *any* size can be approximated by an object of size n, with the quality of the approximation depending only on n. Moreover, the approximation can be computed using sampling with replacement. As a consequence, we are able to approximate the value of a Lipschitz function on an arbitrarily-large input by first computing a random sample whose size only depends on the desired accuracy. For any-dimensional generalization, our results show that if $\widehat{f} \approx f$ on inputs of size n, then $\widehat{f} \approx f$ on inputs of all sizes up to a slack that decays to zero as n increases. In other words, approximating a target function on sufficiently-large input sizes guarantees a good approximation on inputs of all sizes. Theorem 6.1.5 also reduces the analysis of any-dimensional generalization error to generalization error on a finite-dimensional compact set, a classical problem (see [216] for example).

The rates in Theorem 6.1.5 are clearly quite slow. On the other hand, these rates are uniform in the sense that they hold for objects of any size. Moreover, these slow uniform rates are to be expected in general. For example, when Ω_n is the collection of $n \times n$ adjacency matrices of undirected graphs with no self-loops and edge weights in $[-1, 1]$, a rate of $\exp[-\frac{1}{2} \log\log n]$ was shown in [20, Thm 2.9] for the closely related cut metric as a consequence of Szemerédi's regularity lemma (see [146, Exer. 10.33] for the connection between the cut and sampling metrics). Nevertheless, we can substantially imporve the above uniform rates using additional structure in the function f. In particular, many any-dimensional functions of interest depend on their arbitrarily-large inputs only via a fixed-dimensional random sample, as described in Section 6.3. For example, we often have $f(x) = \mathbb{E} g(\mathsf{R}_k x)$ for a fixed function $g\colon \Omega_k \to \mathbb{R}$

We give two illustrations here of some of these improved rates in specific applications. The first pertains to polynomials that are unchanged by suitably-defined duplication of the entries of their inputs. Specifically, a function $p\colon \bigsqcup_n \mathbb{V}_n \to \mathbb{R}$ is a polynomial of degree d if all the restrictions $p|_{\mathbb{V}_n}$ are such polynomials.

Corollary 6.1.6 (Informal; see Corollary 6.3.10)**.** *Suppose that (Ω_n) is a sequence of compact sets closed under sampling, and let $p\colon \bigsqcup_n \Omega_n \to \mathbb{R}$ be a polynomial of degree d that is symmetric and unchanged by duplication. Then there exists $c_1, c_2, c_3 > 0$ only depending on (Ω_n) and p such that with probability at least $1 - e^{-2\epsilon^2}$, we have for any $x \in \bigsqcup_n \Omega_n$ that*

$$|p(x) - p(\mathsf{R}_n x)| \leq \frac{c_1 d^2}{n} + \frac{c_2 d\epsilon}{\sqrt{n}},$$

and if $\widehat{p}$ is another such polynomial,

$$\mathrm{err}_\infty(p, \widehat{p}) \leq \mathrm{err}_n(p, \widehat{p}) + \frac{c_3 d^2}{n}.$$

Examples of polynomials satisfying the above result are polynomials in moments of measures, graph homomorphism densities, and polynomial graph neural networks, see Corollary 6.3.10 and the discussion following it. The proof uses a sampling-based representation for polynomials from [138]. We remark that the above $O(n^{-1})$ generalization rates are a substantial improvement over previous rates, as the latter were proved for larger function classes and exploit less structure. For example, the best rates using the framework of [142] is $O(n^{-1/d})$ for moment polynomials and $O((\log n)^{-1/2})$ for graph homomorphism densities, see [142, §C.3].

The second consequence we highlight is to sketching and generalization for permutation-invariant transformers.

Corollary 6.1.7 (Permutation-Invariant Transformers)**.** *Suppose $\mathbb{V}_n = \mathbb{R}^{d\times n}$ with $d > 2$, and $\Omega_n = \Theta^n$ for compact $\Theta \subseteq [-r, r]^d$. For matrices $Q, K \in \mathbb{R}^{k\times d}$ and $V \in \mathbb{R}^{d\times d}$, define the self-attention mapping $A\colon \mathbb{V}_n \to \mathbb{V}_n$ by*

$$A(x_1,\ldots,x_n)_i = x_i + \frac{1}{Z_i}\sum_{j=1}^n \exp(\langle Qx_i, Kx_j\rangle)Vx_j, \quad \text{where } Z_i = \sum_{j=1}^n \exp(\langle Qx_i, Kx_j\rangle). \tag{6.7}$$

Let $\phi\colon \mathbb{R}^d \to \mathbb{R}^d$ be L_ϕ-Lipschitz, and let it act column-wise on $\mathbb{V}_n$. Finally, let $P\colon \mathbb{V}_n \to \mathbb{V}_1$ be the mean pooling map $P(x_1,\ldots,x_n) = \frac{1}{n}\sum_{i=1}^n x_i$. Define a depth-k transformer to be the mapping $T\colon \mathbb{V}_n \to \mathbb{V}_1$ defined by the composition

$$T = P \circ (\phi \circ A)^{\circ k},$$

Then there exists an L-Lipschitz continuous map $\bar{T}\colon \mathcal{P}(\Theta) \to \mathbb{R}^d$ satisfying $T(x) = \bar{T}(\mathrm{Law}(\mathsf{S}_1 x))$. Consequently, with probability at least $1 - e^{-2\epsilon^2}$ we have

$$|T(x) - T(\mathsf{S}_n x)| \leq Lr\left[\frac{\epsilon}{\sqrt{n}} + \frac{208}{n^{1/d}}\right]. \tag{6.8}$$

For any other L'-Lipschitz map $\bar{F}\colon \mathcal{P}(\Theta) \to \mathbb{R}^d$, we define $F(x) = \bar{F}(\mathrm{Law}(\mathsf{S}_1 x))$ and obtain

$$\mathrm{err}_\infty(T, F) \leq \mathrm{err}_n(T, F) + \frac{208r(L + L')}{n^{1/d}},$$

where $\mathrm{err}_n(T, F) = \sup_{x\in\Theta} \|T(x) - F(x)\|_\infty$ and similarly for err_∞.

The proof is based on the measure-theoretic in-context mapping of [82], and is given in Section 6.3. We remark that evaluating $T(x)$ exactly for $x \in \mathbb{V}_N$ takes $O(N^2)$ operations due to (6.7), but evaluating it to a desired accuracy $\delta > 0$ can be done with high probability in $O(\delta^{-2d})$ time (independent of the length N of the input!) after sampling $n = O(\delta^{-d})$ columns from x uniformly at random and using (6.8).

Rates for Random Binning and Species Sampling

Next, we turn to proving rates for any-dimensional sketching and generalization for the sampling metric (6.2) defined with respect to general random binning $(\mathsf{B}_k\colon \bigsqcup_n \mathbb{V}_n \to \mathbb{V}_k)$. Once again, we fix a sequence of compact subsets $(\Omega_n \subseteq \mathbb{V}_n)$ closed under sampling, this time under random binning so $\mathsf{B}_k \bigsqcup_n \Omega_n \subseteq \Omega_k$ almost surely for all k. Examples include the simplex $\Omega_n = \Delta^{n-1}$ and entrywise ℓ_1 balls.

We would like to use random binning also as a sketching map, similarly to Theorem 6.1.5 where sampling with replacement was both defining our sampling metric and acting as our sketching map. Unfortunately, the distance $d_{\mathrm{samp}}(x, \mathsf{B}_n x)$ does not concentrate as well. In particular, we show in Example 6.4.1 that $d_{\mathrm{samp}}(x, \mathsf{B}_n x)$ is not subgaussian already for $x \in \mathbb{R}^2$. We therefore seek a different sketching map that with high probability yields low-dimensional approximations of arbitrary-dimensional inputs.

Such a sketching map is furnished by generalizing the species sampling map from Example 6.1.4, which turns out to be compatible with random binning.

Theorem 6.1.8 (Informal, see Theorem 6.4.3)**.** *Consider random binning maps $(\mathsf{B}_k\colon \bigsqcup_n \Omega_n \to \Omega_k)$ acting on compact sets closed under binning and sampling metric defined using random binning. There is a generalization of species sampling (E_k) satisfying the following properties.*

(Equivalence of binning and species sampling) *The sampling maps (B_k) and (E_k) define the same topology via (6.2).*

(Any-dimensional approximation) *We have $\mathbb{E} d_{\mathrm{samp}}(x, \mathsf{E}_n x) \leq \frac{c_1}{\sqrt{n}}$. Moreover, the distance $d_{\mathrm{samp}}(x, \mathsf{E}_n x)$ is c_2/n-subgaussian. The constants $c_1, c_2 > 0$ are explicit and only depend on (Ω_n).*

(Any-dimensional sketching) *For any function $f\colon \bigsqcup_n \Omega_n \to \mathbb{R}$ that is L-Lipschitz in d_{samp} and any $x \in \bigsqcup_n \Omega_n$, we have with probability at least $1 - e^{-2\epsilon^2}$ that*

$$|f(x) - f(\mathsf{E}_n x)| \leq \frac{L(c_1 + c_2\epsilon)}{\sqrt{n}}.$$

(Any-dimensional generalization) *If $\widehat{f}\colon \bigsqcup_n \Omega_n \to \mathbb{R}$ is also L-Lipschitz in d_{samp}, then*

$$\mathrm{e}_\infty(f, \widehat{f}) \leq \mathrm{e}_n(f, \widehat{f}) + \frac{2Lc_1}{\sqrt{n}}.$$

In contrast to Theorem 6.1.5, we always get $n^{-1/2}$ rates, regardless of the complexity of the action of random binning. Furthermore, the above rates apply with the same constants to functions exhibiting additional structure.

The following are some of the function classes satisfying the hypotheses of Theorem 6.4.3.

(Polynomials) In analogy to Corollary 6.1.6, there is a large class of polynomial functions that are continuous with respect to random binning. Here, this holds for any polynomial $p\colon \bigsqcup_n \mathbb{V}_n \to \mathbb{R}$ that is symmetric and unchanged by zero-padding, appropriately defined. Such polynomials include multi-symmetric functions on sequences of vectors and homomorphism numbers for graphs of arbitrary size.

(Symmetric Neural Networks) Many permutation-invariant neural network architectures are continuous with respect to random binning, including DeepSets [233], PointNet [182], and graph neural networks [200], with appropriate choices of parameters.

See Section 6.4 for more details.

Chapter Outline

In Section 6.2 we consider a general pseudometric on $\bigsqcup_n \Omega_n$ and prove that the existence of uniform rates for generalization and approximation is equivalent to compactness of this pseudometric space. We then consider sampling metrics more specifically, and show that sequences (Ω_n) of compact sets closed under sampling naturally yield compact pseudometric spaces. We further prove a general equivalence between limit objects in sampling metric and data sources. Finally, we introduce our generalization of sampling with replacement and random binning. In Section 6.3, we prove general rates for sketching and generalization with respect to sampling with replacement. We then instantiate our rates for a variety of function classes. In Section 6.4, we do the same for sketching and generalization with respect to random binning, where we also introduce our generalization of species sampling. We end in Section 6.5 with several questions and directions for future work.

Related Work

There are several machine learning models that are defined for inputs of different sizes. Examples include neural networks processing sets and point clouds of different sizes [30, 182, 233], graphs of different sizes [152, 193, 200], and transformers processing sequences of tokens of any length [82, 145, 226]. Importantly, the ability of each of the above models to generalize to inputs of unseen sizes is much less studied, with a few notable exceptions. The ability of a graph neural network to generalize to graphs of different sizes has been called transferability in the literature, and has been extensively studied by considering appropriate topologies and limits on the space of graphs of all sizes, see [135, 153, 154, 193] for example. We generalize and unify many of the techniques in this literature by considering more general notions of sampling maps. There is also a literature studying the ability of transformers to generalize to inputs of different lengths, see [104, 119, 229, 236] for example. Some of the results in this literature are negative, showing that transformers often do *not* generalize well to longer inputs, while others prove length generalization under specific assumptions on the transformer architecture. Our theory contributes to the latter line of work by giving explicit generalization guarantees for permutation-invariant transformers. The authors of [61] study generalization of symmetric any-dimensional polynomials fitted via least-squares to inputs of different dimensions. They measure generalization using mean-squared error with respect to fixed training and test distributions in fixed, but possibly different dimensions, and their bounds depend explicitly on these distributions. In contrast, our bounds simultaneously control the error on inputs of all possible sizes by the error in a fixed training dimension, which needs to be controlled separately depending on how the polynomial is fitted to data.

Transferability of graph neural networks has been extended in [142] to more general any-dimensional models, where it was defined to be continuity of the model in a certain limiting space obtained from a Banach space containing inputs of all sizes. Importantly, to instantiate the results of [142] one needs to first obtain rates at which finite-dimensional inputs converge to limiting ones which are quoted case-by-case from the literature in [142]. Likewise, some of the results in [142] require compact sets of growing-sized inputs and their limits, obtained on a case-by-case basis again. In contrast, we systematically defined sampling maps that (constructively) approximate arbitrary objects by finite-dimensional ones, and we derive convergence rates for these approximations in a unified fashion. Furthermore, our sampling-based approach naturally yields rich classes of compact sets for which our theory applies. Finally, our approach yields improved rates compared to [142] by exploiting sampling-specific structure in the functions we are learning or sketching.

The rates we obtain for sampling with replacement generalize previous rates obtained for measures [80] (viewed as limits of point clouds) and graphons [20] (viewed as limits of dense simple undirected graphs). In particular, rates of approximation for graphons by finite graphs were shown to follow from compactness of the space of graphons in [149], a connection we generalize in Theorem 6.2.1.

Notation

We denote the distribution of a random variable X by $\mathrm{Law}(X)$. We denote the collection of (Borel) probability distributions on $\Theta \subseteq \mathbb{R}^d$ by $\mathcal{P}(\Theta)$. If $f\colon \mathbb{R}^d \to \mathbb{R}^k$ and $\mu \in \mathcal{P}(\mathbb{R}^d)$, we denote by $f\mu$ the pushforward of μ by f. If $\mu \in \mathcal{P}(\Theta)$ and $f\colon \Theta \to \mathbb{R}$ is a function, we denote by $\mathbb{E}_\mu f = \mathbb{E}_{X\sim\mu} f(X)$ the expectation of f with respect to μ. A sequence of measures $(\mu_n) \subseteq \mathcal{P}(\Theta)$ converge weakly to $\mu \in \mathcal{P}(\Theta)$ if $\mathbb{E}_{\mu_n} f \to \mathbb{E}_\mu f$ for all continuous functions $f\colon \Theta \to \mathbb{R}$. A sequence of random variables (X_n) converge weakly if their distributions $(\mathrm{Law}(X_n))$ converge weakly. If (Ω_n) are sets, we denote by $\bigsqcup_n \Omega_n = \{(n,x) : x \in \Omega_n, n \in \mathbb{N}\}$ their disjoint union. If $\sim$ is an equivalence relation on a set Ω, we denote by $\Omega/{\sim}$ the quotient space consisting of equivalence classes $[x] = \{y \in \Omega : y \sim x\}$.

6.2 Comparing Objects of Different Sizes

In this section, we begin our study of generalization and approximation across dimensions by considering how inputs of different sizes are compared to each other. We start by stating and proving a general equivalence between various notions of generalization and approximation across dimensions and compactness of the metric space consisting of inputs of all sizes and their limits. We then focus on the sampling metric, describe a large and natural family of compact sets in it, and describe a general equivalence between random limit objects and data distributions.

The Role of Compactness

We begin by formally studying a pseudo-metric defined on inputs of all sizes, and prove that compactness of this metric is equivalent to the uniform approximation and any-dimensional generalization properties from Section 3.1. While the argument is elementary, we include it here for completeness.

Suppose δ is a pseudo-metric on $\bigsqcup_n \Omega_n$. Define the equivalence relation $x \sim y$ if $\delta(x,y) = 0$, let $\Omega_\infty = \bigsqcup_n \Omega_n/{\sim}$ and let $\overline{\Omega}_\infty$ by the completion of Ω_∞ with respect to δ. Also let $\Omega_{\leq n} = \bigsqcup_{i\leq n} \Omega_n/{\sim}$. Denote by $\mathcal{F}_1$ the collection of 1-Lipschitz functions on $\bigsqcup_n \Omega_n$, or equivalently, on $\overline{\Omega}_\infty$.

Theorem 6.2.1. *Assume $(\Omega_n/\!\sim, \delta)$ is compact for each n. Then the following are equivalent.*

1. *$(\overline{\Omega}_\infty, \delta)$ is compact.*
2. *For any $\epsilon > 0$, there exists $N(\epsilon) \in \mathbb{N}$ and finite $S_\epsilon \subseteq \Omega_{\leq N(\epsilon)}$ such that S_ϵ is an ϵ-net for $\overline{\Omega}_\infty$.*
3. *$(\overline{\Omega}_\infty, \delta)$ satisfies the uniform approximation property.*
4. *For any $f, f_0 \in \mathcal{F}_1$, we have $\lim_{n\to\infty} \sup_{\substack{f\in\mathcal{F}_1 \\ f|_{\Omega_{\leq n}} = f_0|_{\Omega_{\leq n}}}} \mathrm{err}(f, f_0) = 0$.*
5. *$(\overline{\Omega}_\infty, \delta)$ satisfies the any-dimensional generalization property.*

Proof. Since $\overline{\Omega}_\infty$ is a complete metric space by construction, it is compact if and only if it is totally bounded. Property 2 implies $\overline{\Omega}_\infty$ is totally-bounded. Conversely, if $\overline{\Omega}_\infty$ is totally bounded and $\{y_1, \ldots, y_k\} \subseteq \overline{\Omega}_\infty$ is a finite $\epsilon/2$-net, let $x_i \in \Omega_\infty$ satisfy $d(x_i, y_i) \leq \epsilon/2$ and note that $\{x_1, \ldots, x_k\} \subseteq \Omega_\infty$ is a finite ϵ-net as well. Thus, statements 1 and 2 are equivalent.

We argue statements 2 and 3 are equivalent. Indeed, if statement 2 holds then for any $\epsilon > 0$ we have $\sup_{x\in\Omega_\infty} \mathrm{dist}(x, \Omega_{\leq n}) \leq \epsilon$ for all $n \geq N(\epsilon)$, hence statement 3 holds. Conversely, if statement 3 holds let $N(\epsilon)$ satisfy $\sup_{x\in\Omega_\infty} \mathrm{dist}(x, \Omega_{\leq N(\epsilon)}) \leq \epsilon/2$, and let $S_\epsilon \subseteq \Omega_{\leq N(\epsilon)}$ be a finite $\epsilon/2$-net for $\Omega_{\leq N(\epsilon)}$, which exists by compactness of each Ω_n. Then S_ϵ is an ϵ-net for Ω_∞, so statement 2 holds.

We proceed to argue that statements 3 and 4 are equivalent. If statement 3 holds and $f \in \mathcal{F}_1$ satisfies $f|_{\Omega_{\leq n}} = f_0|_{\Omega_{\leq n}}$, then for any $x \in \Omega_\infty$ we have $|f(x) - f_0(x)| \leq \inf_{y\in\Omega_{\leq n}}\{|f(x)-f(y)|+|f(y)-f_0(y)|+|f_0(y)-f_0(x)|\} \leq 2\mathrm{dist}(x, \Omega_{\leq n})$. Thus, statement 4 holds. Conversely, if statement 4 holds, then set $f_0 = 0$ and $f(x) = \mathrm{dist}(x, \Omega_{\leq n})$. Observe that $f \in \mathcal{F}_1$, that $f|_{\Omega_{\leq n}} = f_0|_{\leq n}$, and that $\|f - f_0\|_\infty = \sup_{x\in\Omega_\infty} \mathrm{dist}(x, \Omega_{\leq n})$. Thus, statement 3 holds.

Finally, we argue that statement 5 is equivalent to the previous equivalent four statements. Statement 5 implies statement 4, since if $f|_{\Omega_{\leq n}} = f_0|_{\Omega_{\leq n}}$ then $\mathrm{err}_{\leq n}(f, f_0) = 0$ and $\|f - f_0\|_\infty \leq R_n$. Conversely, suppose statement 3 (equivalent to statement 4) holds. Then $\|f - g\|_\infty \leq \mathrm{err}_{\leq n}(f, g) + 2\sup_{x\in\Omega_\infty} \mathrm{dist}(x, \Omega_{\leq n})$ by the triangle inequality, hence statement 5 holds. $\square$

We can also characterize any-dimensional generalization using an average error with respect to a data distribution. However, we need to carefully consider what a meaningful notion of a data distribution is. A natural choice might seem to be a distribution on inputs of all sizes $\mu \in \mathcal{P}(\Omega_\infty)$, but any such distribution is necessarily concentrated on bounded-sized inputs since $\lim_n \mu(\Omega_\infty \setminus \Omega_{\leq n}) = 0$. Instead, we consider average error with respect to sequences of distributions supported on inputs of growing size that converge weakly to a distribution on limit objects, as formalized by the following proposition.

Proposition 6.2.2. *Suppose $(\overline{\Omega}_\infty, \delta)$ is compact. For any sequence of distributions $(\mu_n \in \mathcal{P}(\Omega_{\leq n}))$ and $\mu_\infty \in \mathcal{P}(\overline{\Omega}_\infty)$, we have $\mu_n \to \mu_\infty$ weakly if and only if there exists a rate rate $(R_n \geq 0)$ with $R_n \downarrow 0$ such that for any $f, g \in \mathcal{F}_1$, we have $\mathrm{err}_{\mu_\infty}(f,g) \leq \mathrm{err}_{\mu_n}(f,g) + R_n$.*

Proof. Since $\overline{\Omega}_\infty$ is compact, we have $\mu_n \to \mu_\infty$ weakly if and only if $\lim_n W_1(\mu_n, \mu_\infty) = 0$, or equivalently, if and only if $\lim_n \sup_{f \in \mathcal{F}_1} |\mathbb{E}_{\mu_\infty} f - \mathbb{E}_{\mu_n} f| = 0$. Thus, if $\mu_n \to \mu_\infty$ weakly then $\mathrm{err}_{\mu_\infty}(f,g) \leq \mathrm{err}_{\mu_n}(f,g) + W_1(\mu_n, \mu_\infty$ since if $f, g \in \mathcal{F}_1$ then $|f - g| \in \mathcal{F}_1$. Conversely, if such rate R_n exists, note that

$$\begin{aligned}
W_1(\mu_n, \mu_\infty) &= \sup_{f \in \mathcal{F}_1} |\mathbb{E}_{\mu_\infty} f - \mathbb{E}_{\mu_n} f| = \max\left\{\sup_{f \in \mathcal{F}_1} (\mathbb{E}_{\mu_\infty} f - \mathbb{E}_{\mu_n} f),\ \sup_{f \in \mathcal{F}_1} (\mathbb{E}_{\mu_\infty}(-f) - \mathbb{E}_{\mu_n}(-f))\right\} \\
&= \sup_{f \in \mathcal{F}_1} (\mathbb{E}_{\mu_\infty} f - \mathbb{E}_{\mu_n} f) = \sup_{f \in \mathcal{F}_1} (\mathbb{E}_{\mu_\infty}(f - \inf_{x \in \Omega_\infty} f(x)) - \mathbb{E}_{\mu_n}(f - \inf_{x \in \Omega_\infty} f(x))) \\
&= \sup_{\substack{f \in \mathcal{F}_1 \\ f \geq 0}} (\mathbb{E}_{\mu_\infty} f - \mathbb{E}_{\mu_n} f) \leq R_n \to 0,
\end{aligned}$$

and hence $\mu_n \to \mu_\infty$ weakly. Above, in going from the first to the second line we used the fact that $-f \in \mathcal{F}_1$ iff $f \in \mathcal{F}_1$, and the last inequality follows since $\mathrm{err}_\mu(f, 0) = \mathbb{E}_\mu f$ when $f \geq 0$. □

The above results are completely general, and show that any compact pseudo-metric between objects of different sizes enables generalization across dimensions for functions that are continuous with respect to that metric. We now turn to analyzing the sampling metric (6.2) in more detail, and show that it admits rich families of compact sets and a natural correspondence between limit objects and data distributions.

The Sampling Metric

Consider vector space $(\mathbb{V}_n)$ and sampling maps $(\mathsf{S}_k \colon \bigsqcup_n \mathbb{V}_n \to \mathbb{V}_k)$, where each S_k is viewed as a random map.[1] We begin by showing that any sequence of compact sets $(\Omega_n \subseteq \mathbb{V}_n)$ closed under these sampling maps yields a compact limiting space.

Proposition 6.2.3. *Suppose $(\Omega_n \subseteq \mathbb{V}_n)$ is a sequence of compact sets in the usual topology such that $\mathsf{S}_k\Omega_n \subseteq \Omega_k$ almost surely, and consider the sampling metric (6.2) on $\bigsqcup_n \Omega_n$. Then each Ω_n remains compact in d_{samp} and $(\overline{\Omega}_\infty, d_{\mathrm{samp}})$ is compact as well.*

Proof. Endow each $\mathcal{P}(\Omega_n)$ with the weak topology, metrized by the W_1-metric, and endow $\prod_n \mathcal{P}(\Omega_n)$ with the product topology, metrized by $d((\mu_k),(\nu_k)) = \sum_{k\geq 1} 2^{-k} W_1(\mu_k,\nu_k)$. Note that $\prod_n \mathcal{P}(\Omega_n)$ is compact by Tychonoff's theorem, and that the map sending $x \mapsto (\mathrm{Law}(\mathsf{S}_k(x)))_k$ is an isometric embedding of Ω_∞ into $\prod_n \mathcal{P}(\Omega_n)$. In particular, it maps each compact Ω_n to a compact subset in the product topology, and extends to an isometry between $\overline{\Omega}_\infty$ as a closed, and therefore compact, subset of $\prod_n \mathcal{P}(\Omega_n)$. □

We proceed to give several examples of sampling maps and sets satisfying the above hypotheses.

Example 6.2.4 (Sampling entries with replacement)**.** *Suppose $\mathbb{V}_n = \mathbb{R}^{d\times n}$, and for any $x \in \mathbb{V}_n$ the sampling map S_k samples k columns from x uniformly at random with replacement, i.e., $(\mathsf{S}_k x)_{:,i} = x_{:,J_i}$ where $J_1,\ldots,J_k$ are iid uniformly random in $[n]$. Denoting by $\mu_x = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$ the empirical measure of the entries of x, we note that $\mathrm{Law}(\mathsf{S}_k x) = \mu_x^{\otimes k}$. Here $x \sim y$ precisely when x, y differ by permutations and duplication of columns the same number of times. Note that for any compact $\Theta \subseteq \mathbb{R}^d$, the sequence of product sets $\Omega_n = \Theta^n$ is closed under sampling. In this case we have an embedding $\Omega_\infty \hookrightarrow \mathcal{P}(\Theta)$ by sending $x \mapsto \mu_x$, which identifies Ω_∞ with the collection of finitely-supported measures on Θ with rational weights. We also have $\overline{\Omega}_\infty \cong \mathcal{P}(\Theta)$, which is indeed compact.*

Example 6.2.5 (Sampling vertices with replacement)**.** *Suppose $\mathbb{V}_n = \mathbb{R}^{n\times n}$, viewed as the space of weighted and directed graphs, and for any $X \in \mathbb{V}_n$ the sampling map S_k samples k vertices with replacement and extracts the corresponding induced subgraph, i.e., $(\mathsf{S}_k X)_{i,j} = X_{J_i,J_j}$ for $J_1,\ldots,J_k \overset{iid}{\sim} \mathrm{Unif}([n])$. If $\Omega_n = \{0,1\}^{n\times n} \cap \mathbb{S}^n$,*

[1] We can also view S_k as a kernel, i.e., a map $\bigsqcup_n \mathbb{V}_n \to \mathcal{P}(\mathbb{V}_k)$. The two views are equivalent by [114, Lemma 3.2(vii)].

then (Ω_n) *is closed under sampling, and limits in* d_{samp} *correspond to dense graph limits [146]. In particular, we can identify* $\overline{\Omega}_\infty$ *with the space of graphons [146, §7], which was shown to be compact in [149] as a consequence of Szemerédi's regularity lemma.*

Example 6.2.6 (Random Binning)**.** *Suppose* $\mathbb{V}_n = \mathbb{R}^n$ *and for any* $x \in \mathbb{V}_n$*, let the sampling map* S_k *randomly assigning the* n *entries of* x *into* k *bins and summing all the entries in each bin, i.e.,* $\mathsf{S}_k x = \sum_{i=1}^n x_i e_{J_i}$ *where* $J_1, \ldots, J_n \overset{iid}{\sim} \mathrm{Unif}([k])$*. This random binning, also called hashing, has found uses in machine learning as a dimensionality reduction method [51, 203] and in information theory as part of source and channel coding protocols [232].*

If $\Omega_n = \Delta^{n-1}$*, then* (Ω_n) *is closed under sampling. In this case, we can identify the (compact) limit space* $\overline{\Omega}_\infty$ *with the solid simplex* $\Delta^\infty = \{x \in \ell_\infty : x \geq 0, \sum_i x_i \leq 1\}/\sim$ *modulo permutations and the topology induced by the* ℓ_∞ *norm.*

Example 6.2.7 (Random Partition Models)**.** *Suppose* $\mathbb{V}_n = \mathbb{R}^n$ *and for any* $x \in \mathbb{V}_n$*, define the following size-biased sampling. If* $x \in \Delta^{n-1}$*, view* x *as a probability distribution on* $[n]$ *and sample* k *indices* $J_1, \ldots, J_k$ *iid according to this distribution. If* $t_1, \ldots, t_m$ *is a random enumeration of the distinct indices in* $\{J_1, \ldots, J_k\}$ *(where* $m \leq k$*), let*

$$\mathcal{S}_k(x)_i = \frac{|\ell : J_\ell = t_i\}|}{k}, \quad \text{for } i \leq m, \tag{6.9}$$

and $\mathcal{S}_k(x)_i = 0$ *for* $i > m$*. If* $x \in \mathbb{R}^n$ *is a general vector, define* $\mathcal{S}_k(x)_i = \|x\|_1 \mathrm{sign}(x_{t_i}) \mathcal{S}_k(|x|/\|x\|_1)_i$*.*

The sequence of indices $J_1, \ldots, J_k$ *we sample from* x *to form* $\mathcal{S}_k(x)$ *gives a random partition of* $[k]$*, where* $i, j \in [k]$ *lie in the same block of the partition if* $J_i = J_j$*. Likewise, note that* $k\mathcal{S}_k(x) \in \mathbb{Z}_{\geq 0}^k$ *and defines a random partition of the integer* k *for each* $k \in \mathbb{N}$*. Such sequences of random partitions are known as partition structures in the literature [176]. In particular, Kingman [122, 123] shows that any sequence of exchangeable random partitions arises from* $x \in \Delta^\infty$ *as above.*

We show below that the limit space under this choice of sampling maps is the same as in Example 6.2.6, and in fact, that the two sets of sampling maps are compatible with each other in useful ways.

For any choice of sampling maps (S_k) and compact sets (Ω_n) closed under sampling, Theorem 6.2.1 guarantees the existence of universal rates for approximation of objects by low-dimensional ones, and for generalization of functions from fixed to arbitrary input dimensions. In the next two sections, we give such rates for broad families of sampling maps generalizing all the above examples. As in the proof of Theorem 6.2.1, our derivation of these rates is based on bounding the distance $\mathrm{dist}(x, \mathbb{V}_n)$ between an arbitrary limit object $x \in \overline{\Omega}_\infty$ and objects in a fixed dimension. It turns out that in all of the above examples, we have $\mathbb{E}\mathrm{dist}(x, \mathsf{S}_n x) \to 0$ as $n \to \infty$, and analyzing the rate of convergence of the random samples $\mathsf{S}_n x$ back to x yields our desired bounds.

Proposition 6.2.2 also gives us generalization guarantees in average error with respect to data distributions, defined as convergent sequences of distributions on growing-sized inputs. When using a sampling metric, there is a correspondence between distributions on limit objects and such data distributions. Indeed, observe that the map $\Omega_\infty \to \prod_n \mathcal{P}(\Omega_n)$ sending $x \mapsto (\mathrm{Law}(\mathsf{S}_n x))$ is an isometric embedding, and hence we can view limit objects in $\overline{\Omega}_\infty$ as sequences of distributions indexed by dimension. If $\mathbb{E}\mathrm{dist}(x, \mathsf{S}_n x) \to 0$, it is easy to see that $\mathrm{Law}(\mathsf{S}_n x) \to \delta_x$ weakly, thus defining a particular class of data distributions obtained by sampling from a deterministic limit object. We can further take mixtures of such sequences of distributions, yielding the closed convex hull

$$\overline{\mathcal{P}}(\Omega_\bullet) = \overline{\mathrm{conv}}\{(\mathrm{Law}(\mathsf{S}_n x))_{n\in\mathbb{N}} : x \in \Omega_\infty\} \subseteq \prod_n \mathcal{P}(\Omega_n),$$

where the closure is taken with respect to product of weak topologies on each $\mathcal{P}(\Omega_n)$. We proceed to show that all sequences of distributions in $\overline{\mathcal{P}}(\Omega_\bullet)$ can be obtained by sampling from a random limit object.

Endow $\mathcal{P}(\overline{\Omega}_\infty)$ with the Wasserstein-1 metric with respect to d_{samp}. Likewise, endow $\prod_n \mathcal{P}(\Omega_n)$ with the metric $d((\mu_n), (\nu_n)) = \sum_{n\geq 1} 2^{-n} W_1(\mu_n, \nu_n)$.

Proposition 6.2.8. *Fix sampling maps (S_n) and compact sets (Ω_n) closed under sampling. Define the map $\mathcal{L}\colon \mathcal{P}(\overline{\Omega}_\infty) \to \overline{\mathcal{P}}(\Omega_\bullet)$ sending $\mathcal{L}(\nu) = (\mathrm{Law}(\mathsf{S}_n X))_{n\in\mathbb{N}}$ where $X \sim \nu$ is independent of S_n.*

1. *The map $\mathcal{L}$ is a linear 1-Lipschitz, surjective, and an isometry on $\overline{\Omega}_\infty \subseteq \mathcal{P}(\overline{\Omega}_\infty)$.*

2. *Suppose $\mathbb{E}d_{\mathrm{samp}}(x, \mathsf{S}_n x) \to 0$ for every $x \in \overline{\Omega}_\infty$. Then for every $\nu \in \mathcal{P}(\overline{\Omega}_\infty)$ we have $\mathrm{Law}(\mathsf{S}_n X) \to \nu$ weakly in $\mathcal{P}(\overline{\Omega}_\infty)$, where $X \sim \nu$ is independent of S_n. Furthermore, in this case $\mathcal{L}$ is injective and $\mathcal{L}(\overline{\Omega}_\infty) = \mathrm{ext}\overline{\mathcal{P}}(\Omega_\bullet)$.*

Proof. It is clear that $\mathcal{L}$ maps mixtures to mixtures because $\mathrm{Law}(\mathsf{S}_n X) = \int \mathrm{Law}(\mathsf{S}_x x)\, d\nu(x)$ if $X \sim \nu$, i.e., is linear. Let $X_i \sim \nu_i$ for $\nu_1, \nu_2 \in \mathcal{P}(\overline{\Omega}_\infty)$. The map $\mathcal{L}$ is 1-Lipschitz because

$$d(\mathcal{L}(\nu_1), \mathcal{L}(\nu_2)) = \sum_{n\geq 1} 2^{-n} \inf_{\substack{(Y'_{1,n}, Y'_{2,n}) \\ Y'_{i,n} \overset{d}{=} \mathsf{S}_n X_i}} \mathbb{E}\|Y'_{1,n} - Y'_{2,n}\|$$

$$\leq \inf_{\substack{(X'_1, X'_2) \\ X'_i \overset{d}{=} X_i}} \sum_{n\geq 1} 2^{-n} \mathbb{E}_{x_i \sim X'_i} \inf_{\substack{(Y'_{1,n}, Y'_{2,n}) \\ Y'_{i,n} \overset{d}{=} \mathsf{S}_n x_i}} \mathbb{E}\|Y'_{1,n} - Y'_{2,n}\| = W_1(\nu_1, \nu_2).$$

In words, in the first line we minimize over all couplings of $(\mathsf{S}_n X_1, \mathsf{S}_n X_2)$ for each n, whereas in the second line we restrict to couplings obtained by first coupling (X'_1, X'_2), sampling $x_i \sim X'_i$ and then coupling $(\mathsf{S}_n x_1, \mathsf{S}_n x_2)$. Note also that when $\nu_i = \delta_{x_i}$ are deterministic measures on $\overline{\Omega}_\infty$, then $d(\mathcal{L}(\nu_1), \mathcal{L}(\nu_2)) = d_{\mathrm{samp}}(x_1, x_2) = W_1(\nu_1, \nu_2)$, hence $\mathcal{L}$ is an isometry on the extreme points of $\mathcal{P}(\overline{\Omega}_\infty)$. Finally, the image of $\mathcal{L}$ is dense by definition of $\overline{\mathcal{P}}(\Omega_\bullet)$ and the linearity of $\mathcal{L}$. It is also compact since $\mathcal{P}(\overline{\Omega}_\infty)$ is compact and $\mathcal{L}$ is continuous. Thus, the map $\mathcal{L}$ is surjective and the first claim is proved.

For the second claim, consider the coupling of ν and $\mathrm{Law}(\mathsf{S}_n X)$ with $X \sim \nu$, whereby we first sample $X \sim \nu$ and then independently sample S_n to form $(X, \mathsf{S}_n X)$. Note that

$$\limsup_{n\to\infty} W_1(\nu, \mathrm{Law}(\mathsf{S}_n X)) \leq \limsup_{n\to\infty} \mathbb{E}_X \mathbb{E}_{\mathsf{S}_n} d_{\mathrm{samp}}(X, \mathsf{S}_n X) \leq \mathbb{E} \limsup_{n\to\infty} \mathbb{E}_{\mathsf{S}_n} d_{\mathrm{samp}}(X, \mathsf{S}_n X) = 0,$$

where we used Fatou's lemma to interchange the limit and expectation. Thus, if $\mathcal{L}(\nu_1) = \mathcal{L}(\nu_2)$ then $\mathrm{Law}(\mathsf{S}_n X_1) = \mathrm{Law}(\mathsf{S}_n X_2)$ and hence $\nu_1 = \nu_2$, so ν is injective and hence a linear isomorphism. Finally, if $\mathcal{L}(\delta_x)$ is not an extreme point, then by part 1 we have $\mathcal{L}(\delta_x) = \mathcal{L}(\nu)$ for a non-deterministic ν. By injectivity, we have $\delta_x = \nu$, a contradiction. Conversely, we have shown in part 1 that every element of $\overline{\mathcal{P}}(\Omega_\bullet)$ is a mixture of those in $\mathcal{L}(\overline{\Omega}_\infty)$, and hence $\mathcal{L}(\overline{\Omega}_\infty) \supseteq \mathrm{ext}\overline{\mathcal{P}}(\Omega_\bullet)$, proving the second claim. □

Proposition 6.2.8(2) identifies limit objects with extremal sequences of measures, and further shows that $\overline{\mathcal{P}}(\Omega_\bullet)$ is a Choquet simplex. In this case, any distribution on $\overline{\Omega}_\infty$ can be identified with a sequence of distributions indexed by dimensions obtained by first sampling a limit object and then sampling from the limit object. Moreover, We will view such sequences of distributions as our data distributions.

Example 6.2.9 (de Finetti, Aldous–Hoover)**.** *In the setting of Example 6.2.4 with $\Omega_n = \Theta^n$ for compact $\Theta \subseteq \mathbb{R}^d$, we have $\overline{\Omega}_\infty = \mathcal{P}(\Theta)$ so limit objects are probability distributions on Θ, hence $\mathcal{P}(\overline{\Omega}_\infty)$ corresponds to random probability measures on Θ. On the other hand, if $\nu \in \mathcal{P}(\Theta)$ then $\mathrm{Law}(\mathsf{S}_k X) = \nu^{\otimes k}$, and taking mixtures we conclude that $\overline{\mathcal{P}}(\Omega_\bullet)$ corresponds to mixtures of iid arrays. De Finetti's theorem implies that the map $\mathcal{L}$ from Proposition 6.2.8 is indeed an isomorphism, and that $\overline{\mathcal{P}}(\Omega_\bullet)$ also corresponds to infinite exchangeable arrays.*

Similarly, in the setting of Example 6.2.5 with $\Omega_n = \Theta^{n\times n}$, it follows from [138] that $\overline{\mathcal{P}}(\Omega_\bullet)$ corresponds to two-dimensional exchangeable arrays with entries in Θ, and hence limit objects $\overline{\Omega}_\infty$ corresponds to extremal such arrays. These have been characterized by Aldous–Hoover, and for the special case $\Omega_n = \{0,1\}^{n\times n} \cap \mathbb{S}^n$ correspond to symmetric measurable functions $[0,1]^2 \to [0,1]$ called graphons.

In the next two sections, we turn to proving general rates of convergence in sampling metric using various notions of sampling with replacement and random binning, and use these rates to derive generalization and approximation bounds for functions across input dimension.

6.3 Sampling with Replacement

In this section, we define a generalization of sampling with replacement and study rates for approximation and generalization under the resulting sampling metric. Our generalization encompasses sampling entries and vertices as in Examples 6.2.4-6.2.5 as well as sampling from higher-order tensors.

Sampling using Set Maps

We generalize the sampling with replacement used in Examples 6.2.4-6.2.5 using maps between finite sets, similarly to [138]. To motivate our generalization, we observe that any map $f\colon [k] \to [n]$ defines a linear map $\varsigma(f)\colon \mathbb{R}^n \to \mathbb{R}^k$ extracting entries specified by f. Formally for any $x \in \mathbb{R}^n$ the entries of $\varsigma(f)x \in \mathbb{R}^k$ are given by $(\varsigma(f)x)_i = x_{f(i)}$. Moreover, if $F_{n,k}\colon [k] \to [n]$ is a uniformly random such map, so that $F_{n,k}(1), \dots, F_{n,k}(k) \overset{iid}{\sim} \mathrm{Unif}([n])$, then $\varsigma(f_{n,k})x$ is precisely the vector obtained by sampling k coordinates from x uniformly at random with replacement as in Example 6.2.4. Likewise, any map $f\colon [k] \to [n]$ acts on tuples $f^{\times 2}\colon [k]^2 \to [n]^2$ by $f^{\times 2}(i,j) = (f(i), f(j))$, and hence acts linearly on matrices indexed by these tuples via $(\varsigma(f)X)_{i,j} = X_{f(i),f(j)}$. Once again, applying a uniformly random map $\varsigma(f_{n,k})X$ recovers sampling vertices as in Example 6.2.5.

We generalize these and other examples in three steps. First, we consider actions of maps between finite sets on discrete index sets. Then, we extend these actions to vectors indexed by such index sets. Finally, we define sampling maps using the action of random maps between finite sets.

Actions on index sets. We begin by considering actions of maps between finite sets on index sets, and quantifying their complexity.

Definition 6.3.1 (FinSet-indices)**.** *A* FinSet-*index is a sequence of finite sets* $(\mathcal{A}_n)$ *together with maps* $\rho(f_{n,k})\colon \mathcal{A}_k \to \mathcal{A}_n$ *associated to each map* $f_{n,k}\colon [k] \to [n]$ *between finite sets satisfying (i)* $\rho(\mathrm{id}_{[n]}) = \mathrm{id}_{\mathcal{A}_n}$*; and (ii)* $\rho(f_1 \circ f_2) = \rho(f_1) \circ \rho(f_2)$ *whenever the composition is well-defined.* [2]

We have already seen the examples $\mathcal{A}_n = [n]$ and $\mathcal{A}_n = [n]^2$. Other examples include $\mathcal{A}_n = [n]^2 \sqcup [n]$ and $\mathcal{A}_n = \mathrm{Sym}^2([n]) = [n]^2/\sim$ where we identify $(i, j) \sim (k, \ell)$ if either $(i, j) = (k, \ell)$ or $(i, j) = (\ell, k)$, which can be thought of as the collection of multisets of $[n]$ of size 2.

Standard operations on sets can be applied to create new FinSet-indices from previous ones. For example, if $\{(\mathcal{A}_n), \rho_a\}$ and $\{(\mathcal{B}_n), \rho_b\}$ are two FinSet-indices, then so are $\{(\mathcal{A}_n \sqcup \mathcal{B}_n), \rho_n \sqcup \rho_b\}$ and $\{(\mathcal{A}_n \times \mathcal{B}_n), \rho_a \times \rho_b\}$. In this paper, we focus on FinSet-indices that can be formed from the basic one $\mathcal{A}_n = [n]$ using such standard operations, and we quantify the complexity of these indices using the following notion of 'degree'.

Definition 6.3.2 (Degree of a FinSet-index)**.** *A* FinSet-*index* $(\mathcal{A}_n)$ *has degree* D *if* $\mathcal{A}_n = \bigsqcup_{m=1}^{M} [n]^{d_m}/H_m$ *where* $D = \max_m d_m$ *and* $H_m \subseteq \mathsf{S}_{d_m}$ *is a subgroup of permutations acting by permuting the multi-indices in* $[n]^{d_m}$*. A* FinSet-*index has finite degree if it has degree* d *for some* $d \in \mathbb{N}$*.* [3]

For example, the degree of $\mathcal{A}_n = [n] \times [d] = \bigsqcup_{i=1}^{d} [n]$ is 1, while the degrees of $\mathcal{A}_n = [n]^2$, $\mathcal{A}_n = [n]^2 \sqcup [n]$, and $\mathcal{A}_n = \mathrm{Sym}^2([n]) = [n]^2/S_2$ are all 2.

Sampling action on vectors. We now extend the above action of maps between fintie sets on indices to an action on vectors indexed by them. If $(\mathcal{A}_n)$ is a FinSet-index, we can form the vector spaces $\mathbb{V}_n = \mathbb{R}^{\mathcal{A}_n}$ consisting of vectors indexed by

[2] In the language of category theory, this is a functor from the category FinSet of finite sets to itself.

[3] Our notion of degree is closely related to the degree of a polynomial functor [166].

$\mathcal{A}_n$, and let maps between finite sets act on these vector spaces by

$$(\varsigma(f_{n,k})x)_\alpha = x_{\rho(f_{n,k})(\alpha)}, \quad \text{whenever } x \in \mathbb{V}_n \text{ and } f_{n,k}\colon [k] \to [n]. \tag{6.10}$$

In other words, $\varsigma(f_{n,k})x$ extracts a subset of the entries of x specified by $f_{n,k}$. When $\mathcal{A}_n = [n]$ or $[n]^2$ this recovers the actions on vectors and matrices we considered in the beginning of this section. Note that $\varsigma(f_1 \circ f_2) = \varsigma(f_2) \circ \varsigma(f_1)$ whenever this composition is well-defined, so that the action ς reverses the order of compositions. Such composition-reversing linear actions of maps between finite sets were called $\mathrm{FinSet}^{\mathrm{op}}$-representations in [138].

Sampling with replacement. Finally, we define our sampling maps in terms of the action of uniformly random maps between finite sets. If $(\mathcal{A}_n)$ is a FinSet-index and $\mathbb{V}_n = \mathbb{R}^{\mathcal{A}_n}$ is the associated sequence of vector spaces, we then define sampling maps $\mathsf{S}_k\colon \bigsqcup_n \mathbb{V}_n \to \mathbb{V}_k$ by

$$\mathsf{S}_k x = \varsigma(F_{n,k})x \quad \text{if } x \in \mathbb{V}_n, \quad \text{where } F_{n,k}\colon [k] \to [n] \text{ is uniformly random.} \tag{6.11}$$

As discussed above, this generalizes sampling columns from a matrix with replacement when $\mathcal{A}_n = [d] \times [n]$ and sampling vertices with replacement from a graph when $\mathcal{A}_n = [n]^2$, or more generally from a hypergraph when $\mathcal{A}_n = [n]^d$.

Symmetry and relations between dimensions. Permutation symmetry plays a central role in the analysis of the sampling maps (6.11). To see this symmetry, observe that if $\mathfrak{S}_n = \{\sigma\colon [n] \to [n] \text{ bijective}\}$ is the group of permutations on n letters, then $\mathfrak{S}_n$ acts on the nth index set $\mathcal{A}_n$ and hence on the associated vector space $\mathbb{V}_n$. This action makes $\mathbb{V}_n$ into a representation of $\mathfrak{S}_n$.

Proposition 6.3.3 (Permutation Symmetry)**.** *Suppose $(\mathcal{A}_n)$ is a* FinSet*-index, let $\mathbb{V}_n = \mathbb{R}^{\mathcal{A}_n}$ be the associated vector spaces endowed with sampling maps* (6.11)*.*

(Exchangeability) *Each distribution* $\mathrm{Law}(\mathsf{S}_k x)$ *is exchangeable, meaning $\sigma_k \mathsf{S}_k x \overset{d}{=} \mathsf{S}_k x$ for each $\sigma_k \in \mathfrak{S}_k$.*

(Permutation-invariance) *For any $n \in \mathbb{N}$, any $x \in \mathbb{V}_n$, and any permutation $\sigma_n \in \mathfrak{S}_n$, we have $\mathsf{S}_k \sigma_n x \overset{d}{=} \mathsf{S}_k x$ for all k and hence $d_{\mathrm{samp}}(x, \sigma_n x) = 0$.*

Proof. Let $F_{n,k}\colon [k] \to [n]$ be a uniformly random map, and note that $F_{n,k} \circ \sigma_k \overset{d}{=} \sigma_n \circ F_{n,k} \overset{d}{=} F_{n,k}$ since all three maps send each $i \in [k]$ independently to a uniformly random element of $[n]$. Therefore, if $x \in \mathbb{V}_n$ then $\mathsf{S}_k x = \varsigma(F_{n,k})x \overset{d}{=} \varsigma(F_{n,k} \circ \sigma_k) = \varsigma(\sigma_k)\mathsf{S}_k x$ and similarly $\mathsf{S}_k x = \mathsf{S}_k \sigma_n x$. □

In addition to permutation symmetry in each fixed dimension, we can also relate vectors and the distributions of their samples to each other across dimensions. We do so using two distinguished maps between finite sets. The first is the inclusions

$$\iota_{N,n} \colon [n] \hookrightarrow [N], \quad \text{whenever } n \leq N,$$

which define inclusions of $\mathcal{A}_n$ into $\mathcal{A}_N$. Then the map $\varsigma(\iota_{N,n}) \colon \mathbb{V}_N \to \mathbb{V}_n$ acts as the orthogonal projection with respect to the usual inner product. For example, when $\mathcal{A}_n = [n]$ the projection $\varsigma(\iota_{N,n})$ extracts the first n entries from a length-N vector, and when $\mathcal{A}_n = [n]^2$ the projection $\varsigma(\iota_{N,n})$ extracts the induced subgraph on the first n vertices.

The second distinguished set of maps are the equipartitions[4] $d_{n,N} \colon [N] \to [n]$ defined by

$$d_{n,N}(j + (i-1)(N/n)) = i \quad \text{for } i \in [n], j \in [N/n], \quad \text{whenever } n|N. \tag{6.12}$$

The corresponding linear maps $\varsigma(d_{n,N}) \colon \mathbb{V}_n \to \mathbb{V}_N$ are embeddings that duplicate the entries of their inputs in a suitable sense. For example, we have $\varsigma(d_{n,N})x = x \otimes \mathbb{1}_{N/n}$ when $\mathbb{V}_n = \mathbb{R}^n$, which acts by duplicating each entry of a vector x of length n the same number N/n of times. Similarly, we have $\varsigma(d_{n,N})X = X \otimes \mathbb{1}_{N/n}\mathbb{1}_{N/n}^\top$ when $\mathbb{V}_n = \mathbb{R}^{n\times n}$, which acts by duplicating each entry of a matrix X of size $n \times n$ into a block of $(N/n) \times (N/n)$ entries.

Proposition 6.3.4 (Relations Between Dimensions)**.** *Suppose* $(\mathcal{A}_n)$ *is a* FinSet-*index, let* $\mathbb{V}_n = \mathbb{R}^{\mathcal{A}_n}$ *be the associated vector spaces endowed with sampling maps* (6.11).

(Projection) *For any* $n \leq N$ *and* x*, we have* $\varsigma(\iota_{N,n})\mathsf{S}_N x \overset{d}{=} \mathsf{S}_n x$.

(Duplication) *For any* $n|N$ *and* $x \in \mathbb{V}_n$*, we have* $\mathsf{S}_k\varsigma(d_{n,N})x \overset{d}{=} \mathsf{S}_k\varsigma(d_{n,N})x$ *for all* k *and hence* $d_{\mathrm{samp}}(x, \varsigma(d_{n,N})x) = 0$.

Proof. For the first claim, fix $n \leq N$, let $m \in \mathbb{N}$ be arbitrary and $x \in \mathbb{V}_m$, and note that $F_{m,N} \circ \iota_{N,n} \overset{d}{=} F_{m,n}$ so $\varsigma(\iota_{N,n})\mathsf{S}_N x = \varsigma(F_{m,N} \circ \iota_{N,n})x \overset{d}{=} \mathsf{S}_n x$. For the second claim, fix $n|N$, let $k \in \mathbb{N}$ be arbitrary, and note that $d_{n,N} \circ F_{N,k} \overset{d}{=} F_{n,k}$ so $\mathsf{S}_k\varsigma(d_{n,N})x = \varsigma(d_{n,N} \circ F_{N,k}) \overset{d}{=} \mathsf{S}_k x$. ◻

We can abbreviate Propositions 6.3.3 and 6.3.4 by saying that:

[4] An equipartition is a map with equal-sized fibers.

- $\varsigma(\phi_{N,n})\mathsf{S}_N x \overset{d}{=} \mathsf{S}_n x$ for any injection $\phi_{N,n}\colon [n] \to [N]$;
- $\mathsf{S}_k \varsigma(\psi_{n,N}) x \overset{d}{=} \mathsf{S}_k x$ and hence $d_{\mathrm{samp}}(x, \varsigma(\psi_{n,N})x) = 0$ for any equipartition $\psi_{n,N}\colon [n] \to [N]$,

since we can write $\phi_{N,n} = \sigma_N \circ \iota_{N,n}$ and $\psi_{n,N} = d_{n,N}\sigma_N$ for some permutation $\sigma_N \in \mathsf{S}_N$. In particular, if $f\colon \bigsqcup_n \mathbb{V}_n \to \mathbb{R}$ is continuous with respect to d_{samp}, then

$$f(\varsigma(\psi_{n,N})x) = f(x) \quad \text{for all } x \in \mathbb{V}_n \text{ and equipartitions } \psi_{n,N}\colon [N] \to [n],$$

or equivalently, each restriction $f|_{\mathbb{V}_n}$ is permutation-invariant and f is unchanged by duplication of entries defined by $\varsigma(d_{n,N})$.

Approximation and Generalization Rates

In this section, we fix a FinSet-index $(\mathcal{A}_n)$, and consider the associated vector spaces $(\mathbb{V}_n = \mathbb{R}^{\mathcal{A}_n})$ and sampling-with-replacement maps $(\mathsf{S}_k\colon \bigsqcup_n \mathbb{V}_n \to \mathbb{V}_k)$ defined by (6.11). We also fix a sequence of compact sets (Ω_n) closed under sampling, such as the sequence of hypercubes $\Omega_n = [-r, r]^{\mathcal{A}_n}$.

We proceed to analyze the rates at which arbitrary limit objects can be approximated by finite-dimensional ones. We do so by proving that $\mathbb{E} d_{\mathrm{samp}}(x, \mathsf{S}_n x) \to 0$ as $n \to \infty$ and proving the following rate at which this expectation goes to zero.

Theorem 6.3.5. *Suppose $(\mathcal{A}_n)$ has degree D, and suppose $(\Omega_n \subseteq [-r, r]^{\mathcal{A}_n})$ is closed under sampling. Then for any $x \in \overline{\Omega}_\infty$ and $k \leq n$ such that $|\mathcal{A}_k| > 2$, we have*

$$\mathbb{E}_{\widehat{S}_n} W_1(\mathsf{S}_k x, \mathsf{S}_k \widehat{S}_n x) \leq \frac{k(k-1)}{n} r + 208 r \left(\frac{k^2 r^2}{n}\right)^{1/|\mathcal{A}_k|}$$

and $W_1(\mathsf{S}_k x, \mathsf{S}_k \widehat{S}_n x)$ is $\frac{4k^2 r^2}{n}$-subgaussian. Furthermore, we have

$$\mathbb{E} d_{\mathrm{samp}}(x, \widehat{S}_n x) \leq C r \exp\left[-(M^{-1} \log_2 n)^{\frac{1}{1+D}}\right],$$

where $C > 0$ is a universal constant, M is as in Definition 6.3.2, and $d_{\mathrm{samp}}(x, \widehat{S}_n x)$ is $\frac{16 r^2}{n}$-subgaussian.

The proof is based on the observation that a uniformly random map $F_{N,k}\colon [k] \to [N]$ can be identified with k iid uniformly distributed random indices $(F_{N,k}(1), \ldots, F_{N,k}(k))$ from $[N]$, so $\mathrm{Law}(F_{N,k}) = \mu^{\otimes k}$ where μ is the uniform distribution on $[N]$. Meanwhile, if we fix a realization $\widehat{F}_{N,n}$ of $F_{N,n}$, we have $\mathrm{Law}(\widehat{F}_{N,n} \circ F_{n,k}) = \mu_n^{\otimes k}$ where

$\mu_n = \frac{1}{n}\delta_{\widehat{F}_{N,n}(i)}$ is an empirical measure obtained by sampling n iid points from μ. We give the full proof in Section 6.3 below.

Theorem 6.3.5 has two important consequences. First, it shows that $\mathbb{E}d_{\mathrm{samp}}(x, \mathsf{S}_n x) \to 0$ for any $x \in \overline{\Omega}_\infty$, so Proposition 6.2.8 applies and gives us a characterization of data distributions obtained by sampling from random limit objects.

Corollary 6.3.6 (Projection-Consistent Distributions)**.** *Suppose $(\mathcal{A}_n)$ has finite degree, let $\mathbb{V}_n = \mathbb{R}^{\mathcal{A}_n}$, and suppose $\Omega_n \subseteq \mathbb{V}_n$ is compact and closed under sampling. Then the following are equivalent for a sequence of probability distributions $(\mu_n \in \mathcal{P}(\Omega_n))$:*

(Projection consistency) *We have $\varsigma(\phi_{N,n})\mu_N = \mu_n$ whenever $\phi_{N,n} \colon [n] \to [N]$ is injective.*

(Sampling representation) *There exists a distribution $\mu_\infty \in \mathcal{P}(\overline{\Omega}_\infty)$ such that $\mu_n = \mathrm{Law}(\mathsf{S}_n X)$ for all n, where $X \sim \mu_\infty$ is independent of S_n.*

Moreover, the measure μ_∞ above is unique, and sequences of the form $(\mu_n = \mathrm{Law}(\mathsf{S}_n x))$ for deterministic $x \in \overline{\Omega}_\infty$ are extremal in the set of all such sequences.

Proof. Propositions 6.3.3 and 6.3.4 show that if $\mu_n = \mathrm{Law}(\mathsf{S}_n X)$ for random X, then (μ_n) is projection-consistent. Conversely, if (μ_n) is projection-consistent then by [138, Thm. 4.15] there is a sequence $\nu_i \in \mathcal{P}(\Omega_\infty)$ such that $\mathrm{Law}(\mathsf{S}_n X_i) \xrightarrow{i\to\infty} \mu_n$ weakly for each n where $X_i \sim \nu_i$ is independent of S_n. This implies that the sequence (ν_i) converges weakly with respect to d_{samp}, and $\mu_\infty = \lim_i \nu_i \in \mathcal{P}(\overline{\Omega}_\infty)$ then satisfies the desired sampling representation. The last claim follows by Proposition 6.2.8, which applies by Theorem 6.3.5. □

We give a few examples of the data distributions that can be obtained by sampling from limit objects as above.

Example 6.3.7 (Sampling Vectors with Replacement)**.** *Suppose $\mathbb{V}_n = \mathbb{R}^{d\times n}$ and $\Omega_n = [0,1]^{d\times n}$. As discussed in Example 6.2.4, we have an isomorphism $\overline{\Omega}_\infty \cong \mathcal{P}([0,1]^d)$ induced by $x \mapsto \mathrm{Law}(\mathsf{S}_1 x)$. In particular, in this special case $\mathsf{S}_k x = (\mathsf{S}_1 x)^{\otimes k}$ and $d_{\mathrm{samp}}(x, y)$ and $W_1(\mathsf{S}_k x, \mathsf{S}_k y)$ metrize the same topology. The resulting rate $\mathbb{E}W_1(\mathsf{S}_1 x, \mathsf{S}_1 \widehat{S}_n x) = O(n^{-1/d})$ is known to be optimal [80]. In this case, projection-consistent distributions are precisely infinite exchangeable arrays, and*

Corollray 6.3.6 recovers a version of de Finetti's theorem stating that such arrays are unique mixtures of iid arrays.

Example 6.3.8 (Sampling Vertices with Replacement)**.** *Suppose* $\mathbb{V}_n = \mathbb{R}^{n\times n}$ *and* $\Omega_n = \{0,1\}^{n\times n} \cap \mathbb{S}^n$*. Viewing elements of* Ω_n *as adjacency matrices of simple graphs, limits in* d_{samp} *correspond to the dense graph limits of [20, 147] and can be described by symmetric measurable functions on* $[0,1]^2$ *called graphons. Applying the same covering number bounds, we obtain a rate of convergence of* $O(\exp[-\sqrt{\log_2 n}])$*, faster than the* $O(\exp[-\frac{1}{2}\log\log n])$ *showed in [146]. Furthermore, projection-consistent distributions in this case are precisely the* consistent random graph models *studied in [148]. These correspond to countable random graphs whose distributions invariant under permutations of their vertices, and Corollary 6.3.6 recovers the result of [148].*

More generally, we can apply our bounds to limits of weighted and directed graphs by setting $\Omega_n = [0,1]^{n\times n}$ *for instance. Such limits were studied in [1] and can be described by measure-valued functions on* $[0,1]^2$*. In spite of the increased complexity of the limit objects, our bounds seamlessly apply and yield a convergence rate of* $O(\exp[-(\log n)^{1/3}])$*, again improving on the* $O(\exp[-\log\log n])$ *rate proved in [1].*

The same bound applies not only for graphs but also for pairs of graphs and graph signals on them, for which $\mathbb{V}_n = \mathbb{R}^{n\times n} \oplus \mathbb{R}^{n\times d}$ *and where we view an element* $(A, X) \in \mathbb{V}_n$ *as a graph* A *and an* $\mathbb{R}^d$*-valued function* X *on the vertices of the graph. Likewise, we obtain a rate of* $O(\exp[(-\log n)^{\frac{1}{1+d}}])$ *for approximation of general hypergraphs and hypergraph signals with respect to the sampling metric. Limits of simple hypergraphs have been studied previously in [68], and their description requires functions defined on* $O(2^d)$ *variables.*

In these and other examples, a full description of the limit space $\overline{\Omega}_\infty$ may be involved and has to be derived on a case-by-case basis. Our approach via the sampling metric gives a unified view, with associated non-asymptotic bounds, on all these disparate notions of convergence.

The second consequence of Theorem 6.3.5 is sketching and generalization bounds for functions continuous in sampling metric. While the rates we obtain for general Lipschitz-continuous functions are quite slow, we obtain substantially faster rates by exploiting latent low dimensionality in the functions in question.

Corollary 6.3.9 (Sketching and Generalization Rates)**.** *In the setting of Theorem 6.3.5, consider functions $f, \widehat{f} \colon \bigsqcup_n \Omega_n \to \mathbb{R}$.*

(General) *If f is L-Lipschitz in d_{samp}, then for any $x \in \overline{\Omega}_\infty$ with probability at least $1 - e^{-2\epsilon^2}$ we have*

$$|f(x) - f(\mathsf{S}_n x)| \leq LCr \exp\left[-(M^{-1}\log_2 n)^{\frac{1}{1+D}} \right] + \frac{4rL\epsilon}{\sqrt{n}}.$$

If $\widehat{f}$ is also L-Lipschitz in d_{samp}, then

$$\text{err}_\infty(f, \widehat{f}) \leq \text{err}_n(f, \widehat{f}) + 2LCr \exp\left(-(M^{-1}\log_2 n)^{\frac{1}{1+K}} \right).$$

(Fixed-dimensional Sampling) *Suppose $f(x) = F(\text{Law}(\mathsf{S}_k x))$ where $F \colon \mathcal{P}(\Omega_k) \to \mathbb{R}$ is L-Lipschitz in Wasserstein distance and $|\mathcal{A}_k| > 2$. Then for any $x \in \overline{\Omega}_\infty$ with probability at least $1 - e^{-2\epsilon^2}$ we have*

$$|f(x) - f(\mathsf{S}_n x)| \leq Lr\left[\frac{k(k-1)}{n} + \frac{2kr\epsilon}{\sqrt{n}} + 208\left(\frac{k^2r^2}{n}\right)^{1/|\mathcal{A}_k|} \right].$$

If $\widehat{f}$ has the same form as f above, then

$$\text{err}_\infty(f, \widehat{f}) \leq \text{err}_n(f, \widehat{f}) + 2Lr\left[\frac{k(k-1)}{n} + 208\left(\frac{k^2r^2}{n}\right)^{1/|\mathcal{A}_k|} \right].$$

(Sampling Means) *Suppose $f = \sigma(\mathbb{E} g_k \circ \mathsf{S}_k)$ where each coordinate function of $g_k \colon \Omega_k \to \mathbb{R}^\ell$ is B-bounded and $\sigma \colon \mathbb{R}^\ell \to \mathbb{R}$ is L-Lipschitz with respect to the ℓ_1 norm on $\mathbb{R}^\ell$. Then with probability at least $1 - e^{-2\epsilon^2}$ we have*

$$|f(x) - f(\mathsf{S}_n x)| \leq L_\sigma Bk\ell\left(\frac{k-1}{n} + \frac{2\epsilon}{\sqrt{n}} \right).$$

If $\widehat{f}$ has the same form (possibly with a different L-Lipschitz $\widehat{\sigma}$ and B-bounded $\widehat{g}_k$). Then

$$\text{err}_\infty(f, \widehat{f}) \leq \text{err}_n(f, \widehat{f}) + 2L_\sigma Bk\ell\left(\frac{k-1}{n} + \frac{2}{\sqrt{n}} \right).$$

When $\sigma = \text{id}$, the $2/\sqrt{n}$ term above can be omitted.

All of the above estimates also hold for average error with respect to projection-consistent distributions.

The proof is a direct application of Proposition 6.3.12 and Theorem 6.3.5, and we again defer it to Section 6.3.

Note that the general rates for sketching and generalization in Corollary 6.3.9 are very slow, but they improve substantially if we exploit particular representations of the functions at hand in terms of low-dimensional samples of their inputs. We now illustrate these improved rates on several examples. We begin with polynomial regression.

Corollary 6.3.10 (Polynomials)**.** *Suppose $(\mathcal{A}_n)$ has degree D and that $(\Omega_n \subseteq [-r,r]^{\mathcal{A}_n})$ is a sequence of compact sets closed under sampling, and let $p\colon \bigsqcup_n \Omega_n \to \mathbb{R}$ be a polynomial of degree d unchanged by duplication so $p \circ \varsigma(\psi_{n,N}) = p$ for any equipartition $\psi_{n,N}\colon [N] \to [n]$ and any $n|N$. Then there exists $B > 0$ such that with probability at least $1 - e^{-2\epsilon^2}$, we have*

$$|p(x) - p(\mathsf{S}_n x)| \leq BrDd\left[\frac{Dd-1}{n} + \frac{2\epsilon}{\sqrt{n}}\right].$$

If $\widehat{p}$ is another such polynomial, we have

$$\mathrm{err}_\infty(p, \widehat{p}) \leq \mathrm{err}_n(p, \widehat{p}) + \frac{2BrDd(Dd-1)}{n}.$$

Proof. It was shown in [138, Sec. 5] that there exists a polynomial $g_{Dd} \in \mathbb{R}[\mathbb{V}_{Dd}]$ satisfying $p(x) = \mathbb{E} g_{Dd}(\mathsf{S}_{Dd} x)$ for all x. Setting $B = \sup_{x \in \Omega_{Dd}} |g_{Dd}(x)|$, we obtain both conclusions from Corollary 6.3.9. □

Notably, Corollary 6.3.10 applies to any polynomial function and any finite-degree FinSet-index. Here are a few examples encompassed by the above result.

(Moment Polynomials) When $\mathbb{V}_n = \mathbb{R}^{d\times n}$, polynomials unchanged by duplication are precisely moment polynomials $p(x) = q(\mathbb{E}(\mathsf{S}_1 x)^{\alpha_1}, \ldots, \mathbb{E}(\mathsf{S}_1 x)^{\alpha_\ell})$ where $\alpha_i \in \mathbb{Z}^d_{\geq 0}$, where $\mathbb{E}(\mathsf{S}_1 x)^{\alpha} = \frac{1}{n}\sum_{j=1}^n \prod_{\ell=1}^d (x_i)_\ell^{\alpha_\ell}$ for $\alpha \in \mathbb{Z}^d_{\geq 0}$. and $\Omega_n = \Theta^n$ for compact $\Theta \subseteq \mathbb{R}^d$, we can identify limit objects with measures $\overline{\Omega}_\infty = \mathcal{P}(\Theta)$ as in Example 6.2.4

((Hyper)Graph Densities) When $\mathbb{V}_n = \mathbb{R}^{n\times n}$, polynomials unchanged by duplication are precisely linear combinations of graph homomorphism densities. Specifically, if $H \in \mathbb{N}^{k\times k}$ is a multigraph, we define the homomorphism density of H in $G \in \mathbb{V}_n$ by

$$t(H; G) = \frac{1}{n^k} \sum_{f\colon [k]\to[n]} \prod_{i,j=1}^k G_{f(i),f(j)}^{H_{i,j}}.$$

It is called a homomorphism density because when H and G are undirected simple graph, the value $t(H;G)$ is the fraction of maps between vertex sets $V(H) \to V(G)$ that are graph homomorphisms. Similarly, when $\mathbb{V}_n = (\mathbb{R}^n)^{\otimes d}$ is viewed as the space of adjacency tensors of hypergraphs, polynomials unchanged by duplication correspond to hypergraph densities.

(Graph Signals) When $\mathbb{V}_n = \mathbb{R}^{n\times n} \oplus \mathbb{R}^{n\times d}$, viewed as pairs (G, X) of a graph G and a graph signal X assigning a d-dimensional feature to each of n nodes, then polynomials unchanged by duplication include polynomial graphon neural networks, which are compositions of polynomials of the form

$$(G, X) \mapsto \sigma\left(\sum_{\ell=0}^{L} \frac{1}{n^\ell} G^\ell X \Theta_\ell\right), \quad \text{for polynomial } \sigma \text{ and } \Theta_0, \ldots, \Theta_L \in \mathbb{R}^{d\times d}.$$

More generally, in the language of [141, 142], Corollary 6.3.10 applies to any compatible (i.e., unchanged by duplications and permutations) neural network with polynomial activations.

Next, we turn to nonpolynomial function classes.

Corollary 6.3.11 (Normalized DeepSets)**.** *Suppose $\mathcal{A}_n = [d] \times [n]$ so $\mathbb{V}_n = \mathbb{R}^{d\times n}$, and let $(\Omega_n \subseteq [-r, r]^{d\times n})$ be a sequence of compact sets closed under sampling. Suppose each coordinate function of $\rho\colon \mathbb{R}^d \to \mathbb{R}^\ell$ is B-bounded on Ω_1 and $\sigma\colon \mathbb{R}^\ell \to \mathbb{R}$ is L_σ-Lipschitz in ℓ_1-norm, and consider $f\colon \bigsqcup_n \Omega_n \to \mathbb{R}$ defined by*

$$f(x) = \sigma\left(\frac{1}{n}\sum_{i=1}^{n} \rho(x_i)\right) = \sigma(\mathbb{E}\rho(\mathsf{S}_1 x)), \quad \textit{for } x = (x_1, \ldots, x_n) \in \mathbb{V}_n. \tag{6.13}$$

Then with probability at least $1 - e^{-2\epsilon^2}$ we have

$$|f(x) - f(\mathsf{S}_n x)| \le \frac{2L_\sigma B r \ell \epsilon}{\sqrt{n}},$$

and for any other $\widehat{f}$ of the above form, we have

$$\mathrm{err}_\infty(f, \widehat{f}) \le \mathrm{err}_n(f, \widehat{f}) + \frac{4L_\sigma B r \ell}{\sqrt{n}}.$$

Proof. This directly follows from Corollary 6.3.9. □

When σ and ρ are neural networks, the architecture defined by (6.13) is called normalized DeepSets [30, 142]. In particular, when $\Omega_n = \Theta^n$ for compact $\Theta \subseteq \mathbb{R}^d$, we can extend f in (6.13) to measures $\mu \in \mathcal{P}(\Theta)$ by $f(\mu) = \sigma(\mathbb{E}_\mu \rho)$. This extension was shown in [30] to be Lipschitz-continuous in Wasserstein metric on $\mathcal{P}(\Theta)$, and that functions of the form (6.13) can approximate any continuous function on $\mathcal{P}(\Theta)$ uniformly. These results were shown in [142] to imply $O(n^{-1/d})$ generalization rate. This generalization rate can similarly be derived from Corollary 6.3.9, and Corollary 6.3.11 yields improved $O(n^{-1/2})$ generalization rates for functions of the form (6.13).

Missing Proofs from Section 6.3

We begin by proving Theorem 6.4.3. The following is the key proposition we shall need.

Proposition 6.3.12 (Uniform Approximation for $\mathrm{FinSet}^{\mathrm{op}}$-representations)**.** *Fix a bounded $f\colon \Omega_k \to \mathbb{R}$ and $x \in \overline{\Omega}_\infty$, sample $\widehat{S}_n \sim \mathsf{S}_n$ and define the random variable $Z_f = \mathbb{E} f(\mathsf{S}_k \widehat{S}_n x) - \mathbb{E} f(\mathsf{S}_k x)$ where the expectations are only with respect to S_k, which is independent of $\widehat{S}_n$.*

(Bias) *We have $|\mathbb{E}_{\widehat{S}_n} Z_f| \le \frac{k(k-1)}{n} \|f\|_\infty$ whenever $k \le n$.*

(Concentration) *$Z_f - \mathbb{E} Z_f$ is $\frac{4k^2}{n}\|f\|_\infty^2$-subgaussian.*

Proof. It suffices to prove the claim for $x \in \Omega_N$ for some finite $N \in \mathbb{N}$, since if $x \in \overline{\Omega}_\infty$ then we can write $x = \lim_N x_N$ for $x_N \in \Omega_N$, in which case $\mathrm{Law}(\mathsf{S}_n x) = \lim_N \mathrm{Law}(\mathsf{S}_n x_N)$ for all $n \in \mathbb{N}$. By [208], there is a coupling of a uniformly random map $F_{n,k}$ and of a uniformly random injection $\Phi_{n,k}$ such that $\mathbb{P}[F_{n,k} \neq \Phi_{n,k}] \le \frac{k(k-1)}{2n}$. Furthermore, since $F_{N,n} \circ \phi_{n,k} \overset{d}{=} F_{N,k}$ for any injection $\phi_{n,k}\colon [k] \to [n]$, we have $\varsigma(\Phi_{n,k}) \circ \widehat{S}_n \overset{d}{=} \mathsf{S}_k$ if $\widehat{S}_n$ is independent of $\Phi_{n,k}$. In this case, we have

$$|\mathbb{E} Z_f| = |\mathbb{E} f(\mathsf{S}_k \widehat{S}_n x) - \mathbb{E} f(\varsigma(\Phi_{n,k}) \widehat{S}_n x)| \le 2\|f\|_\infty \mathbb{P}[\mathsf{S}_k \neq \varsigma(\phi_{n,k})] \le \|f\|_\infty \frac{k(k-1)}{n},$$

as claimed.

To see the subgaussianity of Z_f, represent a map $[n] \to [N]$ as a sequence $[n]^N$, and observe that $\mathrm{Law}(F_{N,n}) = \mathrm{Unif}[N]^{\otimes n}$. Writing $\widehat{S}_n = \varsigma(\widehat{F}_{N,n})$, let $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{\widehat{F}_{N,n}(i)}$ be the empirical measure obtained by sampling n iid points from $\mu = \mathrm{Unif}([N])$. Now observe that $\mathrm{Law}(\widehat{F}_{N,n} F_{n,k}) = \mu_n^{\otimes k}$ while $\mathrm{Law}(F_{N,k}) = \mu^{\otimes k}$. Furthermore, if we define $\widetilde{f}\colon [N]^k \to \mathbb{R}$ by $\widetilde{f}(F) = f(\varsigma(f)x)$, then

$$\mathbb{E} f(\mathsf{S}_k \widehat{S}_n x) = \mu_n^{\otimes k}(\widetilde{f}) = \frac{1}{n^k} \sum_{i_1,\ldots,i_k=1}^{n} \widetilde{f}\Big(\Big[\widehat{F}_{N,n}(i_1), \ldots, \widehat{F}_{N,n}(i_k)\Big]\Big), \tag{6.14}$$

where $\|\widetilde{f}\|_\infty \le \|f\|_\infty$. Changing the value $\widehat{F}_{N,n}(i_\ell)$ for any fixed $\ell \in [n]$ changes the value of $\mu_n^{\otimes k}(\widetilde{f})$ by at most $\frac{2k}{n}\|f\|_\infty$, hence $Z_f - \mathbb{E}Z_f$ is $\frac{4k^2}{n}\|f\|_\infty^2$-subgaussian by the bounded-difference inequality [223]. □

We turn to proving Theorem 6.3.5 and a related result bounding total variation distance for binary inputs. These rely on the following extension of Proposition 6.3.12 using a standard covering number bound. In what follows, if $\mu, \nu \in \mathcal{P}(\Theta)$ and $\mathcal{F}$ is a collection of functions on $\mathcal{F}$, we denote by $d_{\mathcal{F}}(\mu, \nu) = \sup_{f\in\mathcal{F}} |\mathbb{E}_\mu f - \mathbb{E}_\nu f|$.

Lemma 6.3.13 (Covering Number Bound)**.** *Instantiate the setting of Proposition 6.3.12 and suppose $\mathcal{F}_k$ is a family of functions on Ω_k with $\sup_{f\in\mathcal{F}_k}\sup_{x\in\Omega_k}|f(x)| \le r$. Then*

$$\mathbb{E}_{\widehat{S}_n} d_{\mathcal{F}_k}(\mathsf{S}_k\widehat{S}_n x, \mathsf{S}_k x) = \mathbb{E}\sup_{f\in\mathcal{F}_k} Z_f \le r\frac{k(k-1)}{n} + \inf_{\delta>0}\left\{4\delta + \frac{16\sqrt{2}kr}{\sqrt{n}}\int_\delta^r \sqrt{\log N(\mathcal{F}_k, \|\cdot\|_\infty, \epsilon)}\, d\epsilon\right\},$$

where $N(\mathcal{F}_k, \|\cdot\|_\infty, \epsilon)$ is the ϵ-covering number of the function class $\mathcal{F}_k$ in $\|\cdot\|_\infty$. Moreover, the random variable $\sup_{f\in\mathcal{F}_k} Z_f$ is $\frac{4k^2r^2}{n}$-subgaussian.

Proof. Since $Z_f - Z_g = Z_{f-g}$, we conclude that $\{Z_f - \mathbb{E}Z_f\}_{f\in\mathcal{F}_k}$ is a centered $\frac{4k^2r^2}{n}$-subgaussian process with respect to the uniform metric. Since the diameter of $\mathcal{F}_k$ under the uniform metric is 1 by assumption, Dudley's entropy integral bound [12] gives

$$\mathbb{E}\sup_{f\in\mathcal{F}_k}(Z_f - \mathbb{E}Z_f) \le \inf_{\delta>0}\left\{4\delta + \frac{16\sqrt{2}kr}{\sqrt{n}}\int_\delta^{1/2}\sqrt{\log N(\mathcal{F}_k, \|\cdot\|_\infty, \epsilon)}\, d\epsilon\right\}.$$

Combining the above bound with our bias bound from Proposition 6.3.12 yields the third claim.

Finally, observe that Z_f is a function of n iid random indices $\widehat{F}_{N,n}(1), \ldots, \widehat{F}_{N,n}(n)$, and that changing the value of one of these indices affects at most $n^k - (n-1)^k \le kn^{k-1}$ terms in (6.14) by at most $2\|f\|_\infty/n^k \le 2r/n^k$. Therefore, each Z_f is a $\frac{2kr}{n}$-Lipschitz function with respect to the Hamming distance on $[N]^n$, and hence the same is true of $\sup_{f\in\mathcal{F}_k} Z_f$. The bounded difference inequality [223, Cor. 2.21] yields the final claim. □

Combining Lemma 6.3.13 with covering number bounds for Lipschitz functions yields Theorem 6.3.5.

Proof (Theorem 6.3.5). Set $\mathcal{F}_k$ to be the collection of all functions $f : \Omega_k \to \mathbb{R}$ with $f(0) = 0$ that are 1-Lipschitz with respect to the ℓ_∞-norm that vanish at the origin, and note that $\sup_{f\in\mathcal{F}_k} \|f\|_\infty \leq r$ since $\Omega_k \subseteq [-r, r]^{\mathcal{A}_k}$. Thus, by Lemma 6.3.13

$$\mathbb{E}_{\widehat{S}_n} W_1(\mathsf{S}_k x, \mathsf{S}_k \widehat{S}_n x) = \mathbb{E} \sup_{f\in\mathcal{F}_k} Z_f \leq r\frac{k(k-1)}{n} + \inf_{\delta>0}\left\{4\delta + \frac{16\sqrt{2}kr}{\sqrt{n}}\int_\delta^r \sqrt{\log N(\mathcal{F}_k, \|\cdot\|_\infty, \epsilon)}\, d\epsilon\right\}.$$

By [126, §9] (or by [150, Thm. 17]), we have

$$\log N(\mathcal{F}_k, \|\cdot\|_\infty, \epsilon) \leq (\log 2) N(\Omega_k, \|\cdot\|_\infty, \epsilon/2) + \log\left(2\left\lceil\frac{2r}{\epsilon}\right\rceil + 1\right).$$

Since $\Omega_k \subseteq [-r, r]^{\mathcal{A}_k}$, we have $N(\Omega_k, \|\cdot\|_\infty, \epsilon/2) \leq \lceil 4r/\epsilon\rceil^{|\mathcal{A}_k|}$ and hence $\log N(\mathcal{F}_k, \|\cdot\|_\infty, \epsilon) \leq (1 + \log 2)\left(\frac{5r}{\epsilon}\right)^{|\mathcal{A}_k|}$ for all $\epsilon \in (0, 2r)$. Since $|\mathcal{A}_k| > 2$, we have

$$4\delta + \frac{16\sqrt{2}kr}{\sqrt{n}}\int_\delta^r \sqrt{\log N(\mathcal{F}_k, \|\cdot\|_\infty, \epsilon)}\, d\epsilon \leq 4\delta + 16\sqrt{2(1+\log 2)}\frac{kr}{\sqrt{n}}(5r)^{|\mathcal{A}_k|/2}\frac{\delta^{-|\mathcal{A}_k|/2+1}}{|\mathcal{A}_k|/2 - 1}.$$

Choosing $\delta = 5r\left(4\frac{kr}{\sqrt{n}}\sqrt{2(1+\log 2)}\right)^{2/|\mathcal{A}_k|}$, we obtain the claimed estimate $208r\left(\frac{k^2r^2}{n}\right)^{1/|\mathcal{A}_k|}$. Proposition 6.3.12 also shows that $\sup_{f\in\mathcal{F}_k} Z_f$ is $\frac{4k^2r^2}{n}$-subgaussian.

Turning to the claim bound on d_{samp}, for any $K \leq n$ we have

$$\mathbb{E}d_{\text{samp}}(x, \mathsf{S}_n x) \leq \sum_{k=1}^K 2^{-k}\mathbb{E}W_1(\mathsf{S}_k x, \mathsf{S}_k \widehat{S}_n x) + 2^{-K+1}r \leq Cr\left(\frac{r^2}{n}\right)^{\frac{1}{MK^D}} + 2^{-K}r$$

obtained by estimating $\mathbb{E}W_1(\mathsf{S}_k x, \mathsf{S}_k \widehat{S}_n x) \leq 2r$ for $k \geq K$ and applying the upper bounds on $\mathbb{E}W_1$ from the first part of this corollary. Choosing $K = \left\lceil\left(\frac{\log_2 n}{M}\right)^{\frac{1}{D+1}}\right\rceil$ if $r \leq 1$ and $K = \left\lceil\left(\frac{\log_2(n/r^2)}{M}\right)^{\frac{1}{D+1}}\right\rceil$ gives the desired bound. Finally, since d_{samp} is a function of n independent random variables and is $\sum_{k\geq 1} 2^{-k}\frac{2kr}{n} = \frac{4r}{n}$-Lipschitz in Hamming distance in each of them, the bounded-difference inequality yields the claimed $\frac{16r^2}{n}$-subgaussianity for $d_{\text{samp}}(x, \mathsf{S}_n x)$. □

Next, we prove the sketching and generalization bounds in Corollary 6.3.9.

Proof (Corollary 6.3.9). For general L-Lipschitz $f, \widehat{f}$, for any $x \in \overline{\Omega}_\infty$ we have $|f(x) - f(\mathsf{S}_n x)| \leq L d_{\text{samp}}(x, \mathsf{S}_n x)$ and

$$|f(x) - \widehat{f}(x)| \leq |f(x) - \mathbb{E}f(\mathsf{S}_n x)| + |\mathbb{E}[f(\mathsf{S}_n x) - \widehat{f}(\mathsf{S}_n x)]| + |\mathbb{E}\widehat{f}(\mathsf{S}_n x) - \widehat{f}(x)|. \quad (6.15)$$

Now observe that $|f(x) - \mathbb{E}f(\mathsf{S}_n x)| \leq L\mathbb{E}d_{\text{samp}}(x, \mathsf{S}_n x)$ and similarly for $\widehat{f}$, while $|\mathbb{E}[f(\mathsf{S}_n x) - \widehat{f}(\mathsf{S}_n x)]| \leq \text{err}_n(f, \widehat{f})$. The bounds for general Lipschitz functions now follow from Theorem 6.3.5.

If $f(x) = F(\text{Law}(\mathsf{S}_k x))$ for L-Lipschitz F, we have $|f(x)-f(\mathsf{S}_n x)| \leq LW_1(\mathsf{S}_k x, \mathsf{S}_k\mathsf{S}_n x)$ where we fix a realization of S_n, and $|f(x)-\mathbb{E}f(\mathsf{S}_n x)| \leq L\mathbb{E}W_1(\mathsf{S}_k x, \mathsf{S}_k\mathsf{S}_n x)$. Since $\widehat{f}$ satisfies the same bounds, we obtain the claimed rates from Theorem 6.3.5 using (6.15) again.

Finally, suppose $f(x) = \sigma(\mathbb{E}g_k(\mathsf{S}_k x))$ and note that $|f(x)-f(\mathsf{S}_n x)| \leq L_\sigma \|\mathbb{E}_{\mathsf{S}_k} g_k(\mathsf{S}_k x) - \mathbb{E}_{\mathsf{S}_k} g_k(\mathsf{S}_k\mathsf{S}_n x)\|_1 = \sum_{j=1}^{\ell} |\mathbb{E}_{\mathsf{S}_k}[g_k]_j(\mathsf{S}_k x) - \mathbb{E}_{\mathsf{S}_k}[g_k]_j(\mathsf{S}_k\mathsf{S}_n x)|$ where we fix a realization of S_n and denote by $[g_k]_j$ the jth coordinate function of g_k. Similarly, we have $|f(x) - \mathbb{E}f(\mathsf{S}_n x)| \leq L_\sigma \mathbb{E}_{\mathsf{S}_n} \sum_{j=1}^{\ell} |\mathbb{E}_{\mathsf{S}_k}[g_k]_j(\mathsf{S}_k x) - \mathbb{E}_{\mathsf{S}_k}[g_k]_j(\mathsf{S}_k\mathsf{S}_n x)|$. Since $\widehat{f}$ satisfies the same bounds, we obtain the claimed rates from Proposition 6.3.12 using (6.15) again. If $\sigma = \text{id}$, we can improve the generalization bound because $|f(x)-\mathbb{E}f(\mathsf{S}_n x)| \leq \|\mathbb{E}g_k(\mathsf{S}_k x)-\mathbb{E}g_k(\mathsf{S}_k\mathsf{S}_n x)\|_1$ where both expectations are inside the norm, and hence we only incur an error due to the bias in Proposition 6.3.12. □

We now prove Corollary 6.1.7 using the measure-theoretic expression for attention in [82].

Proof (Corollary 6.1.7). Consider the in-context attention mapping $\bar{A}\colon \mathbb{R}^d \times \mathcal{P}_c(\mathbb{R}^d) \to \mathbb{R}^d$ introduced in [82] and given by

$$\bar{A}(x, \mu) = x + \frac{\mathbb{E}_{Y\sim\mu} \exp(\langle Qx, KY\rangle)VY}{\mathbb{E}_{Y\sim\mu} \exp(\langle Qx, KY\rangle)},$$

which satisfies

$$\mathsf{S}_1 A(x) \stackrel{d}{=} \bar{A}(\mathsf{S}_1 x, \text{Law}(\mathsf{S}_1 x)).$$

Defining $\mathcal{A}\colon \mathcal{P}_c(\mathbb{R}^d) \to \mathcal{P}_c(\mathbb{R}^d)$ by $\mathcal{A}(\mu) = \text{Law}(\bar{A}(X, \mu))$ where $X \sim \mu$, letting ϕ act on measures via pushforward, and defining

$$\bar{F}(\mu) = \mathbb{E}_{X\sim(\phi\circ\mathcal{A})^{\circ k}(\mu)} X,$$

we get $F(x) = \bar{F}(\text{Law}(\mathsf{S}_1 x))$ as desired.

Let $R = \sup_{j\leq k} \sup_{x\in\Theta} \|(\phi \circ A)^{\circ j}(x)\|_\infty < \infty$ and denote $B_R = \{x \in \mathbb{R}^d : \|x\|_\infty \leq R\}$. We can find Lipschitz constants $L_1, L_2 > 0$ such that $\bar{A}(x, \cdot)\colon \mathcal{P}(B_R) \to \mathbb{R}^d$ is L_1-Lipschitz while $A(\cdot, \mu)\colon B_R \to \mathbb{R}^d$ is L_2-Lipschitz for any $x \in B_R$ and $\mu \in \mathcal{P}(B_R)$, see [83, Lemma 2] for example. For any $x, y \in B_R^n$, we then have

$$W_1(\mathsf{S}_1 A(x), \mathsf{S}_1 A(y)) \leq (L_1 + L_2)W_1(\mathsf{S}_1 x, \mathsf{S}_1 y).$$

Likewise, we have $W_1(\mathsf{S}_1\phi(x), \mathsf{S}_1\phi(y)) \leq L_\phi W_1(\mathsf{S}_1 x, \mathsf{S}_1 y)$. Thus, we get

$$W_1(\mathsf{S}_1(\phi\circ A)^{\circ k}(x), \mathsf{S}_1(\phi\circ A)^{\circ k}(y)) \leq L_\phi^k(L_1+L_2)^k W_1(\mathsf{S}_1 x, \mathsf{S}_1 y),$$

so $\bar{F}$ is $L_\phi^k(L_1+L_2)^k$-Lipschitz. The result now follows from Corollary 6.3.9. □

6.4 Random Binning

In this section, we define generalizations for random binning and entry-biased sampling maps from Examples 6.2.6-6.2.7, and show that they lead to the same topology. By exploiting the compatibility between the two notions of sampling, we prove universal rates for approximation and generalization for the corresponding sampling metrics.

Random Binnings as Dual Sampling

We generalize random binning as in Example 6.2.6 by again using maps between finite sets. Specifically, any map $f\colon [n] \to [k]$ defines a binning of n items into k bins, and acts on a vector $x \in \mathbb{R}^n$ by binning its coordinates and summing all coordinates in a given bin, i.e., $(\beta(f)x)_i = \sum_{j\in f^{-1}(i)} x_j$. Observe that this action of maps between finite sets is dual to the one from Section 6.3, in the sense that $\varsigma(f) = \beta(f)^\top$ as matrices. In this case, if $F_{k,n}\colon [n] \to [k]$ is a uniformly random map then $\beta(F_{k,n})x$ is the uniformly random binning of the coordinates of x considered in Example 6.2.6. Once again, we shall generalize this notion of random binning beyond vectors.

If $(\mathcal{A}_n)$ is a FinSet-index and $\mathbb{V}_n = \mathbb{R}^{\mathcal{A}_n}$ is the associated sequence of vector spaces, we define another linear action of maps between finite sets on these vectors by

$$(\beta(f_{k,n})x)_\alpha = \sum_{\beta\in\rho(f_{k,n})^{-1}(\alpha)} x_\beta, \quad \text{whenever } x\in\mathbb{V}_n \text{ and } f_{k,n}\colon [n]\to[k]. \tag{6.16}$$

Note that a map $f_{k,n}\colon [n] \to [k]$ defines two linear maps, one $\beta(f_{k,n})\colon \mathbb{V}_n \to \mathbb{V}_k$ as in (6.16) and one $\varsigma(f_{k,n})\colon \mathbb{V}_k \to \mathbb{V}_n$ as in (6.10). These two maps are adjoints of each other $\beta(f_{k,n}) = \varsigma(f_{k,n})^\star$ with respect to the usual inner products on $\mathbb{R}^{\mathcal{A}_n}$.

We can then define sampling maps $\mathsf{S}_k\colon \bigsqcup_n \mathbb{V}_n \to \mathbb{V}_k$ by applying random maps between finite sets

$$\mathsf{S}_k x = \beta(F_{k,n})x, \quad \text{if } x\in\mathbb{V}_n. \tag{6.17}$$

Once again, symmetries and relations between dimensions play an important role in analyzing this notion of sampling. The proofs of Propositions 6.3.3 and 6.3.4 show that

- $\beta(\psi_{n,N})\mathsf{S}_N x \overset{d}{=} \mathsf{S}_n x$ for any equipartition $\psi_{n,N} \colon [N] \to [n]$;
- $\mathsf{S}_k \beta(\phi_{N,n}) x \overset{d}{=} \mathsf{S}_k x$ and hence $d_{\mathrm{samp}}(x, \beta(\phi_{N,n})x) = 0$ for any injection $\phi_{N,n} \colon [n] \to [N]$.

For the present binning action, an equipartition acts by summing all the coordinates in each part of the partition. For example, if $\mathbb{V}_n = \mathbb{R}^n$ the standard equipartition (6.12) acts via $\beta(d_{n,N})$ by summing each set of N/n consecutive coordinates, and if $\mathbb{V}_n = \mathbb{R}^{n\times n}$ then it sums each consecutive $N/n \times N/n$ block. In contrast, injections act by zero-padding. For example, if $\mathbb{V}_n = \mathbb{R}^n$ the standard inclusion $\beta(\iota_{N,n})$ zero-pads a vector of length n by $N-n$ trailing zeros, and if $\mathbb{V}_n = \mathbb{R}^{n\times n}$ then $\beta(\iota_{N,n})$ zero-pads an $n \times n$ matrix into a $N \times N$ one.

In particular, we have $d_{\mathrm{samp}}(x, y) = 0$ whenever y differs from x by permutations and zero-padding, and if $f \colon \bigsqcup_n \mathbb{V}_n \to \mathbb{R}$ is continuous with respect to d_{samp}, then f is likewise invariant under permutations and zero-padding.

Approximation and Generalization Rates

In this section, we fix a FinSet-index $(\mathcal{A}_n)$, and consider the associated vector spaces $(\mathbb{V}_n = \mathbb{R}^{\mathcal{A}_n})$ and random binning maps $(\mathsf{S}_k \colon \bigsqcup_n \mathbb{V}_n \to \mathbb{V}_k)$ defined by (6.17). We also fix a sequence of compact sets (Ω_n) closed under random binning, such as the sequence of simplices $\Omega_n = \Delta^{\mathcal{A}_n - 1}$.

We turn to proving rates for approximation and generalization with respect to the sampling metric induced by random binning. While we again seek to prove that $\mathbb{E} d_{\mathrm{samp}}(x, \mathsf{S}_n x) \to 0$ and quantify the rate at which it converges to zero, it is considerably more challenging to do so by analyzing random binning alone as it does not concentrate as well as sampling with replacement.

Example 6.4.1 (Anti-concentration for random binning). *Suppose $\mathbb{V}_n = \mathbb{R}^n$ and consider $x = (1/2, 1/2) \in \mathbb{V}_2$. Observe that $\mathsf{S}_n x$ equals one of the standard basis vectors with probability $1/n$, since this is the probability for the two entries of x to map to the same bin. Therefore, we have $d_{\mathrm{samp}}(x, \mathsf{S}_n x) = d_{\mathrm{samp}}(x, e_1) = \sum_{k\geq 2} 2^{-k}(1 - \frac{1}{k})\frac{1}{2} = \frac{1-\log 2}{2}$ with probability $1/n$. This implies that if we sample $x_n \sim \mathsf{S}_n x$ independently for different n, then $d_{\mathrm{samp}}(x, x_n) \geq \frac{1-\log 2}{2}$ for infinitely-many n almost surely by Borel–Cantelli, and hence $x_n \not\to x$ almost surely. Borel–Cantelli further implies that for any rate $(R_n \geq 0)$ with $R_n \downarrow 0$ we have $\sum_{n\geq 1} \mathbb{P}[d_{\mathrm{samp}}(x, \mathsf{S}_n x) \geq R_n] = \infty$.*

Contrast Example 6.4.1 with Theorem 6.3.5 for sampling with replacement, which gives a rate R_n such that $\mathbb{P}[d_{\text{samp}}(x, \mathsf{S}_n x) \geq R_n] \leq e^{-n/8}$.

To circumvent the above difficulties, we introduce a second sampling map generalizing the one in Example 6.2.7, for FinSet-representations of some degree D.

Suppose first that $x \in \Delta^{\mathcal{A}_N - 1}$ is a simplex element. We view such x as a probability distribution over $\mathcal{A}_N$, and consider forming an empirical approximation to it from n iid samples. Specifically, sampling $\alpha_1, \ldots, \alpha_n$ iid according to x, we set

$$\mathscr{E}_n(x) = \frac{1}{n} \sum_{i=1}^{n} e_{\alpha_i}. \tag{6.18}$$

More generally, if $x \in \mathbb{R}^{\mathcal{A}_N}$ is viewed as a signed measure, we set $\mathscr{E}_n(x) = \|x\|_1 \operatorname{sign}(x) \odot \mathscr{E}_n(|x|/\|x\|_1)$ where $\odot$ denotes entrywise product (and $\mathscr{E}_n(0) = 0$ by convention). In general, we have $\mathscr{E}_n(x) \in \mathbb{V}_N$ and hence $\mathscr{E}_n$ does not reduce the dimensionality of its input. However, it does produce a sparse input, as $\mathscr{E}_n(x)$ has at most n nonzero entries regardless of the dimensionality of x. We exploit this fact to prove that when $(\mathcal{A}_n)$ has degree D, we have $\operatorname{dist}(\mathscr{E}_n(x), \mathbb{V}_{Dn}) = 0$ almost surely in sampling metric, and hence $\mathscr{E}_n(x)$ is equivalent to a low-dimensional element.

Proposition 6.4.2. *Suppose $(\mathcal{A}_n)$ has degree D. Then there exist sampling maps $\mathcal{S}_k \colon \bigsqcup_n \mathbb{V}_n \to \mathbb{V}_{kD}$ such that for any $x \in \mathbb{V}_N$ there is a random injection $\Phi_{N,nD} \colon [nD] \to [N]$ satisfying*

$$\mathscr{E}_n(x) = \beta(\Phi_{N,nD}) \mathcal{S}_n(x).$$

The proof is explicit and follows from the definition of a degree of a FinSet-index, we defer it to Section 6.4. Since $\mathsf{S}_k \beta(\phi) \stackrel{d}{=} \mathsf{S}_k$ for any injection ϕ, we conclude that

$$d_{\text{samp}}(\mathscr{E}_n(x), \mathcal{S}_n(x)) = 0, \quad \text{almost surely.}$$

The construction of $\mathcal{S}_n$ from Proposition 6.4.2 is explicit. For example, if $\mathcal{A}_n = [n]$ then we recover the sampling map (6.9) defining random partition models. If $\mathcal{A}_n = [n]^2$ and we view elements of $\mathbb{V}_n = \mathbb{R}^{n \times n}$ as adjacency matrices of graphs on n vertices, then $\mathcal{S}_n$ can be viewed as sampling edges from a weighted graph proportionally to their edge weights, and has been studied in connection with limits of large graphs in [140].

We turn to proving rates for the above nonlinear sampling maps with respect to the sampling metric defined by random binning.

Theorem 6.4.3. *Suppose $(\mathcal{A}_n)$ is a* FinSet*-index of degree D, and consider the sampling metric defined using random binning (S_k) and the W_1-metric with respect to the ℓ_2 norm. Fix $\Omega_n \subseteq \mathbb{V}_n = \mathbb{R}^{\mathcal{A}_n}$ closed under both binning and empirical sampling such that $\sup_{n\in\mathbb{N}} \sup_{x\in\Omega_n} \|x\|_1 \leq r$.*

1. *Whenever $(x_n) \subseteq \Omega_\infty$ converges in d_{samp}, the sequence of distributions $(\mathcal{S}_k(x_n)) \subseteq \Omega_{Dk}$ converges weakly. Thus, the sampling map $\mathcal{S}_n$ extends to all of $\overline{\Omega}_\infty$.*

2. *For any $x \in \overline{\Omega}_\infty$, we have*
$$\mathbb{E}_{\mathcal{S}_n(x)} W_1(\mathsf{S}_k x, \mathsf{S}_k \mathcal{S}_n(x)) \leq \frac{r}{\sqrt{n}},$$
and $\mathbb{E} d_{\text{samp}}(x, \mathcal{S}_n(x)) \leq r/\sqrt{n}$. Moreover, both $W_1(\mathsf{S}_k x, \mathsf{S}_k \mathcal{S}_n(x))$ and $d_{\text{samp}}(x, \mathcal{S}_n(x))$ are $\frac{4r^2}{n}$-subgaussian.

3. *For any $x \in \overline{\Omega}_\infty$, we have*
$$\mathbb{E} d_{\text{samp}}(x, \mathsf{S}_n x) \leq \frac{2(1+D^2)r}{n^{1/5}}. \tag{6.19}$$
With probability $1 - 2e^{-\epsilon^2/2} - \frac{D^2}{2n^{1/5}}$, we also have $d_{\text{samp}}(x, \mathsf{S}_n x) \leq \frac{2(1+2\epsilon)r}{n^{1/5}}$.

The proof of Theorem 6.4.3 is based on the standard bound

$$\mathbb{E}\|x - \mathcal{E}_n(x)\|_2 \leq \|x\|_1 \sqrt{\sum_{\alpha\in\mathcal{A}_N} \mathbb{E}\Big(p_\alpha - \tfrac{1}{n}\mathrm{Binom}(n, p_\alpha)\Big)^2} \leq \frac{\|x\|_1}{\sqrt{n}}, \quad \text{for any } x \in \mathbb{R}^{\mathcal{A}_N},\ N \in \mathbb{N}, \tag{6.20}$$

where $p = \frac{|x|}{\|x\|_1}$. We give the full details in Section 6.4. Examples of sets satisfying the hypotheses of Theorem 6.4.3 include the sequence of simplices $\Omega_n = \Delta^{\mathcal{A}_n - 1}$ and ℓ_1-balls.

As before, Theorem 6.4.3 has two implications. The first is that $\mathbb{E} d_{\text{samp}}(x, \mathsf{S}_n x) \to 0$ as $n \to \infty$ so Proposition 6.2.8 applies and gives us a characterization of sequences of distributions obtained by randomly binning limit objects.

Corollary 6.4.4 (Equipartition-Consistent Distributions)**.** *Suppose $(\mathcal{A}_n)$ has finite degree, let $\mathbb{V}_n = \mathbb{R}^{\mathcal{A}_n}$, and suppose $\Omega_n \subseteq \mathbb{V}_n$ is compact and closed under binning. Then the following are equivalent for a sequence of probability distributions $(\mu_n \in \mathcal{P}(\Omega_n))$:*

(Equipartition consistency) *We have* $\beta(\psi_{n,N})\mu_N = \mu_n$ *whenever* $\psi_{n,N}\colon [N] \to [n]$ *has equal-sized fibers.*

(Sampling representation) *There exists a distribution* $\mu_\infty \in \mathcal{P}(\overline{\Omega}_\infty)$ *such that* $\mu_n = \mathrm{Law}(\mathsf{S}_n X)$ *for all n, where* $X \sim \mu_\infty$ *is independent of* S_n*.*

Moreover, the measure μ_∞ *above is unique, and sequences of the form* $(\mu_n = \mathrm{Law}(\mathsf{S}_n x))$ *for deterministic* $x \in \overline{\Omega}_\infty$ *are extremal in the set of all such sequences.*

Proof. Since $\beta(\psi_{n,N})\mathsf{S}_N x \overset{d}{=} \mathsf{S}_n x$ for any x and any equipartition $\psi_{n,N}$, we conclude that if $\mu_n = \mathrm{Law}(\mathsf{S}_n X)$ for random X indpendent of S_n then (μ_n) is equipartition-consistent. Conversely, if (μ_n) is equipartition-consistent then by [138, Thm. 4.20] there is a sequence $(\nu_i \in \mathcal{P}(\Omega_\infty))$ such that $\mathrm{Law}(\mathsf{S}_n X_i) \xrightarrow{i\to\infty} \mu_n$ weakly for each n, where $X_i \sim \nu_i$ is independent of ν_i. This implies that (ν_i) converges weakly to some $\mu_\infty \in \mathcal{P}(\overline{\Omega}_\infty)$ satisfying the claimed sampling representation. The uniqueness of such μ_∞ and the extremality of sequences of the form $(\mathrm{Law}(\mathsf{S}_n x))$ for $x \in \overline{\Omega}_\infty$ both follow from Proposition 6.2.8, which applies by Theorem 6.4.3. ◻

We give some examples of distributions obtained by randomly binning limit objects.

Example 6.4.5. *If* $\mathcal{A}_n = [n]$ *and* $\Omega_n = \Delta^{n-1}$*, then a sequence* $(\mu_n \in \mathcal{P}(\Delta^{n-1}))$ *is equipartition-consistent if and only if (i) each* μ_n *is exchangeable so* $\pi X_n \overset{d}{=} X_n$ *for any* $\pi \in \mathsf{S}_n$ *when* $X_n \sim \mu_n$*; (ii) whenever* $n|N$ *we have*

$$\left(\sum_{j=1}^{N/n} (X_N)_{j+(i-1)(N/n)}\right)_{i=1}^{n} \overset{d}{=} X_n.$$

For example, the sequence of Dirichlet distributions $\mu_n = \mathrm{Dir}(\alpha \mathbb{1}_n/n)$ *satisfies these conditions for any* $\alpha \in [0,\infty)$*, where* $\alpha = 0$ *corresponds to a uniformly random coordinate vector.*

If $\mathcal{A}_n = [n]^2$*, then equipartition-consistent sequences* $(\mu_n \in \mathcal{P}(\Delta^{n\times n}))$ *were called "equipartition-consistent random graph models" in [140], where they were shown to correspond to certain limits of growing-sized graphs. In particular, the specialization of Corollary 6.4.4 to this case was proved in [140, Thm. 1.9].*

Remark 6.4.6. *Theorem 6.4.3 also implies that the sampling maps* $(\mathcal{S}_k)$ *induce the same topology as random binning* (S_k)*. Indeed, part 1 of the above theorem shows that for a given sequence* $(x_n) \subseteq \Omega_\infty$*, if* $(\mathsf{S}_k x_n)_n$ *converges weakly for each k then*

$(\mathcal{S}_k(x_n))_n$ converges weakly for each k, hence the closure of Ω_∞ with respect to (S_k) is a subset of its closure with respect to $(\mathcal{S}_k)$. Since both closures are compact and contain the dense subset Ω_∞, they must be equal. A special case of this equivalence was shown in [140, Prop. 4.1] for $\mathcal{A}_n = [n]^2$. In particular, Proposition 6.2.8 applies to the sampling maps $(\mathcal{S}_k)$ as well, though the characterization of sequences of the form $(\mathrm{Law}(\mathcal{S}_k(X)))$ for random X is more involved.

The second consequence of Theorem 6.4.3 is the following generalization and sketching bounds.

Corollary 6.4.7. *Let $(\mathbb{V}_n)$ be a* FinSet*-representation of finite degree and $\Omega_n \subseteq B_{\ell_1}^{(n)}(r)$ be closed under sampling where $B_{\ell_1}^{(n)}(r) = \{x \in \mathbb{V}_n : \|x\|_1 \leq r\}$. Suppose $f\colon \bigsqcup_n \Omega_n \to \mathbb{R}$ is L-Lipschitz in either $(x,y) \mapsto W_1(\mathsf{S}_k x, \mathsf{S}_k y)$ or d_{samp}. Then with probability at least $1 - e^{-2\epsilon^2}$ we have*

$$|f(x) - f(\mathcal{S}_n(x))| \leq \frac{Lr(1+2\epsilon)}{\sqrt{n}}. \tag{6.21}$$

If $\widehat{f}$ is another such L-Lipschitz function, then

$$\mathrm{err}_\infty(f, \widehat{f}) \leq \mathrm{err}_n(f, \widehat{f}) + \frac{2Lr}{\sqrt{n}}. \tag{6.22}$$

The above bound also holds for average error with respect to distributions of the form $(\mu_n = \mathrm{Law}(\mathcal{S}_n(X)))$ for random $X \in \overline{\Omega}_\infty$ independent of S_n. In contrast, for average error with respect to an equipartition-consistent sequence we have

$$\mathrm{err}_{\mu_\infty}(f, \widehat{f}) \leq \mathrm{err}_{\mu_n}(f, \widehat{f}) + \frac{4L(1+D^2)r}{n^{1/5}}.$$

As in Section 6.3, a prominent example of functions satisfying the above hypotheses are polynomials.

Corollary 6.4.8. *Suppose $(\mathcal{A}_n)$ has finite degree and that $(\Omega_n \subseteq B_{\ell_1}^{(n)}(r))$ is a sequence of compact sets closed under binning, and let $p\colon \bigsqcup_n \Omega_n \to \mathbb{R}$ be a polynomial of degree d unchanged by zero-padding, so $p \circ \beta(\iota_{N,n}) = p$ for all injections $\iota_{N,n}\colon [n] \to [N]$ and all $n \leq N$. Then p is L-Lipschitz in $(x,y) \mapsto W_1(\mathsf{S}_k x, \mathsf{S}_k y)$ for some $k \in \mathbb{N}$ and hence satisfies (6.21) with probability at least $1 - e^{-2\epsilon^2}$. If $\widehat{p}$ is any other such polynomial, then the generalization bound (6.22) holds.*

Proof. It was shown in [138, Sec. 5] that there exists $g_k \in \mathbb{R}[\mathbb{V}_k]$ satisfying $p(x) = \mathbb{E} g_k(\mathsf{S}_k x)$ for all x. Setting L to be the Lipschitz constant of g_k over Ω_k, the claim follows from Corollary 6.4.7. □

Note that we obtain sketching and generalization rates that does not directly depend on the degree of p or of $(\mathcal{A}_n)$. The following are a few examples of the polynomials to which Corollary 6.4.8 applies, which constitute "unnormalized" analogs of the polynomials unchanged by duplication considered in Corollary 6.3.10.

(Multisymmetric Functions) When $\mathbb{V}_n = \mathbb{R}^{d\times n}$, polynomials unchanged by zero-padding are precisely multisymmetric functions [213], consisting of polynomials of the form $p(x) = q(p_{\alpha_1}(x), \ldots, p_{\alpha_k}(x))$ where $q \in \mathbb{R}[y_1, \ldots, y_k]$ is a fixed polynomial and for $\alpha \in \mathbb{N}^d$ we have $p_\alpha(x) = \sum_{i=1}^n \prod_{j=1}^d x_{j,i}^{\alpha_j}$. When $d = 1$ and $\alpha \in \mathbb{N}$, the polynomials $p_\alpha(x) = \sum_{i=1}^n x_i^\alpha$ are known as power-sum polynomials, and polynomials unchanged by zero-padding are called symmetric functions. These are classic objects of study in combinatorics and representation theory, see [209, Chap. 7] and [151].

((Hyper)Graph Numbers) When $\mathbb{V}_n = \mathbb{R}^{n\times n}$, polynomials unchanged by zero-padding are precisely linear combinations of graph homomorphism numbers. Specifically, if $H \in \mathbb{N}^{d\times d}$ is a multigraph, we define the homomorphism number of H in $G \in \mathbb{V}_n$ by

$$\mathrm{hom}(H; G) = \sum_{f\colon [k]\to[n]} \prod_{i,j=1}^k G_{f(i),f(j)}^{H_{i,j}}.$$

When H, G are simple graphs, the value $\mathrm{hom}(H : G)$ is the number of graph homomorphisms from H to G. Similarly, when $\mathbb{V}_n = (\mathbb{R}^n)^{\otimes d}$ polynomials unchanged by zero-padding are linear combination of hypergraph homomorphism numbers, analogously defined.

(Graph Signals) When $\mathbb{V}_n = \mathbb{R}^{n\times n} \oplus \mathbb{R}^n$, polynomials unchanged by zero-padding include polynomial graph neural networks, which are compositions of polynomials of the form

$$(G, X) \mapsto \sigma\left(\sum_{\ell=0}^L G^\ell X \Theta_\ell\right), \quad \text{for polynomial } \sigma \text{ and } \Theta_0, \ldots, \Theta_L \in \mathbb{R}^{d\times d}.$$

More generally, in the language of [141, 142], Corollary 6.3.10 applies to any compatible (i.e., unchanged by zero-padding) neural network with polynomial activations.

It is challenging to directly verify continuity with respect to random binning for non-polynomial functions. We therefore give the following simple sufficient condition to verify this continuity, and obtain the same sketching and generalization bounds.

Proposition 6.4.9. *Let* $(\mathcal{A}_n)$ *be a* FinSet*-index of finite degree and let* $(f_n\colon \mathbb{R}^{\mathcal{A}_n} \to \mathbb{R})$ *be a sequence of functions satisfying the following conditions:*

(Unchanged by zero-padding) *We have* $f_N \circ \beta(\phi_{N,n}) = f_n$ *for any injection* $\phi_{N,n}\colon [n] \to [N]$*;*

(2-norm Lipschitz) *There is* $L > 0$ *such that* f_n *is* L*-Lipschitz in* $\|\cdot\|_2$ *for all* n*.*

Then $f\colon \bigsqcup_n \mathcal{B}_{\ell_1}^{(n)}(r) \to \mathbb{R}$ *defined by* $f(x) = f_n(x)$ *if* $x \in \mathcal{B}_{\ell_1}^{(n)}(r)$ *is continuous in* d_{samp}*. Moreover, the function* f *satisfies* (6.21) *with probability at least* $1-e^{-2\epsilon^2}$*, and satisfies* (6.22) *for any* $\widehat{f}$ *satisfying the present hypotheses or those of Corollary 6.4.7.*

Proof. Suppose $x = \lim_i x_i$ in d_{samp}. Then $\mathcal{S}_n(x) = \lim_i \mathcal{S}_n(x_i)$ weakly for each $n \in \mathbb{N}$ by Theorem 6.4.3. If the degree of $(\mathcal{A}_n)$ is D, we have

$$\mathbb{E} f(\mathcal{S}_n(x_i)) = \mathbb{E} f_{nD}(\mathcal{S}_n(x_i)) \xrightarrow{i\to\infty} \mathbb{E} f_{nD}(\mathcal{S}_n(x)) = \mathbb{E} f(\mathcal{S}_n(x)),$$

since f_{nD} is Lipschitz-continuous. Moreover, for any $y \in \mathcal{B}_{\ell_1}^{(m)}$ and any m, we have

$$|f(y) - \mathbb{E} f(\mathcal{S}_n(y))| = |f_m(y) - \mathbb{E} f_m(\mathcal{E}_n(y))| \leq L\mathbb{E}\|y - \mathcal{E}_n(y)\|_2 \leq \frac{Lr}{\sqrt{n}}, \quad (6.23)$$

where the first equality follows from Proposition 6.4.2 and the fact that f is unchanged by injections and the second from (6.20). Fixing arbitrary $\epsilon > 0$ and picking $n \geq (Lr/\epsilon)^2$, we conclude that

$$|f(x_i) - f(x_j)| \leq 2\epsilon + |\mathbb{E} f(\mathcal{S}_n(x_i)) - \mathbb{E} f(\mathcal{S}_n(x_j))| \xrightarrow{i,j\to\infty} 2\epsilon.$$

Since $\epsilon > 0$ was arbitrary, we conclude that $(f(x_i))$ is Cauchy and hence f proving the continuity of f. Finally, we have $|f(x) - f(\mathcal{S}_n(x))| \leq L\|x - \mathcal{S}_n(x)\|_2$ and the latter is $\frac{2}{n}$-subgaussian by the bounded-difference inequality, hence the last claim follows. ◻

We apply Proposition 6.4.9 to obtain sketching and generalization guarantees for some non-polynomial functions.

Corollary 6.4.10 (DeepSets)**.** *Let $\mathbb{V}_n = \mathbb{R}^{d\times n}$ and consider maps of the form $f(x_1,\ldots,x_n) = \sigma(\sum_i \rho(x_i))$ for $\rho\colon \mathbb{R}^d \to \mathbb{R}^\ell$ with $\rho(0) = 0$ and $\sigma\colon \mathbb{R}^\ell \to \mathbb{R}$. Then f is unchanged by zero-padding.*

1. *Suppose σ is L_σ-Lipschitz, and that ρ is $L_\rho\sqrt{r}$-Lipschitz with respect to the ℓ_2 norm on $\{x \in \mathbb{R}^d : \|x\|_1 \leq r\}$ for each r. Then f satisfies the hypotheses of Proposition 6.4.9 with $L = 2L_\rho L_\sigma r$.*

2. *The 1-norm $f(x) = \|x\|_1$ is of the above form for Lipschitz-continuous σ, ρ, but f is discontinuous in sampling metric.*

Condition 1 holds for example if ρ is differentiable with $\rho'(0) = 0$ and a $1/2$-Hölder continuous derivative, and condition 2 shows that such a stricter assumption on ρ is necessary.

Proof. The first claim follow from the observation that if $x, y \in \mathbb{V}_n$ then

$$|f(x)-f(y)| \leq L_\sigma \sum_i \|\rho(x_i)-\rho(y_i)\| \leq L_\sigma L_\rho \sum_i \|x_i-y_i\|_2\sqrt{\|x_i\|_1 + \|y_i\|_1} \leq 2L_\sigma L_\rho r\|x-y\|_2.$$

For the second claim, the choice $\sigma = \mathrm{id}$ and $\rho(x) = \|x\|_1$ yields $f(x) = \|x\|_1$. To see that it is discontinuous in sampling metric, consider the sequence $x^{(n)} = (\frac{1}{2n}e_1,\ldots,\frac{1}{2n}e_1,-\frac{1}{2n}e_1,\ldots,-\frac{1}{2n}e_1) \in \mathbb{V}_{2n}$, where both entries are repeated n times. Note that $x^{(n)} \to 0$ in sampling metric. Indeed, we have $\mathsf{S}_k x^{(n)} \overset{d}{=} \frac{1}{2n}\mathsf{S}_k(\mathbb{1}_n e_1^\top) - \frac{1}{2}\mathsf{S}_k'(\mathbb{1}_n e_1^\top)$ where $\mathsf{S}_k, \mathsf{S}_k'$ are independent. Since $\mathsf{S}_k\frac{1}{n}\mathbb{1}_n e_1^\top = (\frac{N_1}{n}e_1,\ldots,\frac{N_k}{n}e_1)$ where $(N_1,\ldots,N_k) \sim \mathrm{Multinom}(n,k,\mathbb{1}_k/k)$, we get $\mathbb{E}\|\mathsf{S}_k x^{(n)}\|_2^2 = \frac{1}{2}\sum_{i=1}^k \mathrm{Var}(N_i/n) = \frac{1-1/k}{2n}$, proving that $\mathsf{S}_k x^{(n)} \to 0$ weakly for each k. However, we have $f(x^{(n)}) \equiv 1$ for all n, proving that f is discontinuous in sampling metric. $\square$

When σ and ρ in Corollary 6.4.10 are neural networks, the resulting architecture is called DeepSets [233]. When $d = 1$ and $\Omega_n = \Delta^{n-1}$, different choices of σ and ρ yield various diversity indices used to quantify the concentration of a discrete probability distribution [235]. For example, setting $\rho(x) = x^q$ and $\sigma(t) = t^{\frac{1}{1-q}}$ for $q \in (0,1)$ gives Hill's effective number of species, which satisfies the hypotheses of Proposition 6.4.9 with $L = \frac{q}{1-q}$. We note however that some diversity indices are not continuous in sampling metric. This is the case for Shannon or Renyí entropies (given by $\sigma = \mathrm{id}$, $\rho(t) = t \log \frac{1}{t}$ for the former and $\sigma(t) = \frac{1}{1-q}\log(t)$, $\rho(t) = t^q$ for the latter) for example, both of which diverge along the sequence $x^{(n)} = \mathbb{1}_n/n$.

Corollary 6.4.11 (PointNet)**.** *Let $\mathbb{V}_n = \mathbb{R}^{d\times n}$ and consider maps of the form $f(x_1,\ldots,x_n) = \sigma(\sup_i \rho(x_i))$ for $\rho\colon \mathbb{R}^d \to \mathbb{R}^\ell_{\geq 0}$ with $\rho(0) = 0$ and $\sigma\colon \mathbb{R}^\ell \to \mathbb{R}$, where the supremum is taken coordinate-wise. When σ and ρ are L_σ and L_ρ-Lipschitz continuous in ℓ_2-norm f satisfies the hypotheses of Proposition 6.4.9 with $L = L_\rho L_\sigma$.*

Proof. It is clear that f is unchanged by zero-padding, and that $|f(x) - f(y)| \leq L_\sigma \sup_i \|\rho(x_i) - \rho(y_i)\| \leq L_\sigma L_\rho \|x - y\|_2$. □

Finally, we consider non-polynomial graph neural networks.

Corollary 6.4.12 (Graph Neural Networks)**.** *Let $\mathbb{V}_n = \mathbb{R}^{n\times n} \oplus \mathbb{R}^{n\times d}$ and consider a composition of maps of the form*

$$F(A, X) = \left(A, \sigma\left(\sum_{d=0}^{D} A^d X \Theta_d\right)\right), \quad \textit{where } \Theta_0, \ldots, \Theta_D \in \mathbb{R}^{d\times d},$$

and $\sigma\colon \mathbb{R}^d \to \mathbb{R}^d$ is L_σ-Lipschitz in ℓ_2 and is applied row-wise. Also define $P\colon \mathbb{V}_n \to \mathbb{R}^d$ by $P(A, X) = X^\top \mathbb{1}_n$ and consider the depth-ℓ graph neural network given by the composition

$$f = P \circ F^{\circ \ell}\colon \bigsqcup_n \mathcal{B}^{(n)}_{\ell_1}(r) \to \mathbb{R}^d,$$

where $\mathcal{B}^{(n)}_{\ell_1}(r) = \{(A, X) \in \mathbb{V}_n : \|A\|_1 + \|X\|_1 \leq r\}$. Then each coordinate function of f satisfies the hypotheses of Proposition 6.4.9.

The proof is an elementary but long computation, and we defer it to Section 6.4.

Missing Proofs from Section 6.4

We begin by proving Proposition 6.4.2, using the definition of a finite-degree FinSet-index.

Proof (Proposition 6.4.2). Fix nonzero $x \in \mathbb{R}^{\mathcal{A}_N}$. Since $(\mathcal{A}_n)$ has degree D, we can write $\mathcal{A}_n = \bigsqcup_{m=1}^{M} [n]^{d_m}/H_m$ for some subgroup $H_m \subseteq \mathsf{S}_{d_m}$ of permutations and with $D = \max_m d_m$. Sampling $\alpha_1, \ldots, \alpha_n$ from the distribution $|x|/\|x\|_1$ defines on $\mathcal{A}_N$, if $\alpha_i \in [n]^{d_i}/H_m$ let $(j_1, \ldots, j_{d_i}) \in [n]^{d_i}$ be a representative of the equivalence class of α_i. Randomly enumerate $\{t_1, \ldots, t_k\} = \bigcup_{i=1}^n \{j_1, \ldots, j_{d_i}\}$ and note that $k \le Dn$ and define the injection $\Phi_{N,nD} \colon [nD] \to [N]$ by sending $\Phi_{N,nD}(i) = t_i$ for $i \le k$ and setting $\Phi_{N,nD}|_{[nD]\setminus[k]}$ be a uniformly random injection into $[N] \setminus \{t_1, \ldots, t_k\}$. Finally, define $\mathscr{S}_n(x) \in \mathbb{R}^{\mathcal{A}_{nD}}$ by $\mathscr{S}_n(x)_J = \mathscr{E}_n(x)_{\Phi^{-1}_{N,nD}(J)}$ if there exists $I \in \mathcal{A}_N$ with $\Phi_{N,nD}(J) = I$ and $\mathscr{S}_n(x)_J = 0$ otherwise. With these definitions, we get the claimed equality. $\square$

We turn to proving Theorem 6.4.3. We shall need the following several lemmas to do so. The first lemma shows a compatibility between the empirical sampling $\mathscr{E}_n$ and the action of maps between finite sets.

Lemma 6.4.13. *Suppose $(\mathcal{A}_n)$ has degree D and $x \in \Delta^{\mathcal{A}_n - 1}$. Then for any map $f \colon [n] \to [m]$, we have $\mathscr{E}_k(\beta(f)x) \overset{d}{=} \beta(f)\mathscr{E}_k(x)$.*

Proof. Observe that the probability of sampling $\alpha \in \mathcal{A}_m$ from $\beta(f)x$ is $\sum_{\substack{\beta \in \mathcal{A}_n \\ \rho(f)(\beta) = \alpha}} x_\beta$. This is also the probability that if $\beta \in \mathcal{A}_n$ is drawn from x then $\rho(f)(\beta) = \alpha$. Thus,

$$\beta(f)\mathscr{E}_k(x) = \frac{1}{k}\sum_{i=1}^k e_{\rho(f)(\beta_i)} \overset{d}{=} \frac{1}{k}\sum_{i=1}^k e_{\alpha_i} = \mathscr{E}_k(\beta(f)x),$$

where $\beta_1, \ldots, \beta_k \in \mathcal{A}_n$ are sampled iid from x while $\alpha_1, \ldots, \alpha_k$ are sampled iid from $\beta(f)x$. $\square$

The second lemma bounds $d_{\mathrm{samp}}(x, \mathsf{S}_n x)$ for a finite-dimensional x.

Lemma 6.4.14. *Let $(\mathcal{A}_n)$ be any* FinSet*-index and let $(\mathbb{V}_n = \mathbb{R}^{\mathcal{A}_n})$ be the associated sequence of vector spaces. For any $x \in \mathbb{V}_m$ and $m \le n$, we have $d_{\mathrm{samp}}(x, \mathsf{S}_n x) = 0$ with probability at least $1 - \frac{m(m-1)}{2n}$, and hence $\mathbb{E} d_{\mathrm{samp}}(x, \mathsf{S}_n x) \le \frac{m(m-1)}{n}\|x\|_1$.*

Proof. Note that $\mathsf{S}_n x = \varsigma(F_{n,m})x$ and that $F_{n,m}$ is injective with probability at least $1 - \frac{m(m-1)}{2n}$, in which case $d_{\mathrm{samp}}(x, \mathsf{S}_n x) = 0$. The expectation bound follows since $W_1(x, \mathsf{S}_n x) \le \mathbb{E}\|x - \mathsf{S}_n x\|_1 \le 2\|x\|_1$. $\square$

The third lemma proves Theorem 6.4.3(2) for the special case of a finite-dimensional x.

Lemma 6.4.15. *Suppose $(\mathcal{A}_n)$ is a* FinSet*-index of finite degree. For any $x \in \mathbb{R}^{\mathcal{A}_N}$ and any $N \in \mathbb{N}$, we have*

$$\mathbb{E}_{\mathscr{S}_n(x)} W_1(\mathsf{S}_k x, \mathsf{S}_k \mathscr{S}_n(x)) \le \|x\|_1 \sqrt{\frac{2}{n}}.$$

Proof. By homogeneity of $\mathcal{S}_n$ and linearity of S_k, it suffices to prove the claim when $\|x\|_1 = 1$, which we assume for the remainder of the proof. Suppose first that $x \geq 0$, so that $x \in \Delta^{\mathcal{A}_N - 1}$. Then

$$\begin{aligned}\mathbb{E}W_1(\mathsf{S}_k x, \mathsf{S}_k \mathcal{S}_n(x)) &= \mathbb{E}W_1(\mathsf{S}_k x, \mathsf{S}_k \mathscr{E}_n(x)) \leq \mathbb{E}_{\mathscr{E}_n(x)} \mathbb{E}_{\mathsf{S}_k} \|\mathsf{S}_k x - \mathsf{S}_k \mathscr{E}_n(x)\|_2 \\ &= \mathbb{E}_{\mathsf{S}_k} \mathbb{E}_{\mathscr{E}_n(x)} \|\mathsf{S}_k x - \mathscr{E}_n(\mathsf{S}_k x)\|_2 \leq \frac{1}{\sqrt{n}},\end{aligned}$$

where the first equality follows from Proposition 6.4.2, the second equality follows from Lemma 6.4.13, and the last inequality follows from the fact that $\mathbb{E}\|p - \mathscr{E}_n(p)\|_2 \leq 1/\sqrt{n}$ for any discrete distribution p.

If $x \leq 0$ then the fact that $\mathscr{E}_n(-x) = -\mathscr{E}_n(x)$ and the linearity of S_k again yields the claimed bound. Therefore, suppose $\max\{0, x\} \neq 0$ and $\max\{0, -x\} \neq 0$ and write $x = \alpha x_+ - \beta x_-$ with $x_\pm \in \Delta^{\mathcal{A}_N - 1}$ and $\alpha, \beta \geq 0$ with $\alpha + \beta = \|x\|_1 = 1$ (explicitly, $\alpha x_+ = \max\{0, x\}$ and $\alpha = \|\max\{0, x\}\|_1$, and similarly for βx_-). By construction of $\mathscr{E}_n$, if $(N_+, N_-) \sim \text{Multinom}(2, n, (\alpha, \beta))$ then we have

$$\mathscr{E}_n(x) = \left[\frac{N_+}{n} \mathscr{E}_{N_+}(x_+) - \frac{N_-}{n} \mathscr{E}_{N_-}(x_-)\right].$$

Conditioning first on (N_+, N_-), we have

$$\mathbb{E}_{\mathscr{E}_n(x)} W_1(\mathsf{S}_k x, \mathsf{S}_k \mathscr{E}_n(x)) \leq \mathbb{E}_{(N_+, N_-)} \left[\frac{N_+}{n} \frac{1}{\sqrt{N_+}} + \frac{N_-}{n} \frac{1}{\sqrt{N_-}}\right] \leq \frac{1}{\sqrt{n}} \left(\sqrt{\alpha} + \sqrt{\beta}\right) \leq \sqrt{\frac{2}{n}},$$

as claimed. □

Finally, we need the following bound on the Lipschitz constant of the sampling map $\mathscr{E}_n$.

Lemma 6.4.16. *Let $\mathcal{A}$ be any finite set and suppose $x, y \in \mathbb{R}^{\mathcal{A}}$, and let W_1 be the Wasserstein-1 metric with respect to the ℓ_2 norm on $\mathbb{R}^{\mathcal{A}}$. Then*

$$W_1(\mathscr{E}_n(x), \mathscr{E}_n(y)) \leq 4\|x - y\|_1.$$

Proof. Suppose first that $\|x\|_1 = \|y\|_1 = 1$, so $|x|, |y| \in \Delta^{\mathcal{A}-1}$. Sampling (α_i, β_i) iid from an optimal coupling of $|x|$ and $|y|$ with respect to the discrete metric on $\mathcal{A}$, so that $\mathbb{P}[\alpha_i = \beta_i] = \frac{1}{2}\||x| - |y|\|_1$. For each $i \in [n]$ we have

$$\|\text{sign}(x_{\alpha_i}) e_{\alpha_i} - \text{sign}(y_{\beta_i}) e_{\beta_i}\|_2 = \begin{cases} 0 & \text{if } \alpha_i = \beta_i \text{ and } \text{sign}(x_{\alpha_i}) = \text{sign}(y_{\beta_i}), \\ 2 & \text{if } \alpha_i = \beta_i \text{ and } \text{sign}(x_{\alpha_i}) \neq \text{sign}(y_{\beta_i}), \\ \sqrt{2} & \text{if } \alpha_i \neq \beta_i. \end{cases}$$

Note also that if $\pi \in \Delta^{\mathcal{A}\times\mathcal{A}}$ is the optimal coupling of $|x|$ and $|y|$, then since $\pi\mathbb{1} = |x|$ and $\mathbb{1}^\top\pi = |y|^\top$ we have $\pi_{\alpha,\alpha} = \mathbb{P}[\alpha_i = \beta_i = \alpha] \le \min\{|x_\alpha|, |y_\alpha|\}$ for each $\alpha \in \mathcal{A}$. Therefore,

$$\begin{aligned} W_1(\mathscr{E}_n(x), \mathscr{E}_n(y)) &\le \frac{1}{n}\sum_{i=1}^n \mathbb{E}\|\mathrm{sign}(x_{\alpha_i})e_{\alpha_i} - \mathrm{sign}(y_{\beta_i})e_{\beta_i}\|_2 \\ &= \sqrt{2}\mathbb{P}[\alpha_1 \ne \beta_1] + 2\mathbb{P}[\alpha_1 = \beta_1, \mathrm{sign}(x_{\alpha_1}) \ne \mathrm{sign}(y_{\beta_1})] \\ &\le \frac{1}{\sqrt{2}}\||x| - |y|\|_1 + \sum_{\substack{\alpha\in\mathcal{A} \\ \mathrm{sign}(x_\alpha)\ne\mathrm{sign}(y_\alpha)}} 2\min\{|x_\alpha|, |y_\alpha|\}. \end{aligned}$$

Since $a + b = |a - b| + 2\min\{a, b\}$ for any $a, b \ge 0$, we have

$$\begin{aligned} \|x - y\|_1 &= \sum_{\substack{\alpha\in\mathcal{A} \\ \mathrm{sign}(x_\alpha)=\mathrm{sign}(y_\alpha)}} \Big||x_\alpha| - |y_\alpha|\Big| + \sum_{\substack{\alpha\in\mathcal{A} \\ \mathrm{sign}(x_\alpha)\ne\mathrm{sign}(y_\alpha)}} (|x_\alpha| + |y_\alpha|) \\ &= \||x| - |y|\|_1 + \sum_{\substack{\alpha\in\mathcal{A} \\ \mathrm{sign}(x_\alpha)\ne\mathrm{sign}(y_\alpha)}} 2\min\{|x_\alpha|, |y_\alpha|\}. \end{aligned}$$

Thus, we have

$$W_1(\mathscr{E}_n(x), \mathscr{E}_n(y)) \le \frac{1}{\sqrt{2}}\||x| - |y|\|_1 + \|x - y\|_1 - \||x| - |y|\|_1 \le \|x - y\|_1,$$

when $\|x\|_1 = \|y\|_1 = 1$.

For general nonzero $x, y \in \mathbb{R}^{\mathcal{A}}$, we have $\mathscr{E}_n(x) = \|x\|_1\mathscr{E}_n(x/\|x\|_1)$ and similarly for $\mathscr{E}_n(y)$. Therefore,

$$\begin{aligned} W_1(\mathscr{E}_n(x), \mathscr{E}_n(y)) &\le \|x\|_1 W_1(\mathscr{E}_n(x/\|x\|_1), \mathscr{E}_n(y/\|y\|_1)) + W_1(\|x\|_1\mathscr{E}_n(y/\|y\|_1), \|y\|_1\mathscr{E}_n(y/\|y\|_1)) \\ &\le \|x\|_1 \left\|\frac{x}{\|x\|_1} - \frac{y}{\|y\|_1}\right\|_1 + 2\Big|\|x\|_1 - \|y\|_2\Big| \\ &\le \|x - y\|_1 + \left\|y - \frac{\|x\|_1}{\|y\|_1}y\right\|_1 + 2\|x - y\|_1 \\ &\le 4\|x - y\|_1, \end{aligned}$$

as claimed. □

We are now ready to prove Theorem 6.4.3.

Proof (Theorem 6.4.3). We begin by proving statement 1. Suppose $x \in \overline{\Omega}_\infty \setminus \{0\}$ and write $x = \lim_i x_i$ for $x_i \in \Omega_\infty$. First $\mathcal{S}_n(x)$ is well-defined for any n and equals the weak limit of $(\mathcal{S}_n(x_i))_i$. Indeed, the latter is Cauchy because

$$W_1(\mathcal{S}_n(x_i), \mathcal{S}_n(x_j)) \le W_1(\mathcal{S}_n(x_i), \mathsf{S}_k\mathcal{S}_n(x_i)) + W_1(\mathsf{S}_k\mathcal{S}_n(x_i), \mathsf{S}_k\mathcal{S}_n(x_j)) + W_1(\mathsf{S}_k\mathcal{S}_n(x_j), \mathcal{S}_n(x_j)),$$

where we set S_k to be independent of $\mathcal{S}_n(x_i)$ and $\mathcal{S}_n(x_j)$. The first and third terms above are bounded by $\frac{(nD)^2}{k}$ by Lemma 6.4.14. For the second term, we have

$$\begin{aligned} W_1(\mathsf{S}_k\mathcal{S}_n(x_i), \mathsf{S}_k\mathcal{S}_n(x_j)) = W_1(\mathsf{S}_k\mathcal{E}_n(x_i), \mathsf{S}_k\mathcal{E}_n(x_i)) &= W_1(\mathcal{E}_n(\mathsf{S}_k x_i), \mathcal{E}_n(\mathsf{S}_k x_j)) \\ &\le \sqrt{|\mathcal{A}_k|}W_1(\mathsf{S}_k x_i, \mathsf{S}_k x_j), \end{aligned}$$

where the first equality follows by Proposition 6.4.2, and the second by Lemma 6.4.13 and the fact that $\|y\|_1 \le \sqrt{|\mathcal{A}_k|}\|y\|_2$ for $y \in \mathbb{R}^{\mathcal{A}_k}$. Now fix $\epsilon > 0$ and $k \ge (nD)^2/\epsilon$, and let $N(\epsilon)$ satisfy $W_1(\mathsf{S}_k x_i, \mathsf{S}_k x_j) \le \epsilon/\sqrt{|\mathcal{A}_k|}$ for all $i, j \ge N(\epsilon)$, which exists as $(\mathsf{S}_k x_i)$ converge weakly to $\mathsf{S}_k x$. We conclude that $W_1(\mathcal{S}_n(x_i), \mathcal{S}_n(x_j)) \le 3\epsilon$ for all $i, j \ge N(\epsilon)$, proving the first statement.

The second statement was proved for $x \in \Omega_\infty$ in Lemma 6.4.15. For $x \in \overline{\Omega}_\infty$, write $x = \lim_i x_i$ and note that

$$\mathbb{E}_{\mathcal{S}_n} W_1(\mathsf{S}_k x, \mathsf{S}_k\mathcal{S}_n(x)) \le W_1(\mathsf{S}_k x, \mathsf{S}_k x_i) + \mathbb{E}_{\mathcal{S}_n} W_1(\mathsf{S}_k x_i, \mathsf{S}_k\mathcal{S}_n(x_i)) + \mathbb{E}_{\mathcal{S}_n} W_1(\mathsf{S}_k\mathcal{S}_n(x_i), \mathsf{S}_k\mathcal{S}_n(x)).$$

Taking $i \to \infty$ and observing that the first and last term converge to zero as argued above, while the middle is bounded by $\sqrt{2/n}$ independently of i, we obtain the claimed bounds for $x \in \overline{\Omega}_\infty$.

Finally, when $x \in \Omega_\infty$ note that $\mathcal{S}_n(x)$ is a function of n iid indices sampled from x, and changing any one of them affects two entries by $1/n$. Thus, the claimed subgaussianity follows from the bounded-difference inequality. When $x \in \overline{\Omega}_\infty$, we have $\mathcal{S}_n(x) = \lim_i \mathcal{S}_n(x_i)$ for some $x_i \in \Omega_\infty$ by the first part and hence its claimed subgaussianity likewise follows.

For the third statement, we may assume $\|x\|_1 = 1$ by homogeneity. For any $m \le n$ we have

$$\begin{aligned} \mathbb{E} d_{\mathrm{samp}}(x, \mathsf{S}_n x) &\le \mathbb{E} d_{\mathrm{samp}}(x, \mathcal{S}_m(x)) + \mathbb{E} d_{\mathrm{samp}}(\mathcal{S}_m(x), \mathsf{S}_n\mathcal{S}_m(x)) + \mathbb{E}\underbrace{d_{\mathrm{samp}}(\mathsf{S}_n\mathcal{S}_m(x), \mathsf{S}_n x)}_{=d_{\mathrm{samp}}(\mathcal{S}_m(\mathsf{S}_n x), \mathsf{S}_n x)} \\ &\le \frac{2}{\sqrt{m}} + \frac{D^2 m^2}{2n}, \end{aligned}$$

where S_n is independent of $\mathcal{S}_m$ and we used the fact that $d_{\mathrm{samp}}(\mathsf{S}_n\mathcal{S}_m(x), \mathcal{S}_m(\mathsf{S}_n x)) = 0$ conditioned on S_n. Choosing $m = \lceil n^{2/5} \rceil \le n$, we get

$$\mathbb{E} d_{\mathrm{samp}}(x, \mathsf{S}_n x) \le \frac{2}{n^{1/5}} + \frac{D^2(n^{2/5}+1)^2}{2n} \le \frac{2(1+D^2)}{n^{1/5}},$$

as claimed. Finally, with probability $1 - 2e^{-2\epsilon^2} - \frac{D^2}{2n^{1/5}}$ and m chosen as above we have

$$d_{\mathrm{samp}}(x, \mathsf{S}_n x) \le d_{\mathrm{samp}}(x, \mathcal{S}_m(x)) + d_{\mathrm{samp}}(\mathcal{S}_m(\mathsf{S}_n x), \mathsf{S}_n x) \le \frac{2(1+2\epsilon)}{n^{1/5}},$$

by the $\frac{4}{n}$-subgaussianity from Theorem 6.4.3. □

Finally, we prove Corollary 6.4.12.

Proof (Corollary 6.4.12). It is straightforward to verify that f is unchanged by zero-padding. If $\|A\|_1 + \|X\|_1 \leq r$, we have

$$\|F(A, X) - F(B, Y)\|_2 \leq \|A - B\|_2 + L_\sigma \left\| \sum_{d=0}^{D} A^d X \Theta_d - \sum_{d=0}^{D} B^d Y \Theta_d \right\|_2$$

$$\leq \|A - B\|_2 + L_\sigma \sum_{d=0}^{D} \left(\|A^d X \Theta_d - A^d Y \Theta_d\|_2 + \|A^d Y \Theta_d - B^d Y \Theta_d\|_2 \right)$$

$$\leq \|A - B\|_2 + L_\sigma \sum_{d=0}^{D} \|A\|_2^d \|\Theta_d\|_{\mathrm{op}} \|X - Y\|_2 + \sum_{d=0}^{D} \|A^d - B^d\|_{\mathrm{op}} \|Y\|_2 \|\Theta_d\|_{\mathrm{op}}$$

$$\leq \|A - B\|_2 + L_\sigma \max_{d=0,\ldots,D} \{(r^d + r)\|\Theta_d\|_{\mathrm{op}}\} \left[(D+1)\|X - Y\|_2 + \frac{D(D+1)}{2} \|A - B\|_2 \right],$$

$$\leq \underbrace{\left(1 + L_\sigma \frac{D(D+1)}{\sqrt{2}} \max_{d=0,\ldots,D} \{(r^d + r)\|\Theta_d\|_{\mathrm{op}}\} \right)}_{=L(r)} \|(A, X) - (B, Y)\|_2.$$

Finally, note that if $\|A\|_1 + \|X\|_1 \leq r$, then

$$\|F(A, X)\|_1 \leq \|A\|_1 + \left\| \sum_{d=0}^{D} A^d X \Theta_d \right\|_1 \leq \|A\|_1 + \sum_{d=0}^{D} \|\Theta_d\|_1 \|A^d\|_1 \|X\|_1 \leq r + \sum_{d=0}^{D} r^{d+1} \|\Theta_d\|_1 = R(r).$$

Therefore, after j applications of $F(A, X)$ we have $\|F^{\circ j}(A, X)\|_1 \leq R^{\circ j}(r)$, and hence is Lipschitz with constant $\prod_{j=1}^{\ell} L(R^{\circ(j-1)}(r))$. Since $f = P \circ F^{\circ \ell}$ and P is 1-Lipschitz, we obtain the claim. □

6.5 Discussion

We have considered the closely-related problems of any-dimensional generalization and sketching of any-dimensional functions. We have seen that both problems are intimately tied to the problem of approximating an arbitrarily-large input to these functions by a fixed-dimensional input, which can often be achieved by randomly sketching it. To tackle these problems, we compare objects of different sizes by comparing distributions of random samples of them of the same size. By using the right sampling maps, depending on the application domain and the relations between inputs of different sizes there, we obtain rich families of compact sets on which we get uniform rates for generalization and sketching, and a correspondence between limit objects and data distributions. Focusing on specific generalizations of sampling with replacement, random binning, and species sampling, we then obtain precise quantitative rates for approximation, sketching, and generalization. Some of the function classes for which our framework applies include polynomials, permutation-invariant transformers, and several architectures defined for sets, point clouds, and graphs of all sizes. We end with a few directions for future work.

(Rate for Binning) Is the rate $n^{-1/5}$ in (6.19) optimal for binning? Can it be improved to $n^{-1/2}$?

(Distributions of species samples) Can we characterize the collection of sequences $(\mathrm{Law}(\mathsf{E}_k(X)))_k$ for random X in terms of the relations of these distributions across dimensions, analogously to Corollaries 6.3.6 and 6.4.4? These sequences of distributions include random partitions of integers [122, 123, 176, 190] and edge-exchangeable random graph models [110, 140].

(Other Sampling Maps) While we focused on sampling with replacement, random binning, and species sampling in this chapter, there are other notions of sampling that might be appropriate for different applications. For example, in the context of natural language sentences, which are not permutation-invariant, is there a different notion of sampling that can summarize long sentences by short ones?

(Set-based Summaries) In this chapter, we consider summarizing objects using random sampling, and comparing these random summaries in Wasserstein distance. Another type of summary studied in the context of certain graph limits involves summarizing an object by forming sets consisting of all possible projections of it, suitably defined, and comparing these sets in Hausdorff distance, see [146, Chap. 12] and [135] for example. Can we develop a general framework and rates for such summaries?

Chapter 7

CONCLUSIONS AND FUTURE DIRECTIONS

We briefly summarize the contributions of this thesis. This thesis has developed a framework for working with any-dimensional data. We began by systematically deriving finitely-parametrized families of convex sets, functions, and neural networks that accept inputs in any dimension. We have also shown that broad families of sequences of symmetric convex problems of growing size can be reduced to constant-sized programs. Our finitely-parametrized descriptions are obtained by imposing symmetries and relations between functions and sets in different dimensions, allowing us to exploit the phenomenon of representation stability. We then turn to computationally proving inequalities between any-dimensional polynomial functions. We do so by constructing new sequences of polynomial optimization problems giving increasingly-better such inequalities. Our construction again uses representation stability, as well as new generalizations of de Finetti's theorem. One of our generalizations involves the analysis of a new extension of random binning, and we use this extension to study random quotients of graphs and its associated notion of limit of growing-sized graphs. Our new limits of graphs, which we call grapheurs, allow us to test properties of large graphs pertaining to their hub structure. Finally, we systematically study the generalization and sketching of any-dimensional functions, both of which involve comparing the inputs to these functions across dimensions. We compare objects of different sizes by comparing the distribution of their random samples, where the choice of sampling maps depends on the application domain. Focusing on generalizations of sampling with replacement and random binning, we prove new and improved quantitative rates for approximation of arbitrarily-large objects by fixed-dimensional ones, and associated rates for generalization and sketching of any-dimensional functions.

We end by listing a few higher-level directions for future work.

Statistical Inference over Spaces of Algorithms: My work on any-dimensional learning can be viewed as an approach to learn an *algorithm*, i.e., a map accepting and returning objects of any size, by fitting maps from different parametric families [137, 139, 141, 142]. Other approaches in the literature may also be viewed similarly, with the underlying families described in terms of different domain-specific primitives.

Adopting an inferential perspective, the algorithms learned from data using existing paradigms come with limited generalization guarantees to inputs of arbitrary size. The fundamental challenge in obtaining such guarantees is that of precisely specifying how input problem instances of different sizes are related. My previous work addresses this question for problem instances related by their symmetries. However, such structure it not present or explicitly evident in many other application domains. I aim to identify and exploit the mathematical structure underlying these broader types of relations, and ultimately developing generalization guarantees across problem sizes for algorithms learned from data.

New Information Inequalities: Proving inequalities satisfied by entropies of random variables and their quantum analogues is a fundamental problem in information theory, with new inequalities leading to the design of better communication networks and error-correcting codes. Proving entropy inequalities is an instance of an any-dimensional optimization problem. Indeed, entropies of random variables taking values in a finite alphabet are invariant under relabelling of the alphabet, and are defined for all alphabet sizes. Similarly, quantum entropies are invariant to unitary transformations on the Hilbert space describing a system of qudits, and are defined for systems described by a space of any finite dimension. Thus, I expect that extensions of our framework from [138] will enable us to computationally search over and prove such entropy inequalities.

BIBLIOGRAPHY

[1] Romain Abraham, Jean-François Delmas, and Julien Weibel. "Probability-graphons: Limits of large dense weighted graphs". In: *Innovations in Graph Theory* 2 (2025), pp. 25–117. DOI: 10.5802/igt.7
URL: https://hal.science/hal-04361443.

[2] Jose Acevedo and Grigoriy Blekherman. "Power mean inequalities and sums of squares". In: *Discrete & Computational Geometry* (2024), pp. 1–47.

[3] Jose Acevedo et al. "Symmetric nonnegative functions, the tropical Vandermonde cell and superdominance of power sums". In: *arXiv preprint arXiv:2408.04616* (2024).

[4] Jose Acevedo et al. "The Wonderful Geometry of the Vandermonde map". In: *arXiv preprint arXiv:2303.09512* (2023).

[5] José Acevedo and Mauricio Velasco. "Test sets for nonnegativity of polynomials invariant under a finite reflection group". In: *Journal of Pure and Applied Algebra* 220.8 (2016), pp. 2936–2947. ISSN: 0022-4049. DOI: https://doi.org/10.1016/j.jpaa.2016.01.010
URL: https://www.sciencedirect.com/science/article/pii/S0022404916000190.

[6] David Aldous and J Michael Steele. "The objective method: probabilistic combinatorial optimization and local weak convergence". In: *Probability on discrete structures*. Springer, 2004, pp. 1–72.

[7] David J Aldous. "Representations for partially exchangeable arrays of random variables". In: *Journal of Multivariate Analysis* 11.4 (1981), pp. 581–598.

[8] Yulia Alexandr, Joe Kileel, and Bernd Sturmfels. "Moment Varieties for Mixtures of Products". In: *arXiv preprint arXiv:2301.09068* (2023).

[9] Colin R. Blyth and. "Expected Absolute Error of the Usual Estimator of the Binomial Parameter". In: *The American Statistician* 34.3 (1980), pp. 155–157. DOI: 10.1080/00031305.1980.10483022. eprint: https://www.tandfonline.com/doi/pdf/10.1080/00031305.1980.10483022
URL: https://www.tandfonline.com/doi/abs/10.1080/00031305.1980.10483022.

[10] Tim Austin. "On exchangeable random variables and the statistics of large graphs and hypergraphs". In: *Probability Surveys* 5.none (2008), pp. 80–145. DOI: 10.1214/08-PS124
URL: https://doi.org/10.1214/08-PS124.

[11] Maria-Florina Balcan et al. "Learning to Branch". In: *Proceedings of the 35th International Conference on Machine Learning*. Ed. by Jennifer Dy and Andreas Krause. Vol. 80. Proceedings of Machine Learning Research. PMLR, 2018, pp. 344–353
URL: https://proceedings.mlr.press/v80/balcan18a.html.

[12] Peter Bartlett. *Theoretical Statistics, Lecture 14*. https://www.stat.berkeley.edu/~bartlett/courses/2013spring-stat210b/notes/14notes.pdf. 2013.

[13] Danielle S Bassett and Olaf Sporns. "Network neuroscience". In: *Nature neuroscience* 20.3 (2017), pp. 353–364.

[14] Aharon Ben-Tal and Arkadi Nemirovski. *Lectures on Modern Convex Optimization*. Society for Industrial and Applied Mathematics, 2001. DOI: 10.1137/1.9780898718829. eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9780898718829
URL: https://epubs.siam.org/doi/abs/10.1137/1.9780898718829.

[15] Itai Benjamini and Oded Schramm. "Recurrence of Distributional Limits of Finite Planar Graphs". In: *Electronic Journal of Probability* 6.none (2001), pp. 1–13. DOI: 10.1214/EJP.v6-96
URL: https://doi.org/10.1214/EJP.v6-96.

[16] Grigoriy Blekherman, Pablo A Parrilo, and Rekha R Thomas. *Semidefinite optimization and convex algebraic geometry*. SIAM, 2012.

[17] Andreas Bluhm, Anna Jenčová, and Ion Nechita. "Incompatibility in general probabilistic theories, generalized spectrahedra, and tensor norms". In: *Communications in Mathematical Physics* 393.3 (2022), pp. 1125–1198.

[18] Andreas Bluhm and Ion Nechita. "Joint measurability of quantum effects and the matrix diamond". In: *Journal of Mathematical Physics* 59.11 (2018), p. 112202. DOI: 10.1063/1.5049125. eprint: https://doi.org/10.1063/1.5049125
URL: https://doi.org/10.1063/1.5049125.

[19] Alexander Bogatskiy et al. "Lorentz Group Equivariant Neural Network for Particle Physics". In: *Proceedings of the 37th International Conference on Machine Learning*. Ed. by Hal Daumé III and Aarti Singh. Vol. 119. Proceedings of Machine Learning Research. PMLR, 2020, pp. 992–1002
URL: https://proceedings.mlr.press/v119/bogatskiy20a.html.

[20] C. Borgs et al. "Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing". In: *Advances in Mathematics* 219.6 (2008), pp. 1801–1851. ISSN: 0001-8708. DOI: https://doi.org/10.1016/j.aim.2008.07.008
URL: https://www.sciencedirect.com/science/article/pii/S0001870808002053.

[21] Christian Borgs et al. “An L^p theory of sparse graph convergence I: Limits, sparse random graph models, and power law distributions”. In: *Transactions of the American Mathematical Society* 372.5 (2019), pp. 3019–3062.

[22] Christian Borgs et al. “Identifiability for graphexes and the weak kernel metric”. In: *Building Bridges II: Mathematics of László Lovász*. Springer, 2020, pp. 29–157.

[23] Christian Borgs et al. “Limits of sparse configuration models and beyond: Graphexes and MultiGraphexes”. In: *The Annals of Probability* 49.6 (2021), pp. 2830–2873. DOI: 10.1214/21-AOP1508
URL: https://doi.org/10.1214/21-AOP1508.

[24] Christian Borgs et al. “Sampling perspectives on sparse exchangeable graphs”. In: *The Annals of Probability* 47.5 (2019), pp. 2754–2800. DOI: 10.1214/18-AOP1320
URL: https://doi.org/10.1214/18-AOP1320.

[25] Christian Borgs et al. “Sparse Exchangeable Graphs and Their Limits via Graphon Processes”. In: *Journal of Machine Learning Research* 18.210 (2018), pp. 1–71
URL: http://jmlr.org/papers/v18/16-421.html.

[26] Marcin Bownik and John Jasper. “The Schur-Horn theorem for operators with finite spectrum”. In: *Transactions of the American Mathematical Society* 367.7 (2015), pp. 5099–5140.

[27] Michael M Bronstein et al. “Geometric deep learning: Grids, groups, graphs, geodesics, and gauges”. In: *arXiv preprint arXiv:2104.13478* (2021).

[28] Daniel Brosch. “Symmetry reduction in convex optimization with applications in combinatorics”. English. PhD thesis. Tilburg University, 2022. DOI: 10.26116/12v4-k024.

[29] Deborah L. Bryan and Morton E. O’Kelly. “Hub-and-Spoke Networks in Air Transportation: An Analytical Review”. In: *Journal of Regional Science* 39.2 (1999), pp. 275–295. DOI: https://doi.org/10.1111/1467-9787.00134. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-9787.00134
URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9787.00134.

[30] Christian Bueno and Alan Hylton. “On the representation power of set pooling networks”. In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 17170–17182.

[31] Sabine Burgdorf, Igor Klep, Janez Povh, et al. *Optimization of polynomials in non-commuting variables*. Vol. 2. Springer, 2016.

[32] Stefan A. Burr and Vera Rosta. "On the Ramsey multiplicities of graphs—problems and recent results". In: *Journal of Graph Theory* 4.4 (1980), pp. 347–361. DOI: https://doi.org/10.1002/jgt.3190040403. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jgt.3190040403 URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/jgt.3190040403.

[33] Diana Cai, Trevor Campbell, and Tamara Broderick. "Edge-exchangeable graphs and sparsity". In: *Advances in Neural Information Processing Systems*. Ed. by D. Lee et al. Vol. 29. Curran Associates, Inc., 2016 URL: https://proceedings.neurips.cc/paper_files/paper/2016/file/1a0a283bfe7c549dee6c638a05200e32-Paper.pdf.

[34] Trevor Campbell, Diana Cai, and Tamara Broderick. "Exchangeable trait allocations". In: *Electronic Journal of Statistics* 12.2 (2018), pp. 2290–2322. DOI: 10.1214/18-EJS1455 URL: https://doi.org/10.1214/18-EJS1455.

[35] Utkan Onur Candogan and Venkat Chandrasekaran. "Finding Planted Subgraphs with Few Eigenvalues using the Schur–Horn Relaxation". In: *SIAM Journal on Optimization* 28.1 (2018), pp. 735–759. DOI: 10.1137/16M1075144. eprint: https://doi.org/10.1137/16M1075144 URL: https://doi.org/10.1137/16M1075144.

[36] Pierre Cardaliaguet. *Notes on mean field games*. Tech. rep. Technical report, 2010.

[37] François Caron and Emily B. Fox. "Sparse Graphs Using Exchangeable Random Measures". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 79.5 (2017), pp. 1295–1366. ISSN: 1369-7412. DOI: 10.1111/rssb.12233. eprint: https://academic.oup.com/jrsssb/article-pdf/79/5/1295/49269335/jrsssb_79_5_1295.pdf URL: https://doi.org/10.1111/rssb.12233.

[38] Rui Castro et al. "Network Tomography: Recent Developments". In: *Statistical Science* 19.3 (2004), pp. 499–517. DOI: 10.1214/088342304000000422 URL: https://doi.org/10.1214/088342304000000422.

[39] Venkat Chandrasekaran, Pablo A. Parrilo, and Alan S. Willsky. "Convex Graph Invariants". In: *SIAM Review* 54.3 (2012), pp. 513–541. DOI: 10.1137/100816900. eprint: https://doi.org/10.1137/100816900 URL: https://doi.org/10.1137/100816900.

[40] Venkat Chandrasekaran and Parikshit Shah. "Relative Entropy Relaxations for Signomial Optimization". In: *SIAM Journal on Optimization* 26.2 (2016), pp. 1147–1173. DOI: 10.1137/140988978. eprint: https://doi.org/10.1137/140988978 URL: https://doi.org/10.1137/140988978.

[41] Venkat Chandrasekaran et al. "The convex geometry of linear inverse problems". In: *Foundations of Computational mathematics* 12.6 (2012), pp. 805–849.

[42] Ángel Chávez, Stephan Ramon Garcia, and Jackson Hurley. "Norms on complex matrices induced by random vectors". In: *Canadian Mathematical Bulletin* 66.3 (2023), pp. 808–826. DOI: 10.4153/S0008439522000741.

[43] Zhengdao Chen et al. "Can Graph Neural Networks Count Substructures?" In: *Advances in Neural Information Processing Systems*. Ed. by H. Larochelle et al. Vol. 33. Curran Associates, Inc., 2020, pp. 10383–10395 URL: https://proceedings.neurips.cc/paper/2020/file/75877cb75154206c4e65e76b88a12712-Paper.pdf.

[44] Christopher H Chiu et al. "Sym-Noetherianity for powers of GL-varieties". In: *arXiv preprint arXiv:2212.05790* (2022).

[45] Thomas Church and Jordan Ellenberg. "Homology of FI-modules". In: *Geometry & Topology* 21.4 (2017), pp. 2373–2418. DOI: 10.2140/gt.2017.21.2373 URL: https://doi.org/10.2140/gt.2017.21.2373.

[46] Thomas Church, Jordan S. Ellenberg, and Benson Farb. "FI-modules and stability for representations of symmetric groups". In: *Duke Mathematical Journal* 164.9 (2015), pp. 1833–1910. DOI: 10.1215/00127094-3120274 URL: https://doi.org/10.1215/00127094-3120274.

[47] Thomas Church and Benson Farb. "Representation theory and homological stability". In: *Advances in Mathematics* 245 (2013), pp. 250–314. ISSN: 0001-8708. DOI: https://doi.org/10.1016/j.aim.2013.06.016 URL: https://www.sciencedirect.com/science/article/pii/S0001870813002259.

[48] Taco Cohen and Max Welling. "Group Equivariant Convolutional Networks". In: *Proceedings of The 33rd International Conference on Machine Learning*. Ed. by Maria Florina Balcan and Kilian Q. Weinberger. Vol. 48. Proceedings of Machine Learning Research. New York, New York, USA: PMLR, 2016, pp. 2990–2999 URL: https://proceedings.mlr.press/v48/cohenc16.html.

[49] Taco S. Cohen and Max Welling. "Steerable CNNs". In: *International Conference on Learning Representations*. 2017 URL: https://openreview.net/forum?id=rJQKYt5ll.

[50] Steven J Cook et al. "Whole-animal connectomes of both Caenorhabditis elegans sexes". In: *Nature* 571.7763 (2019), pp. 63–71.

[51] Graham Cormode and S. Muthukrishnan. "An improved data stream summary: the count-min sketch and its applications". In: *Journal of Algorithms* 55.1 (2005), pp. 58–75. ISSN: 0196-6774. DOI: https://doi.org/10.

1016/j.jalgor.2003.12.001
URL: https://www.sciencedirect.com/science/article/pii/S0196677403001913.

[52] Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. John Wiley & Sons, 1999.

[53] Harry Crane and Walter Dempsey. "Edge exchangeable models for network data". In: *arXiv preprint arXiv:1603.04571* (2016).

[54] Harry Crane and Walter Dempsey. "Relational exchangeability". In: *Journal of Applied Probability* 56.1 (2019), pp. 192–208. DOI: 10.1017/jpr.2019.13.

[55] Kenneth R Davidson. *C^*-algebras by example*. Vol. 6. American Mathematical Society, 1996.

[56] Bruno De Finetti. "Funzione caratteristica di un fenomeno aleatorio". In: *Atti del Congresso Internazionale dei Matematici: Bologna del 3 al 10 de settembre di 1928*. 1929, pp. 179–190.

[57] Bruno De Finetti. "La prévision: ses lois logiques, ses sources subjectives". In: *Annales de l'institut Henri Poincaré*. Vol. 7. 1937, pp. 1–68.

[58] Sebastian Debus and Cordian Riener. "Reflection groups and cones of sums of squares". In: *arXiv preprint arXiv:2011.09997* (2020).

[59] P. Diaconis and D. Freedman. "Finite Exchangeable Sequences". In: *The Annals of Probability* 8.4 (1980), pp. 745–764. DOI: 10.1214/aop/1176994663
URL: https://doi.org/10.1214/aop/1176994663.

[60] Persi Diaconis and Svante Janson. "Graph limits and exchangeable random graphs". In: *arXiv preprint arXiv:0712.2749* (2007).

[61] Mateo Díaz et al. "Invariant kernels: Rank stabilization and generalization across dimensions". In: *arXiv preprint arXiv:2502.01886* (2025).

[62] Jean Dieudonné. *Foundations of Modern Analysis*. New York: Academic Press, 1969.

[63] Martin Doležal et al. "Cut distance identifying graphon parameters over weak* limits". In: *Journal of Combinatorial Theory, Series A* 189 (2022), p. 105615. ISSN: 0097-3165. DOI: https://doi.org/10.1016/j.jcta.2022.105615
URL: https://www.sciencedirect.com/science/article/pii/S0097316522000231.

[64] Jan Draisma. "Noetherianity up to Symmetry". In: *Combinatorial Algebraic Geometry: Levico Terme, Italy 2013, Editors: Sandra Di Rocco, Bernd Sturmfels*. Cham: Springer International Publishing, 2014, pp. 33–61. ISBN: 978-3-319-04870-3. DOI: 10.1007/978-3-319-04870-3_2
URL: https://doi.org/10.1007/978-3-319-04870-3_2.

[65] Evgenii Borisovich Dynkin. "Classes of equivalent random quantities". In: *Uspekhi Matematicheskikh Nauk* 8.2 (1953), pp. 125–130.

[66] Edward G. Effros and Soren Winkler. "Matrix Convexity: Operator Analogues of the Bipolar and Hahn–Banach Theorems". In: *Journal of Functional Analysis* 144.1 (1997), pp. 117–152. ISSN: 0022-1236. DOI: https://doi.org/10.1006/jfan.1996.2958
URL: https://www.sciencedirect.com/science/article/pii/S0022123696929584.

[67] Gábor Elek. "Note on limits of finite graphs". In: *Combinatorica* 27.4 (2007), pp. 503–507.

[68] Gábor Elek and Balázs Szegedy. "A measure-theoretic approach to the theory of dense hypergraphs". In: *Advances in Mathematics* 231.3 (2012), pp. 1731–1772. ISSN: 0001-8708. DOI: https://doi.org/10.1016/j.aim.2012.06.022
URL: https://www.sciencedirect.com/science/article/pii/S0001870812002435.

[69] Pál Erdös. "On the number of complete subgraphs and circuits contained in graphs". eng. In: *Časopis pro pěstování matematiky* 094.3 (1969), pp. 290–296
URL: http://eudml.org/doc/19428.

[70] Benson Farb. "Representation stability". In: *arXiv preprint arXiv:1404.4065* (2014).

[71] Hamza Fawzi, Omar Fawzi, and Samuel O Scalet. "Certified algorithms for equilibrium states of local quantum Hamiltonians". In: *arXiv preprint arXiv:2311.18706* (2023).

[72] Hamza Fawzi, Omar Fawzi, and Samuel O. Scalet. "Entropy constraints for ground energy optimization". In: *Journal of Mathematical Physics* 65.3 (2024), p. 032201. ISSN: 0022-2488. DOI: 10.1063/5.0159108. eprint: https://pubs.aip.org/aip/jmp/article-pdf/doi/10.1063/5.0159108/19702881/032201_1_5.0159108.pdf
URL: https://doi.org/10.1063/5.0159108.

[73] Hamza Fawzi and James Saunderson. "Optimal Self-Concordant Barriers for Quantum Relative Entropies". In: *SIAM Journal on Optimization* 33.4 (2023), pp. 2858–2884. DOI: 10.1137/22M1500216. eprint: https://doi.org/10.1137/22M1500216
URL: https://doi.org/10.1137/22M1500216.

[74] Hamza Fawzi, James Saunderson, and Pablo A Parrilo. "Semidefinite approximations of the matrix logarithm". In: *Foundations of Computational Mathematics* 19.2 (2019), pp. 259–296.

[75] Hamza Fawzi, James Saunderson, and Pablo A. Parrilo. "Equivariant Semidefinite Lifts and Sum-of-Squares Hierarchies". In: *SIAM Journal on Optimization* 25.4 (2015), pp. 2212–2243. DOI: 10.1137/140966265. eprint: https://doi.org/10.1137/140966265
URL: https://doi.org/10.1137/140966265.

[76] Hamza Fawzi, James Saunderson, and Pablo A. Parrilo. "Equivariant Semidefinite Lifts of Regular Polygons". In: *Mathematics of Operations Research* 42.2 (2017), pp. 472–494. DOI: 10.1287/moor.2016.0813. eprint: https://doi.org/10.1287/moor.2016.0813
URL: https://doi.org/10.1287/moor.2016.0813.

[77] Hamza Fawzi et al. "Geometry and Optimization in Quantum Information". In: *Oberwolfach Reports* 18.4 (2022), pp. 2665–2720.

[78] Hamza Fawzi et al. "Lifting for Simplicity: Concise Descriptions of Convex Sets". In: *SIAM Review* 64.4 (2022), pp. 866–918. DOI: 10.1137/20M1324417. eprint: https://doi.org/10.1137/20M1324417
URL: https://doi.org/10.1137/20M1324417.

[79] Marc Finzi, Max Welling, and Andrew Gordon Gordon Wilson. "A Practical Method for Constructing Equivariant Multilayer Perceptrons for Arbitrary Matrix Groups". In: *Proceedings of the 38th International Conference on Machine Learning*. Ed. by Marina Meila and Tong Zhang. Vol. 139. Proceedings of Machine Learning Research. PMLR, 2021, pp. 3318–3328
URL: https://proceedings.mlr.press/v139/finzi21a.html.

[80] Nicolas Fournier. "Convergence of the empirical measure in expected wasserstein distance: non-asymptotic explicit bounds in $\mathbb{R}^d$". In: *ESAIM: Probability and Statistics* 27 (2023), pp. 749–775.

[81] William Fulton and Joe Harris. *Representation theory: a first course*. Vol. 129. Springer, 2013.

[82] Takashi Furuya, Maarten V. de Hoop, and Gabriel Peyré. "Transformers are Universal In-context Learners". In: *The Thirteenth International Conference on Learning Representations*. 2025
URL: https://openreview.net/forum?id=6S4WQD1LZR.

[83] Takashi Furuya, Davide Murari, and Carola-Bibiane Schönlieb. "Approximation Theory for Lipschitz Continuous Transformers". In: *arXiv preprint arXiv:2602.15503* (2026).

[84] Nir Gadish. "Categories of FI type: A unified approach to generalizing representation stability and character polynomials". In: *Journal of Algebra* 480 (2017), pp. 450–486. ISSN: 0021-8693. DOI: https://doi.org/10.

1016/j.jalgebra.2017.03.010
URL: https://www.sciencedirect.com/science/article/pii/S0021869317301849.

[85] Wee Liang Gan and Liping Li. "An inductive machinery for representations of categories with shift functors". In: *Transactions of the American Mathematical Society* 371.12 (2019), pp. 8513–8534.

[86] Karin Gatermann and Pablo A. Parrilo. "Symmetry groups, semidefinite programs, and sums of squares". In: *Journal of Pure and Applied Algebra* 192.1 (2004), pp. 95–128. ISSN: 0022-4049. DOI: https://doi.org/10.1016/j.jpaa.2003.12.011
URL: https://www.sciencedirect.com/science/article/pii/S0022404904000131.

[87] Justin Gilmer et al. "Neural Message Passing for Quantum Chemistry". In: *Proceedings of the 34th International Conference on Machine Learning*. Ed. by Doina Precup and Yee Whye Teh. Vol. 70. Proceedings of Machine Learning Research. PMLR, 2017, pp. 1263–1272
URL: https://proceedings.mlr.press/v70/gilmer17a.html.

[88] Michel X Goemans. "Smallest compact formulation for the permutahedron". In: *Mathematical Programming* 153.1 (2015), pp. 5–11.

[89] Michel X Goemans and David P Williamson. "Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming". In: *Journal of the ACM (JACM)* 42.6 (1995), pp. 1115–1145.

[90] Craig Gotsman and Sivan Toledo. "On the Computation of Null Spaces of Sparse Rectangular Matrices". In: *SIAM Journal on Matrix Analysis and Applications* 30.2 (2008), pp. 445–463. DOI: 10.1137/050638369. eprint: https://doi.org/10.1137/050638369
URL: https://doi.org/10.1137/050638369.

[91] Joao Gouveia, Pablo A Parrilo, and Rekha R Thomas. "Lifts of convex sets and cone factorizations". In: *Mathematics of Operations Research* 38.2 (2013), pp. 248–264.

[92] Sander Gribling, David De Laat, and Monique Laurent. "Lower bounds on matrix factorization ranks via noncommutative polynomial optimization". In: *Foundations of Computational Mathematics* 19.5 (2019), pp. 1013–1070.

[93] Sander Gribling, David de Laat, and Monique Laurent. "Bounds on entanglement dimensions and quantum graph parameters via noncommutative polynomial optimization". In: *Mathematical programming* 170 (2018), pp. 5–42.

[94] I. Gutman and N. Trinajstić. "Graph theory and molecular orbitals. Total ϕ-electron energy of alternant hydrocarbons". In: *Chemical Physics Letters* 17.4 (1972), pp. 535–538. ISSN: 0009-2614. DOI: https://doi.org/10.

1016/0009-2614(72)85099-1
URL: https://www.sciencedirect.com/science/article/pii/0009261472850991.

[95] Nebojša Gvozdenović and Monique Laurent. "Semidefinite bounds for the stability number of a graph via sums of squares of polynomials". In: *Mathematical Programming* 110 (2007), pp. 145–173.

[96] Mahdi Hashemi. "Enlarging smaller images before inputting into convolutional neural network: zero-padding vs. interpolation". In: *Journal of Big Data* 6.1 (2019), pp. 1–13.

[97] Hamed Hatami and Serguei Norine. "Undecidability of linear inequalities in graph homomorphism densities". In: *Journal of the American Mathematical Society* 24.2 (2011), pp. 547–565.

[98] David Haussler. "Sphere packing numbers for subsets of the Boolean *n*-cube with bounded Vapnik–Chervonenkis dimension". In: *Journal of Combinatorial Theory, Series A* 69.2 (1995), pp. 217–232. ISSN: 0097-3165. DOI: https://doi.org/10.1016/0097-3165(95)90052-7
URL: https://www.sciencedirect.com/science/article/pii/0097316595900527.

[99] Xionglei He and Jianzhi Zhang. "Why Do Hubs Tend to Be Essential in Protein Networks?" In: *PLOS Genetics* 2.6 (2006), pp. 1–9. DOI: 10.1371/journal.pgen.0020088
URL: https://doi.org/10.1371/journal.pgen.0020088.

[100] J Helton, Igor Klep, and Scott McCullough. "Matrix convex hulls of free semialgebraic sets". In: *Transactions of the American Mathematical Society* 368.5 (2016), pp. 3105–3139.

[101] J. William Helton, Igor Klep, and Scott McCullough. "Chapter 8: Free Convex Algebraic Geometry". In: *Semidefinite Optimization and Convex Algebraic Geometry*. Society for Industrial and Applied Mathematics, 2012, pp. 341–405. DOI: 10.1137/1.9781611972290.ch8. eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9781611972290.ch8
URL: https://epubs.siam.org/doi/abs/10.1137/1.9781611972290.ch8.

[102] Edwin Hewitt and Leonard J. Savage. "Symmetric Measures on Cartesian Products". In: *Transactions of the American Mathematical Society* 80.2 (1955), pp. 470–501. ISSN: 00029947, 10886850
URL: http://www.jstor.org/stable/1992999 (visited on 05/29/2025).

[103] Douglas N Hoover. "Relations on Probability Spaces and Arrays of Random Variables". In: *Institute for Advanced Study* (1979).

[104] Xinting Huang et al. "A Formal Framework for Understanding Length Generalization in Transformers". In: *International Conference on Learning Representations*. Ed. by Y. Yue et al. Vol. 2025. 2025, pp. 58095–58179 URL: https://proceedings.iclr.cc/paper_files/paper/2025/file/928170bcb050fe64a63fe781b82265aa-Paper-Conference.pdf.

[105] Felix Huber. "Positive maps and trace polynomials from the symmetric group". In: *Journal of Mathematical Physics* 62.2 (2021), p. 022203. ISSN: 0022-2488. DOI: 10.1063/5.0028856. eprint: https://pubs.aip.org/aip/jmp/article-pdf/doi/10.1063/5.0028856/16099609/022203_1_online.pdf URL: https://doi.org/10.1063/5.0028856.

[106] D. B. Hunter. "The positive-definiteness of the complete symmetric functions of even order". In: *Mathematical Proceedings of the Cambridge Philosophical Society* 82.2 (1977), pp. 255–258. DOI: 10.1017/S030500410005386X.

[107] Matthew O. Jackson and Alison Watts. "The Evolution of Social and Economic Networks". In: *Journal of Economic Theory* 106.2 (2002), pp. 265–295. ISSN: 0022-0531. DOI: https://doi.org/10.1006/jeth.2001.2903 URL: https://www.sciencedirect.com/science/article/pii/S0022053101929035.

[108] Svante Janson. "Graphons, cut norm and distance, couplings and rearrangements". In: *New York Journal of Mathematics* 4 (2013). NYJM Monographs.

[109] Svante Janson. "On convergence for graphexes". In: *European Journal of Combinatorics* 104 (2022), p. 103549. ISSN: 0195-6698. DOI: https://doi.org/10.1016/j.ejc.2022.103549 URL: https://www.sciencedirect.com/science/article/pii/S0195669822000452.

[110] Svante Janson. "On edge exchangeable random graphs". In: *Journal of statistical physics* 173.3 (2018), pp. 448–484.

[111] John Jumper et al. "Highly accurate protein structure prediction with AlphaFold". In: *Nature* 596.7873 (2021), pp. 583–589.

[112] Olav Kallenberg. "Characterization and convergence of random measures and point processes". In: *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 27.1 (1973), pp. 9–21.

[113] Olav Kallenberg. "Exchangeable random measures in the plane". In: *Journal of Theoretical Probability* 3.1 (1990), pp. 81–136.

[114] Olav Kallenberg. *Foundations of modern probability*. Springer, 1997.

[115] Olav Kallenberg et al. *Probabilistic symmetries and invariance principles*. Vol. 9. Springer, 2005.

[116] Raymond Kan. *Transition Matrices between Symmetric Polynomials*. (https://www.mathworks.com/matlabcentral/fileexchange/136299-transition-matrices-between-symmetric-polynomials). MATLAB Central File Exchange. [Retrieved May 18, 2025]. 2023.

[117] George Em Karniadakis et al. "Physics-informed machine learning". In: *Nature Reviews Physics* 3.6 (2021), pp. 422–440.

[118] Leo Katz. "A New Status Index Derived from Sociometric Analysis". In: *Psychometrika* 18.1 (1953), pp. 39–43. DOI: 10.1007/BF02289026.

[119] Amirhossein Kazemnejad et al. "The Impact of Positional Encoding on Length Generalization in Transformers". In: *Advances in Neural Information Processing Systems*. Ed. by A. Oh et al. Vol. 36. Curran Associates, Inc., 2023, pp. 24892–24928
URL: https://proceedings.neurips.cc/paper_files/paper/2023/file/4e85362c02172c0c6567ce593122d31c-Paper-Conference.pdf.

[120] Apoorva Khare and Terence Tao. "On the sign patterns of entrywise positivity preservers in fixed dimension". In: *American Journal of Mathematics* 143.6 (2021), pp. 1863–1929.

[121] Diederik P. Kingma and Jimmy Ba. "Adam: A Method for Stochastic Optimization". In: *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*. Ed. by Yoshua Bengio and Yann LeCun. 2015
URL: http://arxiv.org/abs/1412.6980.

[122] J. F. C. Kingman. "The Representation of Partition Structures". In: *Journal of the London Mathematical Society* s2-18.2 (1978), pp. 374–380. ISSN: 0024-6107. DOI: 10.1112/jlms/s2-18.2.374. eprint: https://academic.oup.com/jlms/article-pdf/s2-18/2/374/2788610/s2-18-2-374.pdf
URL: https://doi.org/10.1112/jlms/s2-18.2.374.

[123] John Frank Charles Kingman. "Random partitions in population genetics". In: *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* 361.1704 (1978), pp. 1–20. ISSN: 0080-4630. DOI: 10.1098/rspa.1978.0089. eprint: https://royalsocietypublishing.org/rspa/article-pdf/361/1704/1/62915/rspa.1978.0089.pdf
URL: https://doi.org/10.1098/rspa.1978.0089.

[124] Igor Klep, Victor Magron, and Jurij Volčič. "Sums of squares certificates for polynomial moment inequalities". In: *Foundations of Computational Mathematics* (2025), pp. 1–43.

[125] Igor Klep, Špela Špenko, and Jurij Volčič. "Positive trace polynomials and the universal Procesi–Schacher conjecture". In: *Proceedings of the London Mathematical Society* 117.6 (2018), pp. 1101–1134. DOI: https://

doi.org/10.1112/plms.12156. eprint: https://londmathsoc.onlinelibrary.wiley.com/doi/pdf/10.1112/plms.12156
URL: https://londmathsoc.onlinelibrary.wiley.com/doi/abs/10.1112/plms.12156.

[126] Andrei Nikolaevich Kolmogorov and Vladimir Mikhailovich Tikhomirov. "ε-entropy and ε-capacity of sets in function spaces". In: *Uspekhi Matematicheskikh Nauk* 14.2 (1959), pp. 3–86.

[127] Risi Kondor, Zhen Lin, and Shubhendu Trivedi. "Clebsch–Gordan Nets: a Fully Fourier Space Spherical Convolutional Neural Network". In: *Advances in Neural Information Processing Systems*. Ed. by S. Bengio et al. Vol. 31. Curran Associates, Inc., 2018
URL: https://proceedings.neurips.cc/paper_files/paper/2018/file/a3fc981af450752046be179185ebc8b5-Paper.pdf.

[128] Risi Kondor and Shubhendu Trivedi. "On the Generalization of Equivariance and Convolution in Neural Networks to the Action of Compact Groups". In: *Proceedings of the 35th International Conference on Machine Learning*. Ed. by Jennifer Dy and Andreas Krause. Vol. 80. Proceedings of Machine Learning Research. PMLR, 2018, pp. 2747–2755
URL: https://proceedings.mlr.press/v80/kondor18a.html.

[129] Pawel Kowal. *Null space of a sparse matrix*. https://www.mathworks.com/matlabcentral/fileexchange/11120-null-space-of-a-sparse-matrix. [Retrieved July 12, 2022]. 2006.

[130] Tom-Lukas Kriel. "An introduction to matrix convex sets and free spectrahedra". In: *Complex Analysis and Operator Theory* 13.7 (2019), pp. 3251–3335.

[131] Ani Kristo et al. "The case for a learned sorting algorithm". In: *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*. 2020, pp. 1001–1016.

[132] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: *Commun. ACM* 60.6 (2017), pp. 84–90. ISSN: 0001-0782. DOI: 10.1145/3065386
URL: https://doi.org/10.1145/3065386.

[133] Jean B. Lasserre. "Global Optimization with Polynomials and the Problem of Moments". In: *SIAM Journal on Optimization* 11.3 (2001), pp. 796–817. DOI: 10.1137/S1052623400366802. eprint: https://doi.org/10.1137/S1052623400366802
URL: https://doi.org/10.1137/S1052623400366802.

[134] Monique Laurent and Luis Felipe Vargas. "Exactness of Parrilo's Conic Approximations for Copositive Matrices and Associated Low Order Bounds for the Stability Number of a Graph". In: *Mathematics of Operations Research* 48.2 (2023), pp. 1017–1043. DOI: 10.1287/moor.2022.1290. eprint:

https://doi.org/10.1287/moor.2022.1290
URL: https://doi.org/10.1287/moor.2022.1290.

[135] Thien Le and Stefanie Jegelka. "Limits, approximation and size transferability for GNNs on sparse graphs via graphops". In: *Advances in Neural Information Processing Systems*. Ed. by A. Oh et al. Vol. 36. Curran Associates, Inc., 2023, pp. 41305–41342
URL: https://proceedings.neurips.cc/paper_files/paper/2023/file/8154c89c8d3612d39fd1ed6a20f4bab1-Paper-Conference.pdf.

[136] Joonkyung Lee and Bjarne Schülke. "Convex graphon parameters and graph norms". In: *Israel Journal of Mathematics* 242.2 (2021), pp. 549–563.

[137] Eitan Levin and Venkat Chandrasekaran. "Any Dimensional Learning by Sampling". In: *In Preparation* (2026).
E.L participated in the conception of the project, performed the analysis, and participated in the writing of the manuscript.

[138] Eitan Levin and Venkat Chandrasekaran. "Any-Dimensional Polynomial Optimization via de Finetti Theorems". In: *arXiv preprint arXiv:2507.15632* (2025).
E.L participated in the conception of the project, performed the analysis, undertook the numerical experiments, and participated in the writing of the manuscript.
DOI: 10.48550/arXiv.2507.15632.

[139] Eitan Levin and Venkat Chandrasekaran. "Dimension-free descriptions of convex sets". In: *arXiv preprint arXiv:2307.04230* (2023).
E.L participated in the conception of the project, performed the analysis, undertook the numerical experiments, and participated in the writing of the manuscript.
DOI: 10.48550/arXiv.2307.04230.

[140] Eitan Levin and Venkat Chandrasekaran. "Limits of Weighted Graphs via Random Quotients". In: *arXiv preprint arXiv:2512.23149* (2025).
E.L participated in the conception of the project, performed the analysis, and participated in the writing of the manuscript.
DOI: 10.48550/arXiv.2512.23149.

[141] Eitan Levin and Mateo Díaz. "Any-dimensional equivariant neural networks". In: *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*. Ed. by Sanjoy Dasgupta, Stephan Mandt, and Yingzhen Li. Vol. 238. Proceedings of Machine Learning Research.
E.L participated in the conception of the project, performed the analysis, and participated in the writing of the manuscript.
PMLR, 2024, pp. 2773–2781
URL: https://proceedings.mlr.press/v238/levin24a.html.

[142] Eitan Levin et al. "On Transferring Transferability: Towards a Theory for Size Generalization". In: *Advances in Neural Information Processing Systems*. Ed. by D. Belgrave et al. Vol. 38.
E.L participated in the conception of the project, performed some of the analysis, and participated in the writing of the manuscript.
Curran Associates, Inc., 2025, pp. 67015–67083
URL: https://proceedings.neurips.cc/paper_files/paper/2025/file/60b393761eea35bf491687c46e4f1a5a-Paper-Conference.pdf.

[143] Cong Han Lim and Stephen Wright. "Beyond the Birkhoff Polytope: Convex Relaxations for Vector Permutation Problems". In: *Advances in Neural Information Processing Systems*. Ed. by Z. Ghahramani et al. Vol. 27. Curran Associates, Inc., 2014
URL: https://proceedings.neurips.cc/paper_files/paper/2014/file/208e43f0e45c4c78cafadb83d2888cb6-Paper.pdf.

[144] Lek-Heng Lim and Bradley J. Nelson. "What is an equivariant neural network?" In: *Notices Amer. Math. Soc.* 70.4 (2023), pp. 619–625. ISSN: 0002-9920. DOI: 10.1090/noti2666
URL: https://doi.org/10.1090/noti2666.

[145] Tianyang Lin et al. "A survey of transformers". In: *AI Open* 3 (2022), pp. 111–132. ISSN: 2666-6510. DOI: https://doi.org/10.1016/j.aiopen.2022.10.001
URL: https://www.sciencedirect.com/science/article/pii/S2666651022000146.

[146] László Lovász. *Large networks and graph limits*. Vol. 60. American Mathematical Soc., 2012.

[147] László Lovász and Balázs Szegedy. "Limits of dense graph sequences". In: *Journal of Combinatorial Theory, Series B* 96.6 (2006), pp. 933–957. ISSN: 0095-8956. DOI: https://doi.org/10.1016/j.jctb.2006.05.002
URL: https://www.sciencedirect.com/science/article/pii/S0095895606000517.

[148] László Lovász and Balázs Szegedy. "Random graphons and a weak Positivstellensatz for graphs". In: *Journal of Graph Theory* 70.2 (2012), pp. 214–225. DOI: https://doi.org/10.1002/jgt.20611. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jgt.20611
URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/jgt.20611.

[149] László Lovász and Balázs Szegedy. "Szemerédi's lemma for the analyst". In: *GAFA Geometric And Functional Analysis* 17.1 (2007), pp. 252–270.

[150] Ulrike von Luxburg and Olivier Bousquet. "Distance-based classification with Lipschitz functions". In: *Journal of Machine Learning Research* 5.Jun (2004), pp. 669–695.

[151] Ian Grant Macdonald. *Symmetric functions and Hall polynomials*. Oxford university press, 1998.

[152] Haggai Maron et al. "Invariant and Equivariant Graph Networks". In: *International Conference on Learning Representations*. 2019 URL: https://openreview.net/forum?id=Syx72jC9tm.

[153] Sohir Maskey, Ron Levie, and Gitta Kutyniok. "Transferability of graph neural networks: an extended graphon approach". In: *Applied and Computational Harmonic Analysis* 63 (2023), pp. 48–83.

[154] Sohir Maskey et al. "Generalization Analysis of Message Passing Neural Networks on Large Random Graphs". In: *Advances in Neural Information Processing Systems*. Ed. by S. Koyejo et al. Vol. 35. Curran Associates, Inc., 2022, pp. 4805–4817 URL: https://proceedings.neurips.cc/paper_files/paper/2022/file/1eeaae7c89d9484926db6974b6ece564-Paper-Conference.pdf.

[155] Charles A. Micchelli and Massimiliano Pontil. "On Learning Vector-Valued Functions". In: *Neural Computation* 17.1 (2005), pp. 177–204. DOI: 10.1162/0899766052530802.

[156] Tomáš Mikolov et al. "Extensions of recurrent neural network language model". In: *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2011, pp. 5528–5531. DOI: 10.1109/ICASSP.2011.5947611.

[157] James A Mingo and Roland Speicher. *Free probability and random matrices*. Vol. 35. Springer, 2017.

[158] Vishal Monga, Yuelong Li, and Yonina C. Eldar. "Algorithm Unrolling: Interpretable, Efficient Deep Learning for Signal and Image Processing". In: *IEEE Signal Processing Magazine* 38.2 (2021), pp. 18–44. DOI: 10.1109/MSP.2020.3016905.

[159] T. S. Motzkin and E. G. Straus. "Maxima for Graphs and a New Proof of a Theorem of Turán". In: *Canadian Journal of Mathematics* 17 (1965), pp. 533–540. DOI: 10.4153/CJM-1965-053-6.

[160] Philippe Moustrou et al. "Symmetry Reduction in AM/GM-Based Optimization". In: *SIAM Journal on Optimization* 32.2 (2022), pp. 765–785. DOI: 10.1137/21M1405691. eprint: https://doi.org/10.1137/21M1405691 URL: https://doi.org/10.1137/21M1405691.

[161] Riley Murray, Venkat Chandrasekaran, and Adam Wierman. "Newton polytopes and relative entropy optimization". In: *Foundations of Computational Mathematics* 21.6 (2021), pp. 1703–1737.

[162] Riley Murray, Venkat Chandrasekaran, and Adam Wierman. "Signomial and polynomial optimization via relative entropy and partial dualization". In: *Mathematical Programming Computation* 13.2 (2021), pp. 257–295.

[163] Netzschleuder. *Complete C. elegans neurons (2019)*. https://networks.skewed.de/net/celegans_2019. Accessed: 2025-12-17.

[164] Mark Newman. *Networks*. Oxford University Press, 2018. ISBN: 9780198805090. DOI: 10.1093/oso/9780198805090.001.0001 URL: https://doi.org/10.1093/oso/9780198805090.001.0001.

[165] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010. DOI: 10.1017/CBO9780511976667.

[166] Nelson Niu and David I. Spivak. *Polynomial Functors: A Mathematical Theory of Interaction*. London Mathematical Society Lecture Note Series. Cambridge University Press, 2025.

[167] Morton E O'Kelly. "A geographer's analysis of hub-and-spoke networks". In: *Journal of Transport Geography* 6.3 (1998), pp. 171–186. ISSN: 0966-6923. DOI: https://doi.org/10.1016/S0966-6923(98)00010-6 URL: https://www.sciencedirect.com/science/article/pii/S0966692398000106.

[168] Gregory Ongie et al. "Deep learning techniques for inverse problems in imaging". In: *IEEE Journal on Selected Areas in Information Theory* 1.1 (2020), pp. 39–56.

[169] Peter Orbanz. "Projective limit random probabilities on Polish spaces". In: *Electronic Journal of Statistics* 5.none (2011), pp. 1354–1373. DOI: 10.1214/11-EJS641.

[170] Christopher C Paige and Michael A Saunders. "LSQR: An algorithm for sparse linear equations and sparse least squares". In: *ACM Transactions on Mathematical Software (TOMS)* 8.1 (1982), pp. 43–71.

[171] Olaf Parczyk et al. "New Ramsey multiplicity bounds and search heuristics". In: *Foundations of Computational Mathematics* (2024), pp. 1–38.

[172] Pablo A Parrilo. "Semidefinite programming relaxations for semialgebraic problems". In: *Mathematical Programming* 96 (2003), pp. 293–320.

[173] Pablo A. Parrilo. "Chapter 3: Polynomial Optimization, Sums of Squares, and Applications". In: *Semidefinite Optimization and Convex Algebraic Geometry*. Society for Industrial and Applied Mathematics, 2012, pp. 47–157. DOI: 10.1137/1.9781611972290.ch3. eprint: https://epubs.siam.

org/doi/pdf/10.1137/1.9781611972290.ch3
URL: https://epubs.siam.org/doi/abs/10.1137/1.9781611972290.ch3.

[174] Pablo A. Parrilo. "Polynomial games and sum of squares optimization". In: *Proceedings of the 45th IEEE Conference on Decision and Control*. 2006, pp. 2855–2860. DOI: 10.1109/CDC.2006.377261.

[175] S. Pironio, M. Navascués, and A. Acín. "Convergent Relaxations of Polynomial Optimization Problems with Noncommuting Variables". In: *SIAM Journal on Optimization* 20.5 (2010), pp. 2157–2180. DOI: 10.1137/090760155. eprint: https://doi.org/10.1137/090760155
URL: https://doi.org/10.1137/090760155.

[176] Jim Pitman. "Exchangeable and partially exchangeable random partitions". In: *Probability theory and related fields* 102.2 (1995), pp. 145–158.

[177] Sven C. Polak. "Symmetry Reduction to Optimize a Graph-based Polynomial From Queueing Theory". In: *SIAM Journal on Applied Algebra and Geometry* 6.2 (2022), pp. 243–266. DOI: 10.1137/21M1413298. eprint: https://doi.org/10.1137/21M1413298
URL: https://doi.org/10.1137/21M1413298.

[178] Renata Possobon and Christian S. Rodrigues. "Geometric properties of disintegration of measures". In: *Ergodic Theory and Dynamical Systems* 45.5 (2025), pp. 1619–1648. DOI: 10.1017/etds.2024.74.

[179] Andres Potapczynski et al. "CoLA: Exploiting Compositional Structure for Automatic and Efficient Numerical Linear Algebra". In: *Advances in Neural Information Processing Systems*. Ed. by A. Oh et al. Vol. 36. Curran Associates, Inc., 2023, pp. 43894–43917
URL: https://proceedings.neurips.cc/paper_files/paper/2023/file/88c3c482430a62d35e03926a22e4b67e-Paper-Conference.pdf.

[180] Jonathan D Power et al. "Evidence for hubs in human functional brain networks". In: *Neuron* 79.4 (2013), pp. 798–813.

[181] C Procesi. "Positive symmetric functions". In: *Advances in Mathematics* 29.2 (1978), pp. 219–225. ISSN: 0001-8708. DOI: https://doi.org/10.1016/0001-8708(78)90011-7
URL: https://www.sciencedirect.com/science/article/pii/0001870878900117.

[182] Charles R. Qi et al. "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017.

[183] Mohan Ravichandran. "The Schur-Horn theorem in von Neumann algebras". In: *arXiv preprint arXiv:1209.0909* (2012).

[184] Annie Raymond, Mohit Singh, and Rekha R Thomas. "Symmetry in Turán sums of squares polynomials from flag algebras". In: *Algebraic Combinatorics* 1.2 (2018), pp. 249–274.

[185] Annie Raymond et al. "Symmetric sums of squares over k-subset hypercubes". In: *Mathematical Programming* 167.2 (2018), pp. 315–354.

[186] Alexander A. Razborov. "Flag algebras". In: *The Journal of Symbolic Logic* 72.4 (2007), pp. 1239–1282. DOI: 10.2178/jsl/1203350785.

[187] Cordian Riener. "Symmetric semi-algebraic sets and non-negativity of symmetric polynomials". In: *Journal of Pure and Applied Algebra* 220.8 (2016), pp. 2809–2815.

[188] Cordian Riener et al. "Exploiting symmetries in SDP-relaxations for polynomial optimization". In: *Mathematics of Operations Research* 38.1 (2013), pp. 122–141.

[189] R Tyrrell Rockafellar. *Convex analysis*. Vol. 18. Princeton university press, 1970.

[190] Abel Rodriguez and Peter Müller. "Nonparametric bayesian inference". In: *NSF-CBMS Regional Conference Series in Probability and Statistics*. Vol. 9. JSTOR. 2013, pp. i–110.

[191] Walter Rudin. *Functional Analysis*. 2nd ed. New York: McGraw-Hill, 1991.

[192] Walter Rudin. *Real and Complex Analysis*. 3rd ed. New York: McGraw-Hill, 1987.

[193] Luana Ruiz, Luiz Chamon, and Alejandro Ribeiro. "Graphon Neural Networks and the Transferability of Graph Neural Networks". In: *Advances in Neural Information Processing Systems*. Ed. by H. Larochelle et al. Vol. 33. Curran Associates, Inc., 2020, pp. 1702–1712 URL: https://proceedings.neurips.cc/paper_files/paper/2020/file/12bcd658ef0a540cabc36cdf2b1046fd-Paper.pdf.

[194] Steven Sam and Andrew Snowden. "GL-equivariant modules over polynomial rings in infinitely many variables". In: *Transactions of the American Mathematical Society* 368.2 (2016), pp. 1097–1158.

[195] Steven Sam and Andrew Snowden. "Gröbner methods for representations of combinatorial categories". In: *Journal of the American Mathematical Society* 30.1 (2017), pp. 159–203.

[196] Steven V Sam. "Structures in representation stability". In: *Notices of the American Mathematical Society* 67.1 (2020).

[197] Steven V Sam and Andrew Snowden. "STABILITY PATTERNS IN REPRESENTATION THEORY". In: *Forum of Mathematics, Sigma* 3 (2015), e11. DOI: 10.1017/fms.2015.10.

[198] Raman Sanyal and James Saunderson. "Spectral polyhedra". In: *arXiv preprint arXiv:2001.04361* (2020).

[199] J. Saunderson, P. A. Parrilo, and A. S. Willsky. "Semidefinite Descriptions of the Convex Hull of Rotation Matrices". In: *SIAM Journal on Optimization* 25.3 (2015), pp. 1314–1343. DOI: 10.1137/14096339X. eprint: https://doi.org/10.1137/14096339X
URL: https://doi.org/10.1137/14096339X.

[200] Franco Scarselli et al. "The Graph Neural Network Model". In: *IEEE Transactions on Neural Networks* 20.1 (2009), pp. 61–80. DOI: 10.1109/TNN.2008.2005605.

[201] Alberto Seeger. "Convex Analysis of Spectrally Defined Matrix Functions". In: *SIAM Journal on Optimization* 7.3 (1997), pp. 679–696. DOI: 10.1137/S1052623495288866. eprint: https://doi.org/10.1137/S1052623495288866
URL: https://doi.org/10.1137/S1052623495288866.

[202] Jean-Pierre Serre et al. *Linear representations of finite groups*. Vol. 42. Springer, 1977.

[203] Qinfeng Shi et al. "Hash kernels for structured data". In: *Journal of Machine Learning Research* 10.11 (2009).

[204] Alexander F Sidorenko. "Cycles in graphs and functional inequalities". In: *Matematicheskie Zametki* 46.5 (1989), pp. 72–79.

[205] Alexander F Sidorenko. "Inequalities for functionals generated by bipartite graphs". In: *Diskretnaya Matematika* 3.3 (1991), pp. 50–65.

[206] Richard Socher et al. "Parsing natural scenes and natural language with recursive neural networks". In: *Proceedings of the 28th international conference on machine learning (ICML-11)*. 2011, pp. 129–136.

[207] Ben Stabler, Hillel Bar-Gera, and Elizabeth Sall. *Transportation Networks for Research*. https://github.com/bstabler/TransportationNetworks. Accessed: 2025-11-18.

[208] A. J. Stam. "Distance between sampling with and without replacement". In: *Statistica Neerlandica* 32.2 (1978), pp. 81–91. DOI: https://doi.org/10.1111/j.1467-9574.1978.tb01387.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9574.1978.tb01387.x
URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9574.1978.tb01387.x.

[209] Richard P. Stanley and Sergey Fomin. *Enumerative Combinatorics*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1999.

[210] Shazia Tabassum et al. “Social network analysis: An overview”. In: *WIREs Data Mining and Knowledge Discovery* 8.5 (2018), e1256. DOI: https://doi.org/10.1002/widm.1256. eprint: https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/widm.1256
URL: https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/widm.1256.

[211] Vlad Timofte. “On the positivity of symmetric polynomial functions.: Part I: General results”. In: *Journal of Mathematical Analysis and Applications* 284.1 (2003), pp. 174–190.

[212] Raphael J. L. Townshend et al. “Geometric deep learning of RNA structure”. In: *Science* 373.6558 (2021), pp. 1047–1051. DOI: 10.1126/science.abe5650. eprint: https://www.science.org/doi/pdf/10.1126/science.abe5650
URL: https://www.science.org/doi/abs/10.1126/science.abe5650.

[213] Francesco Vaccarino. “The ring of multisymmetric functions”. en. In: *Annales de l’Institut Fourier* 55.3 (2005), pp. 717–731. DOI: 10.5802/aif.2111
URL: https://www.numdam.org/articles/10.5802/aif.2111/.

[214] Martijn P Van den Heuvel and Olaf Sporns. “Network hubs in the human brain”. In: *Trends in cognitive sciences* 17.12 (2013), pp. 683–696.

[215] Dinh Van Le and Tim Römer. “Theorems of Carathéodory, Minkowski-Weyl, and Gordan up to symmetry”. In: *arXiv preprint arXiv:2110.10657* (2021).

[216] Vladimir Vapnik. *The Nature of Statistical Learning Theory*. Springer, 1999. DOI: 10.1007/978-1-4757-3264-1.

[217] Victor Veitch and Daniel M Roy. “The class of random graphs arising from exchangeable random measures”. In: *arXiv preprint arXiv:1512.03099* (2015).

[218] Victor Veitch and Daniel M. Roy. “Sampling and estimation for (sparse) exchangeable graphs”. In: *The Annals of Statistics* 47.6 (2019), pp. 3274–3299. DOI: 10.1214/18-AOS1778
URL: https://doi.org/10.1214/18-AOS1778.

[219] Clément Vignac, Andreas Loukas, and Pascal Frossard. “Building powerful and equivariant graph neural networks with structural message-passing”. In: *Advances in Neural Information Processing Systems*. Ed. by H. Larochelle et al. Vol. 33. Curran Associates, Inc., 2020, pp. 14143–14155
URL: https://proceedings.neurips.cc/paper_files/paper/2020/file/a32d7eeaae19821fd9ce317f3ce952a7-Paper.pdf.

[220] Cédric Villani et al. *Optimal transport: old and new*. Vol. 338. Springer, 2008.

[221] Soledad Villar et al. "Dimensionless machine learning: Imposing exact units equivariance". In: *Journal of Machine Learning Research* 24.109 (2023), pp. 1–32
URL: http://jmlr.org/papers/v24/22-0680.html.

[222] Soledad Villar et al. "Scalars are universal: Equivariant machine learning, structured like classical physics". In: *Advances in Neural Information Processing Systems*. Ed. by M. Ranzato et al. Vol. 34. Curran Associates, Inc., 2021, pp. 28848–28863
URL: https://proceedings.neurips.cc/paper_files/paper/2021/file/f1b0775946bc0329b35b823b86eeb5f5-Paper.pdf.

[223] Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*. Vol. 48. Cambridge university press, 2019.

[224] K.S. Wang and L.C. Grove. "Realizability of representations of finite groups". In: *Journal of Pure and Applied Algebra* 54.2 (1988), pp. 299–310. ISSN: 0022-4049. DOI: https://doi.org/10.1016/0022-4049(88)90036-9
URL: https://www.sciencedirect.com/science/article/pii/0022404988900369.

[225] Maurice Weiler et al. "3D steerable CNNs: Learning rotationally equivariant features in volumetric data". In: *Advances in Neural Information Processing Systems* 31 (2018).

[226] Qingsong Wen et al. "Transformers in time series: a survey". In: *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*. IJCAI '23. Macao, P.R.China, 2023. ISBN: 978-1-956792-03-4. DOI: 10.24963/ijcai.2023/759
URL: https://doi.org/10.24963/ijcai.2023/759.

[227] Jennifer C.H. Wilson. "$\mathrm{FI}_{\mathcal{W}}$-modules and stability criteria for representations of classical Weyl groups". In: *Journal of Algebra* 420 (2014), pp. 269–332. ISSN: 0021-8693. DOI: https://doi.org/10.1016/j.jalgebra.2014.08.010
URL: https://www.sciencedirect.com/science/article/pii/S0021869314004505.

[228] Jenny Wilson. "An introduction to FI–modules and their generalizations". In: *Michigan Representation Stability Week* (2018).

[229] Andy Yang et al. "Length Generalization Bounds for Transformers". In: *arXiv preprint arXiv:2603.02238* (2026).

[230] Mihalis Yannakakis. "Expressing combinatorial optimization problems by Linear Programs". In: *Journal of Computer and System Sciences* 43.3 (1991), pp. 441–466. ISSN: 0022-0000. DOI: https://doi.org/10.1016/0022-0000(91)90024-Y
URL: https://www.sciencedirect.com/science/article/pii/002200009190024Y.

[231] Weichi Yao, Afonso S. Bandeira, and Soledad Villar. "Experimental performance of graph neural networks on random instances of max-cut". In: *Wavelets and Sparsity XVIII*. Ed. by Dimitri Van De Ville, Manos Papadakis, and Yue M. Lu. Vol. 11138. International Society for Optics and Photonics. SPIE, 2019, 111380S. DOI: 10.1117/12.2529608
URL: https://doi.org/10.1117/12.2529608.

[232] Mohammad Hossein Yassaee, Mohammad Reza Aref, and Amin Gohari. "Achievability proof via output statistics of random binning". In: *IEEE Transactions on Information Theory* 60.11 (2014), pp. 6760–6786.

[233] Manzil Zaheer et al. "Deep Sets". In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon et al. Vol. 30. Curran Associates, Inc., 2017
URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/f22e4747da1aa27e363d86d40ff442fe-Paper.pdf.

[234] Bin Zhang and Steve Horvath. "A General Framework for Weighted Gene Co-Expression Network Analysis". In: *Statistical applications in genetics and molecular biology* 4.1 (2005). DOI: doi:10.2202/1544-6115.1128.

[235] Zhiyi Zhang and Michael Grabchak. "Entropic representation and estimation of diversity indices". In: *Journal of Nonparametric Statistics* 28.3 (2016), pp. 563–575. DOI: 10.1080/10485252.2016.1190357. eprint: https://doi.org/10.1080/10485252.2016.1190357
URL: https://doi.org/10.1080/10485252.2016.1190357.

[236] Yongchao Zhou et al. "Transformers Can Achieve Length Generalization But Not Robustly". In: *ICLR 2024 Workshop on Mathematical and Empirical Understanding of Foundation Models*. 2024
URL: https://openreview.net/forum?id=DWkWIh3vFJ.

Appendix A

REPRESENTATIONS OF CATEGORIES

As Remark 2.2.10 shows, the set of embeddings from low to high dimensions in a consistent sequence $\mathscr{V}$ of $\{\mathsf{G}_n\}$-representations, determined by $\{\mathsf{G}_n\}$ and the centralizing subgroups $\{\mathsf{H}_{n,d}\}$ from Definition 2.2.4, play a central role in our framework. These sets of embeddings are conveniently encoded in a category, whose representations are precisely the $\mathscr{V}$-modules of Definition 2.2.5. Morphisms between such representations in the categorical sense coincide with morphisms of sequences. This categorical approach to representation stability was introduced in [46] for the case $\mathsf{G}_n = \mathsf{S}_n$ and the $\mathsf{H}_{n,d}$ from Example 2.2.8, and has since been extended to other groups in [84, 195, 227].

Definition A.0.1. *A (real)* representation *of a category C, also called a C*-module*, is a functor $C \to \mathrm{Inn}_{\mathbb{R}}$ from C to the category whose objects are real inner product spaces and whose morphisms are isometries.* [1]

In other words, a C-module is an assignment of a vector space $\mathbb{V}_n$ to each object $n \in C$ and a linear map $\varphi_{N,n} \colon \mathbb{V}_n \to \mathbb{V}_N$ to each morphism in $\mathrm{Hom}_C(n, N)$ such that compositions are respected. Each $\mathbb{V}_n$ is then a representation of the group $\mathsf{G}_n = \mathrm{End}_C(n)^\times$ of the automorphisms of n in C. Every consistent sequence is a representation of a suitable category.

Definition A.0.2. *Given a consistent sequence $\mathscr{V} = \{(\mathbb{V}_n, \varphi_n)\}$ of $\{\mathsf{G}_n\}$-representations, define a category $C_{\mathscr{V}}$ whose set of objects is $\mathbb{N}$ and whose morphisms are $\mathrm{Hom}_{C_V}(n, N) = \{g\varphi_{N-1}\cdots\varphi_n : g \in \mathsf{G}_N\}$ for $n \leq N$ and zero otherwise. Note that $\mathrm{Hom}_{C_V}(n, N) = \mathsf{G}_N/\mathsf{H}_{N,n}$ where $\mathsf{H}_{N,n} \subseteq \mathsf{G}_N$ is the subgroup of elements acting trivially on $\mathbb{V}_n$.*

If $n \leq N \leq M$, we compose morphisms by setting

$$(g_M\varphi_{M-1}\cdots\varphi_N) \circ (g_N\varphi_{N-1}\cdots\varphi_n) = (g_M g_N)\varphi_{M-1}\cdots\varphi_n,$$

which is well-defined and satisfies the necessary axioms since $\varphi_N, \ldots, \varphi_{M-1}$ are G_N-equivariant when acting on the vector spaces in $\mathscr{V}$.

[1] Representations of categories are typically defined as functors to the category of all vector spaces over a given field. The results in this section continue to hold for functors to the category of all real vector spaces if we allow consistent sequences consisting of non-orthogonal representations and non-isometric linear maps between them.

This definition clearly extends to consistent sequences indexed by posets (Remark 2.4.4). If $\mathscr{U} = \{(\mathbb{U}_n, \psi_n)\}$ is a $\mathscr{V}$-module (Definition 2.2.5), then $\mathscr{U}$ is a $C_{\mathscr{V}}$-module, since sending $n \in \mathbb{N}$ to $\mathbb{U}_n$ and $g\varphi_{N-1}\cdots\varphi_n$ to the map $g\psi_{N-1}\cdots\psi_n$ for each $g \in \mathsf{G}_N$ is a well-defined functor $C_{\mathscr{V}} \to \mathrm{Vect}_{\mathbb{R}}$. Conversely, if $\mathscr{U}$ is a $C_{\mathscr{V}}$-module then it is a $\mathscr{V}$-module since $\mathsf{H}_{n,d}$ acts trivially on the image of $\psi_{N-1}\cdots\psi_d$ by definition of a functor. Furthermore, if $\mathscr{W}, \mathscr{U}$ are $\mathscr{C}_{\mathscr{V}}$-modules, then a morphism of functors $\mathscr{W} \to \mathscr{U}$ (also called a natural transformation) coincides with a morphism of sequences in Definition 2.2.1. Applying the constructions in Section 2.2 to C-modules yields other C-modules.

Example A.0.3. *Here are some examples of the categories resulting from Definition A.0.2.*

(a) The category corresponding to Example 2.1.3 with $\mathsf{G}_n = \mathsf{S}_n$ is (the skeleton of) $C = \mathrm{FI}$, the category whose objects are finite sets and whose morphisms are injections.

(b) The category corresponding to Example 2.1.3 with $\mathsf{G}_n = \mathsf{B}_n$ (resp., D_n) is $C = \mathrm{FI}_{BC}$ (resp., $C = \mathrm{FI}|_D$) defined in [227], whose objects are the sets $[\pm n] := \{\pm 1, \ldots, \pm n\}$ for $n \in \mathbb{N}$ and whose morphisms are injections $f\colon [\pm n] \hookrightarrow [\pm N]$ satisfying $f(-i) = -f(i)$ (and reverse evenly-many signs if $\mathsf{G}_n = \mathsf{D}_n$).

(c) The category corresponding to the graphon sequence in Section 2.4 is the opposite category $C = \mathcal{P}_2^{\mathrm{op}}$ of the category $\mathcal{P}_2$ with objects $[2^n]$ and morphisms which are 2^{N-n}-to-one surjections $[2^N] \to [2^n]$, or equivalently, partitions of $[2^N]$ into 2^n equal parts.

Following [227], we say $C = \mathrm{FI}|_{\mathcal{W}}$ if $C = \mathrm{FI}, \mathrm{FI}|_{BC}$ or $\mathrm{FI}|_D$. (Algebraically) free C-modules are defined exactly as in Definition 2.2.6, see [46, Def. 2.2.2] and [84, Def. 1.8, 3.1] for example. The theory of [84] gives the following result for $C = \mathrm{FI}|_{\mathcal{W}}$, which extends to categories of FI-type introduced in [84].

Theorem A.0.4 ([84, Thm. B(1)])**.** *Tensor products of free $\mathrm{FI}|_{\mathcal{W}}$-modules are free.*

The following result illustrates two further properties of $\mathrm{FI}|_{\mathcal{W}}$-modules, the second of which is included in our Theorem 2.2.11 stated above.

Theorem A.0.5 (Noetherianity and tensor products)**.** *Let $C = \mathrm{FI}|_{\mathcal{W}}$.*

(Noetherianity) *Any submodule of a finitely-generated C-module is finitely-generated.*

(Tensor products) *If $\mathscr{V}_1$ and $\mathscr{V}_2$ are C-modules generated in degrees d_1 and d_2, respectively, then $\mathscr{V}_1 \otimes \mathscr{V}_2$ is generated in degree $d_1 + d_2$.*

Proof. Noetherianity is shown for FI in [46, Thm. 1.13] and for $\mathrm{FI}|_{BC}, \mathrm{FI}|_D$ in [227, Thm. 4.21]. The generation degree bound is shown in [46, Prop. 2.3.6] for FI and in [227, Prop. 5.2] for $\mathrm{FI}|_{BC}, \mathrm{FI}|_D$. □

Noetherianity helps explain the ubiquity of representation stability, while the generation degree bound for tensor products allows us to bound the generation degrees of complicated sequences from degrees of simple ones. The two properties in Theorem A.0.5 hold over more general categories than $\mathrm{FI}|_{\mathcal{W}}$, including for categories of FI-type and certain quasi-Gröbner categories introduced in [195]. We remark that these two types of categories only include representations of finite groups, and do not include the graphon category in Example A.0.3(c), whose properties would be interesting to study in future work.

Definition A.0.6 (Property (TFG))**.** *We say that a category C satisfies* property (TFG) *if:*

(a) tensor products of free C-modules are free;

(b) if $\mathscr{V}_1$ and $\mathscr{V}_2$ are C-modules generated in degrees d_1 and d_2, respectively, then $\mathscr{V}_1 \otimes \mathscr{V}_2$ is generated in degree $d_1 + d_2$.

An example of a category not satisfying (TFG) is given in [195, Rmk. 7.4.3]. We can use property (TFG) to obtain a calculus for presentation degrees from which Theorem 2.2.11 may be deduced.

Proposition A.0.7. *Suppose C is a category satisfying (TFG). If $\mathscr{V}, \mathscr{U}$ are C-modules which are generated in degrees d_V, d_U and presented in degrees k_V, k_U, respectively, then $\mathscr{V} \otimes \mathscr{U}$ is presented in degree* $\max\{d_V + k_U, d_U + k_V\}$.

Proof. Suppose $\mathscr{F}_V, \mathscr{F}_U$ are free C-modules generated in degrees d_V, d_U, respectively, and $\mathscr{F}_V \to \mathscr{V}$ and $\mathscr{F}_U \to \mathscr{U}$ are surjective morphisms whose kernels $\mathscr{K}_V$ and $\mathscr{K}_U$ are generated in degrees r_V, r_U, respectively. Then $\mathscr{F}_V \otimes \mathscr{F}_U$ is a free C-module generated in degree $d_V + d_U$ by (TFG), and the morphism $\mathscr{F}_V \otimes \mathscr{F}_U \to \mathscr{V} \otimes \mathscr{U}$

is surjective with kernel $\mathscr{K}_V \otimes \mathscr{F}_U + \mathscr{F}_V \otimes \mathscr{K}_U$. Since $\mathscr{K}_V \otimes \mathscr{F}_U$ is generated in degree $r_V + d_U$ and similarly for $\mathscr{F}_U \otimes \mathscr{K}_V$, their sum is generated in degree $\max\{r_V + d_U, d_V + r_U\}$. $\square$

Our next goal is to understand presentation degrees for images, and in particular, for Schur functors. We begin with a number of elementary lemmas.

Lemma A.0.8. *Let $\mathscr{V}$ and $\mathscr{U}$ be C-modules. If $\mathscr{V}, \mathscr{U}$ are generated in degrees d_V, d_U and presented in degrees k_V, k_U, respectively, then $\mathscr{V} \oplus \mathscr{U}$ is generated in degree* $\max\{d_V, d_U\}$ *and presented in degree* $\max\{k_V, k_U\}$.

Proof. The claim about the generation degree is immediate from its definition. Suppose that $\mathscr{F}_V \to \mathscr{V}$ and $\mathscr{F}_U \to \mathscr{U}$ are surjective morphisms with $\mathscr{F}_V, \mathscr{F}_U$ being free C-modules generated in degrees d_V, d_U with kernels $\mathscr{K}_V, \mathscr{K}_U$ generated in degrees k_V, k_U, respectively. Then $\mathscr{F}_V \oplus \mathscr{F}_U$ is free (by definition) and generated in degree $\max\{d_V, d_U\}$, and surjects onto $\mathscr{V} \oplus \mathscr{U}$ with kernel $\mathscr{K}_V \oplus \mathscr{K}_U$ which is generated in degree $\max\{k_V, k_U\}$. $\square$

Lemma A.0.9. *Let $\mathscr{V} = \{\mathbb{V}_n\}$ and $\mathscr{U} = \{\mathbb{U}_n\}$ be two C-modules, let $\mathscr{A} = \{A_n\}\colon \mathscr{V} \to \mathscr{U}$ be a surjective morphism, and let $\mathscr{W} = \{\mathbb{W}_n \subseteq \mathbb{U}_n\}$ be a C-submodule of $\mathscr{U}$. If* $\ker\mathscr{A}$ *is generated in degree d and $\mathscr{W}$ is generated in degree d_W, then $\mathscr{A}^{-1}(\mathscr{W}) = \{A_n^{-1}(\mathbb{W}_n)\}$ is a C-module generated in degree* $\max\{d, d_W\}$.

Proof. Suppose $\{y_i\} \subseteq \mathbb{W}_{d_W}$ are generators for $\mathscr{W}$. Since A_{d_W} is surjective, we can find $\{x_i\} \subseteq \mathbb{V}_{d_W}$ such that $A_{d_W} x_i = y_i$ for all i. Also let $\{z_j\} \subseteq \ker A_d$ be generators for $\ker \mathscr{A}$. We claim that $\{x_i\} \cup \{z_j\} \subseteq A^{-1}_{\max\{d,d_W\}}(\mathbb{W}_{\max\{d,d_W\}})$ generate $\mathscr{A}^{-1}(\mathscr{W})$. Indeed, if $n \geq \max\{d, d_W\}$ and $x \in A_n^{-1}(\mathbb{W}_n)$, then we can write $A_n x = \sum_i \alpha_i g_i y_i$ for some $\alpha_i \in \mathbb{R}$ and $g_i \in \mathsf{G}_n$, in which case $x - \sum_i \alpha_i g_i x_i \in \ker A_n$ and thus, we have $x = \sum_i \alpha_i g_i x_i + \sum_j \beta_j g_j z_j$ for some $\beta_j \in \mathbb{R}$ and $g_j \in \mathsf{G}_n$. $\square$

Lemma A.0.10. *Suppose $\mathscr{V} = \{\mathbb{V}_n\}$ and $\{\mathbb{U}_n\}$ are two C-modules, and $\mathscr{A} = \{A_n\}\colon \mathscr{V} \to \mathscr{U}$ is a morphism. If $\mathscr{V}$ is generated in degree d, then* $\mathrm{im}\mathscr{A}$ *is generated in degree d. If, moreover, $\mathscr{A}^* = \{A_n^*\}$ is a morphism, then* $\ker\mathscr{A}$ *is also generated in degree d.*

Proof. The first claim follows from $A_n(\mathbb{V}_n) = A_n(\mathbb{R}[\mathsf{G}_n]\mathbb{V}_d) = \mathbb{R}[\mathsf{G}_n]A_n(\mathbb{V}_d) = \mathbb{R}[\mathsf{G}_n]A_d(\mathbb{V}_d)$, where we used the equivariance of A_n and the fact that $A_n|_{\mathbb{V}_d} = A_d$. For the second claim, note that if $\mathscr{A}^*$ is a morphism, then $\{\mathrm{im}\, A_n^* = (\ker A_n)^\perp\}$ is a

C-submodule of $\mathscr{V}$. Therefore, $\{\mathcal{P}_{\ker A_n}\}\colon \mathscr{V} \to \mathscr{V}$ is a morphism, and its image is precisely $\ker \mathscr{A}$. $\square$

Proposition A.0.11. *Suppose $\mathscr{V} = \{\mathbb{V}_n\}, \mathscr{U} = \{\mathbb{U}_n\}$ are two C-modules and both $\mathscr{A} = \{A_n\colon \mathbb{V}_n \to \mathbb{U}_n\}$ and $\{A_n^*\colon \mathbb{U}_n \to \mathbb{V}_n\}$ are morphisms. If $\mathscr{V}$ is generated in degree d and presented in degree k, then* $\mathrm{im}\mathscr{A} = \{A_n(\mathbb{V}_n)\}$ *is generated in degree d and presented in degree k.*

Proof. Let $\mathscr{F} = \{\mathbb{F}_n\}$ be a free C-module generated in degree d and let $\mathscr{B} = \{B_n\}\colon \mathscr{F} \to \mathscr{V}$ be a surjective morphism whose kernel $\mathscr{K} = \{\mathbb{K}_n\}$ is generated in degree k. The composition $\mathscr{F} \xrightarrow{\mathscr{B}} \mathscr{V} \xrightarrow{\mathscr{A}} \mathrm{im}\mathscr{A}$ is a surjective morphism from the free C-module $\mathscr{F}$ whose kernel is $\mathscr{B}^{-1}(\ker\mathscr{A})$ and is generated in degree $\max\{d,k\} = k$ by Lemmas A.0.9 and A.0.10. $\square$

Corollary A.0.12. *Suppose $\mathscr{C}$ satisfies property (TFG). If $\mathscr{V}$ is a C-module generated in degree d and presented in degree k, and λ is a partition, then $\mathbb{S}^\lambda\mathscr{V}$ is generated in degree $d|\lambda|$ and presented in degree $k + d(|\lambda| - 1)$.*

Schur functors generalize symmetric and alternating algebras, see [81, §6.1]. Their generation degree for C = FI was bounded using a similar approach in [46, Prop. 3.4.3].

Proof. By Proposition A.0.7, the C-module $\mathscr{V}^{\otimes|\lambda|}$ is generated in degree $d|\lambda|$ and presented in degree $d(|\lambda|-1)+k$. Let $\mathsf{S}_{|\lambda|}$ act on each $\mathbb{V}_n^{\otimes|\lambda|}$ by permuting its factors $\sigma\cdot(v_1\otimes\cdots\otimes v_{|\lambda|}) = v_{\sigma(1)}\otimes\cdots\otimes v_{\sigma(|\lambda|)}$, which is an orthogonal action commuting with the embeddings $\mathbb{V}_n^{\otimes|\lambda|} \subseteq \mathbb{V}_{n+1}^{\otimes|\lambda|}$. In this way, any element $c_\lambda \in \mathbb{R}[\mathsf{S}_{|\lambda|}]$ defines a morphism $c_\lambda\colon \mathscr{V}^{\otimes|\lambda|} \to \mathscr{V}^{\otimes|\lambda|}$ such that $c_\lambda^* \in \mathbb{R}[\mathsf{S}_{|\lambda|}]$ is also a morphism. If $c_\lambda \in \mathbb{R}[\mathsf{S}_{|\lambda|}]$ is the Young symmetrizer corresponding to partition λ, then $\mathrm{im}\, c_\lambda = \mathbb{S}^\lambda\mathscr{V}$. The result follows from Proposition A.0.11. $\square$

We conclude this appendix by summarizing our calculus for generation and presentation degrees. Instantiating the following theorem with $C = \mathrm{FI}|_{\mathcal{W}}$ yields Theorem 2.2.11.

Theorem A.0.13 (Calculus for generation and presentation degrees)**.** *Let $\mathscr{V} = \{\mathbb{V}_n\}, \mathscr{U} = \{\mathbb{U}_n\}$ be C-modules generated in degrees d_V, d_U and presented in degrees k_V, k_U, respectively.*

(Sums) *$\mathscr{V} \oplus \mathscr{U}$ is generated in degree* $\max\{d_V, d_U\}$ *and presented in degree* $\max\{k_V, k_U\}$.

(Images and kernels) *If $\mathscr{A} : \mathscr{V} \to \mathscr{U}$ and $\mathscr{A}^*$ are morphisms, then* $\operatorname{im}\mathscr{A}$ *and* $\ker \mathscr{A}$ *are generated in degree d_V and presented in degree k_V.*

Suppose C satisfies (TFG). Then

(Tensors) *$\mathscr{V} \otimes \mathscr{U}$ is generated in degree $d_V + d_U$ and presented in degree* $\max\{k_V + d_U, k_U + d_V\}$.

(Schur functors) *$\mathbb{S}^\lambda \mathscr{V}$ is generated in degree $d_V|\lambda|$ and presented in degree $d_V(|\lambda| - 1) + k_V$ for any partition λ.*

Appendix B

DERIVING THE PRESENTATION DEGREE FROM EXTENDABILITY

In this appendix, we show how the definitions of algebraically free consistent sequences and the presentation degree naturally arise when trying to extend a fixed equivariant linear map to a morphism of sequences, necessary for our algorithm in Section 2.6.

Let $\mathscr{V} = \{\mathbb{V}_n\}$, $\mathscr{U} = \{\mathbb{U}_n\}$ be consistent sequences of $\{\mathsf{G}_n\}$-representations, fix $n_0 \in \mathbb{N}$ and consider a linear map $A_{n_0} \in \mathcal{L}(\mathbb{V}_{n_0}, \mathbb{U}_{n_0})^{\mathsf{G}_{n_0}}$. When can we extend A_{n_0} to a morphism of sequences $\{A_n\}$? We seek conditions on A_{n_0} which are easy to enforce computationally, so that they can be used to parametrize and search over compatible sequences of convex sets computationally. The following proposition gives an equivalent characterization for the existence of such an extension.

Proposition B.0.1. *Let $\mathscr{V} = \{\mathbb{V}_n\}, \mathscr{U} = \{\mathbb{U}_n\}$ be consistent sequences such that $\mathscr{V}$ is generated in degree d, and fix $A_{n_0} \in \mathcal{L}(\mathbb{V}_{n_0}, \mathbb{U}_{n_0})^{\mathsf{G}_{n_0}}$ for $n_0 \geq d$.*

(a) There exists $\{A_n \in \mathcal{L}(\mathbb{V}_n, \mathbb{U}_n)^{\mathsf{G}_n}\}_{n<n_0}$ satisfying $A_{n+1}|_{\mathbb{V}_n} = A_n$ for all $n < n_0$ if and only if $A_{n_0}(V_j) \subseteq U_j$ for $j \leq d$.

(b) There exists $\{A_n \in \mathcal{L}(\mathbb{V}_n, \mathbb{U}_n)^{\mathsf{G}_n}\}_{n>n_0}$ satisfying $A_{n+1}|_{\mathbb{V}_n} = A_n$ for all $n \geq n_0$ if and only if the following implication holds

$$\sum_i g_i x_i = 0 \implies \sum_i g_i A_{n_0} x_i = 0, \qquad \text{for all } g_i \in \mathsf{G}_n,\ x_i \in \mathbb{V}_d,\ n \in \mathbb{N}. \tag{B.1}$$

If an extension $\{A_n\}$ of A_{n_0} exists, then it is unique.

Proof. (a) If such $\{A_n\}_{n<n_0}$ exists, then it is uniquely given in terms of A_{n_0} by $A_n = A_{n_0}|_{\mathbb{V}_n}$. Therefore, we have $A_{n_0}(V_j) = A_j(V_j) \subseteq U_j$ for all $j \leq d$. Conversely, suppose $A_{n_0}(V_j) \subseteq U_j$ for $j \leq d$. We claim that $A_{n_0}(\mathbb{V}_n) \subseteq \mathbb{U}_n$ for all $d \leq n \leq n_0$ as well. Indeed, because $\mathscr{V}$ is generated in degree d, we have $A_{n_0}(\mathbb{V}_n) = A_{n_0}(\mathbb{R}[\mathsf{G}_n]\mathbb{V}_d) = \mathbb{R}[\mathsf{G}_n]A_{n_0}(\mathbb{V}_d) \subseteq \mathbb{R}[\mathsf{G}_n]\mathbb{U}_d \subseteq \mathbb{U}_n$ for such $n \geq d$. Defining $A_n = A_{n_0}|_{\mathbb{V}_n}$ for each $n < n_0$ yields the desired extension to lower dimensions.

(b) If such $\{A_n\}_{n>n_0}$ exists, it is unique and is explicitly given in terms of A_{n_0} as follows. For any $n > n_0$ and $x \in \mathbb{V}_n$, we can write $x = \sum_i g_i x_i$ for some $g_i \in \mathsf{G}_n$ and $x_i \in \mathbb{V}_d \subseteq \mathbb{V}_{n_0}$ by definition of the generation degree. Because $A_n \colon \mathbb{V}_n \to \mathbb{U}_n$ is G_n-equivariant and satisfies $A_n|_{\mathbb{V}_{n_0}} = A_{n_0}$, we have

$$A_n x = A_n \left(\sum_i g_i x_i\right) = \sum_i g_i (A_{n_0} x_i), \tag{B.2}$$

which expresses A_n in terms of A_{n_0}. The expression (B.2) shows that (B.1) is satisfied. Conversely, suppose that (B.1) is satisfied. For each $n > n_0$ define $A_n \colon \mathbb{V}_n \to \mathbb{U}_n$ as follows. For any $x \in \mathbb{V}_n$, write $x = \sum_i g_i x_i$ for some $g_i \in \mathsf{G}_n$ and $x_i \in \mathbb{V}_d$, which is possible because $\mathscr{V}$ is generated in degree d, and set $A_n x$ to the right-hand side of (B.2). This is well-defined because if $x = \sum_i g_i x_i = \sum_j g'_j x'_j$ for $g_i, g'_j \in \mathsf{G}_n$ and $x_i, x'_j \in \mathbb{V}_d$ then $\sum_i g_i A_{n_0} x_i = \sum_j g'_j A_{n_0} x'_j$ by (B.1). Moreover, A_n is linear, G_n-equivariant, and extends A_{n_0} by construction, so $\{A_n\}_{n>n_0}$ is the desired extension of A_{n_0}. □

The conditions on A_{n_0} in Proposition B.0.1(a) are easy to impose computationally, and we do so in Section 2.6. In contrast, while condition (B.1) in Proposition B.0.1(b) fully characterizes extendability of A_{n_0} to higher dimensions, it is unclear how to impose it computationally. We therefore proceed to study it further. In algebraic terms, elements $x_i \in \mathbb{V}_d$ are called the *generators* of $\mathscr{V}$, and expressions of the form $\sum_i g_i x_i = 0$ with $g_i \in \mathsf{G}_n$ are called *relations* between those generators over the group G_n. Proposition B.0.1(b) shows that A_{n_0} extends to a morphism iff any relation satisfied by the generators of $\mathscr{V}$ is also satisfied by their images under A_{n_0}. We therefore need to understand the relations among the generators in $\mathbb{V}_d$.

We study these relations in two stages. First, we identify two simple types of relations that are satisfied by the images of the generators under A_{n_0} for appropriate $\mathscr{V}, \mathscr{U}$. We then define algebraically free consistent sequences whose generators satisfy only these two types of relations. Second, we express an arbitrary consistent sequence as the quotient of an algebraically free one. The kernel of this quotient morphism is a consistent sequence that encodes all additional relations. To capture the degree starting from which both the generators and the relations between them stabilize, we define the presentation degree of a consistent sequence as the maximum of the generation degree of the sequence itself and that of the above kernel, see Definition 2.2.7. The presentation degree plays a prominent role in our structural results in Section 2.3.

We begin carrying out the first stage of the above program and identify two simple types of relations. The first source for relations between generators in $\mathbb{V}_d$ arises from the action of G_d on $\mathbb{V}_d$ itself, without involving any higher dimensions. Indeed, if $\sum_i g_i x_i = 0$ for $g_i \in \mathsf{G}_d$ and $x_i \in \mathbb{V}_d$ is such a relation then $\sum_i g_i A_{n_0} x_i = 0$ so the image $A_{n_0} x_i$ satisfies the same relation. Moreover, the above relations in $\mathbb{V}_d$ give rise to relations in $\mathbb{V}_n$ for $n > d$ since if $\sum_i g_i x_i = 0$ as above then $\sum_i g g_i x_i = 0$ in $\mathbb{V}_n$ for any $g \in \mathsf{G}_n$. Once again, these higher-dimensional relations are always satisfied by $A_{n_0} x_i$. A second source for such relations arises from subgroups of G_n acting trivially on $\mathbb{V}_d$, which are precisely the centralizing subgroups of Definition 2.2.4. Centralizing subgroups yield a second source for relations, namely, the relations $(h - \mathrm{id})x = 0$ for all $x \in \mathbb{V}_d$ and $h \in \mathsf{H}_{n,d}$. Thus, if A_{n_0} extends to a morphism then $A_{n_0}(\mathbb{V}_d) \subseteq \bigcap_{n \geq d} \mathbb{U}_n^{\mathsf{H}_{n,d}}$. Rather than attempt to enforce these constraints computationally, we make a standard simplifying assumption from the representation stability literature. Specifically, we assume that $\mathbb{U}_d$ is fixed by $\mathsf{H}_{n,d}$ (or a subgroup of it, see below) for all $n \geq d$, in which case there is a simple sufficient condition for the above constraints that we can impose computationally. This is precisely the assumption that $\mathscr{U}$ is a $\mathscr{V}$-module as in Definition 2.2.5. This terminology comes from a categorical approach to representation stability, see Appendix A. If $\mathscr{U}$ is a $\mathscr{V}$-module, then imposing $A_{n_0}(\mathbb{V}_d) \subseteq \mathbb{U}_d$ is sufficient to guarantee $A_{n_0}(\mathbb{V}_d) \subseteq \bigcap_{n \geq d} \mathbb{U}_n^{\mathsf{H}_{n,d}}$ since $\mathbb{U}_d$ is contained in the right-hand side of this inclusion. Imposing this sufficient condition can done computationally, as we do in Section 2.6. This concludes the first stage.

To go beyond the above two simple types of relations satisfied by the generators, we define algebraically free consistent sequences (Definition 2.2.6), whose generators do not satisfy any additional types of relations. We then write any consistent sequence as the image under a morphism of sequences of an algebraically free one. The kernel of this morphism precisely captures the relations satisfied by the generators. To that end, note that any finitely-generated consistent sequence is the image under a morphism of sequences of an algebraically free sequence.

Proposition B.0.2. *Let $\mathscr{V}$ be a consistent sequence of $\{\mathsf{G}_n\}$-representations and let $\mathscr{U} = \{\mathbb{U}_n\}$ be a $\mathscr{V}$-module. Then $\mathscr{U}$ is generated in degree d if and only if there exists an algebraically free $\mathscr{V}$-module $\mathscr{F}$ generated in degree d and a surjective morphism of sequences $\mathscr{F} \to \mathscr{U}$.*

Proof. If $\mathscr{F} = \{\mathbb{F}_n\}$ is a $\mathscr{V}$-module and $\{A_n \colon \mathbb{F}_n \to \mathbb{U}_n\}$ is a surjective morphism, then for any $n \geq d$ we have $\mathbb{U}_n = A_n(\mathbb{F}_n) = A_n(\mathbb{R}[\mathsf{G}_n]\mathbb{F}_d) = \mathbb{R}[\mathsf{G}_n] A_n(\mathbb{F}_d) =$

$\mathbb{R}[\mathsf{G}_n]A_d(\mathbb{F}_d) = \mathbb{R}[\mathsf{G}_n]\mathbb{U}_d$, where we used the fact that $\mathscr{F}$ is generated in degree d; the equivariance of A_n; the fact that $A_n|_{\mathbb{F}_d} = A_d$ since $\{A_n\}$ is a morphism; and the surjectivity of A_d. This shows $\mathscr{U}$ is generated in degree d.

Conversely, if $\mathscr{U}$ is generated in degree d, define the algebraically free $\mathscr{V}$-module $\mathscr{F} = \bigoplus_{i=1}^d \mathrm{Ind}_{\mathsf{G}_i}(U_i)$ and consider the morphism $\mathscr{F} \to \mathscr{U}$ defined by $g \otimes u \mapsto g \cdot u$ for each $g \in \mathsf{G}_n, u \in U_i$, and $i \in [d]$ (see Section 2.1). The image of this morphism in $\mathbb{U}_n$ is precisely $\sum_{i=1}^{\min\{d,n\}} \mathbb{R}[\mathsf{G}_n]U_i$, hence it is surjective for all n. Finally, $\mathrm{Ind}_{\mathsf{G}_i}(U_i)$ is generated in degree i, so $\mathscr{F}$ is generated in degree d. □

The kernel of the morphism in Proposition B.0.2 precisely encodes all the additional relations beyond the two simple types above satisfied by the generators of $\mathscr{V}$. The generation degree of this kernel then captures the point at which relations stabilize. We therefore define the presentation degree (Definition 2.2.7) as the maximum of the generation degree of $\mathscr{V}$ and that of this kernel, which captures stabilization of the generators as well as of the relations between them.

The presentation degree allows us to ensure condition (B.1) is satisfied and hence to extend a fixed linear map to a morphism of sequences in Theorem 2.3.4, thus answering the question posed in the beginning of this section. Indeed, comparing Theorem 2.3.4 with Proposition B.0.1, we see that condition (B.1) is satisfied by any fixed equivariant map in dimension n_0 exceeding the presentation degree. Thus, the presentation degree appears in our extendability result for convex sets (Theorem 2.3.5), and in our algorithm for computationally parametrizing such sets (Algorithm 1).

Appendix C

SCHUR–HORN ORBITOPES IN AF ALGEBRAS

The permutahedra and Schur–Horn orbitopes studied in Section 2.4 are special cases of a more general construction arising in the field of operator algebras, and which is naturally analyzed using our framework. In this appendix, we present this unified view. Let $\mathbb{A}_n \subseteq \mathbb{M}_n := \mathbb{C}^{m2^n \times m2^n}$ be a unital $\mathbb{C}$-subalgebra with the faithful tracial state $\tau_n(x) = \mathrm{Tr}(x)/m2^n$ which induces the inner product $\langle x, y\rangle_n = \tau_n(x^* y)$, and the max-singular value operator norm $\|\cdot\|_{(n)}$. Let $\varphi_n \colon \mathbb{M}_n \hookrightarrow \mathbb{M}_{n+1}$ be the unital algebra embedding $\varphi_n(x) = x \otimes I_2$, and note that $\|x \otimes I_2\|_{(n+1)} = \|x\|_{(n)}$ and $\tau_{n+1} \circ \varphi_n = \tau_n$, so φ_n is an isometry with respect to the above inner product. Thus, both the operator norm $\|\cdot\|$ and the normalized trace τ do not depend on n and extend to the limit $\mathbb{A}_\infty = \bigcup_n \mathbb{A}_n$, and we omit their subscripts above. Let $\overline{\mathbb{A}_\infty}$ be the closure of $\mathbb{A}_\infty$ with respect to the operator norm, which is now a C^* algebra called an approximately finite-dimensional (AF) algebra [55, Chap. 3]. Note that τ extends continuously to $\overline{\mathbb{A}_\infty}$ since $|\tau(x)\| \leq \|x\|$ for all $x \in \mathbb{A}_\infty$, and hence τ defines a tracial state on $\overline{\mathbb{A}_\infty}$.

The spectral theorem for C^* algebras gives an (isometric) isomorphism of algebras between continuous functions on the spectrum $\sigma(x) = \{\lambda \in \mathbb{C} : x - \lambda I \text{ not invertible}\}$ of any self-adjoint $x \in \overline{\mathbb{A}_\infty}$ and the closed subalgebra it generates [191, Thm. 11.19]. This allows us to apply continuous functions $f \colon \sigma(x) \to \mathbb{C}$ to x itself to obtain $f(x) \in \overline{\mathbb{A}_\infty}$. Using this functional calculus, for each self-adjoint $x \in \overline{\mathbb{A}_\infty}$ we get the positive unital linear functional on such functions sending $f \mapsto \tau(f(x))$. By the Riesz representation theorem [192, Thm. 6.19], there is a unique probability measure μ_x^τ on $\sigma(x)$ satisfying $\mu_x^\tau(f) = \tau(f(x))$ for all continuous $f \colon \sigma(x) \to \mathbb{C}$, which is called the spectral measure of x with respect to τ. Finally, if $\sigma(x) = \{\lambda_1, \ldots, \lambda_q\}$ is finite then we have a spectral decomposition $x = \sum_{i=1}^q \lambda_i p_i$ where $\{p_i\}$ are orthogonal projectors satisfying $p_i p_j = \delta_{i,j} p_i$, $\sum_i p_i = 1$, and $\mu_x^\tau = \sum_{i=1}^q \tau(p_i)\delta_{\lambda_i}$.

Let $\mathbb{V}_n = \mathbb{A}_n \cap \mathbb{M}_n^{\mathrm{s.a.}}$ be the (finite-dimensional real) vector space of self-adjoint elements in $\mathbb{A}_n$, and note that $\varphi_n(\mathbb{V}_n) \subseteq \mathbb{V}_{n+1}$ and that $\overline{\mathbb{V}_\infty}$ is the collection of self-adjoints in $\overline{\mathbb{A}_\infty}$. Let $\mathsf{G}_n \subseteq \mathrm{U}(m2^n)$ be a group of unitaries acting on $\mathbb{V}_n$ by conjugation such that $\mathsf{G}_n \subseteq \mathsf{G}_{n+1}$ under the embedding $g \mapsto g \otimes I_2$, so that $\{\mathbb{V}_n\}$ is a consistent sequence of $\{\mathsf{G}_n\}$-representations.

Lemma C.0.1. *Let $\mathcal{P}_n\colon \mathbb{V}_\infty \to \mathbb{V}_n$ be orthogonal projections. Then $\|\mathcal{P}_n x\| \leq \|x\|$ and $\tau \circ \mathcal{P}_n = \tau$ for all n.*

Proof. Note that $\varphi_n^*\colon \mathbb{V}_{n+1} \to \mathbb{V}_n$ satisfies

$$\varphi_n^*(x) = \frac{1}{2}\sum_{i=1}^{2}(I \otimes e_i)^\top x(I \otimes e_i), \qquad \text{hence } \|\varphi_n^* x\| \leq \|x\|, \tag{C.1}$$

for all $x \in \mathbb{V}_{n+1}$. Because the orthogonal projection $\mathcal{P}_n\colon \mathbb{V}_\infty \to \mathbb{V}_n$ satisfies $\mathcal{P}_n|_{\mathbb{V}_N} = \varphi_n^* \circ \ldots \circ \varphi_{N-1}^*$ for each $N > n$, we conclude that $\|\mathcal{P}_n x\| \leq \|x\|$ for all $x \in \mathbb{V}_\infty$. Also, for any $x \in \mathbb{V}_\infty$ we have $\tau(\mathcal{P}_n x) = \langle \mathcal{P}_n x, 1\rangle = \langle x, \mathcal{P}_n 1\rangle = \langle x, 1\rangle = \tau(x)$, hence the second claim follows. ◻

In particular, the norm $\|\cdot\|$ satisfies the condition of Definition 2.2.21. Let $\lambda \in \mathbb{R}^q$ and $\widetilde{\lambda} = [\lambda_1 \mathbb{1}_{m_i}^\top, \ldots, \lambda_q \mathbb{1}_{m_q}^\top]^\top \in \mathbb{R}^m$ has entry λ_i repeated m_i times (so $m = \sum_i m_i$). Then $x_1 = \operatorname{diag}(\widetilde{\lambda})$ has spectral measure $\mu_{x_1}^\tau = \sum_{i=1}^q \frac{m_i}{m}\delta_{\lambda_i}$. The Shur-Horn orbitopes associated to x_1 is then

$$\mathrm{SH}(x_1)_n = \operatorname{conv}\{x \in \mathbb{V}_n : \mu_x^\tau = \mu_{x_1}^\tau\} = \operatorname{conv}\left\{\sum_{i=1}^q \lambda_i p_i : p_i \in \mathbb{V}_n,\ p_i p_j = \delta_{i,j} p_i,\ \sum_i p_i = 1,\ \tau(p_i) = \frac{m_i}{m}\right\}$$

To obtain conic descriptions for the above orbitopes, let $\mathcal{K}_n = \{xx^* : x \in \mathbb{A}_n\} \subseteq \mathbb{V}_n$ be the cone of positive-semidefinite Hermitian matrices in $\mathbb{A}_n$, so that $\mathcal{K}_n \subseteq K_{n+1}$ and $\overline{\mathcal{K}_\infty} = \{xx^* : x \in \overline{\mathbb{A}_\infty}\}$. We also have $\mathcal{P}_n(K_{n+1}) = \mathcal{K}_n$ by (C.1), hence $\mathcal{P}_n(\overline{\mathcal{K}_\infty}) = \mathcal{K}_n$. Therefore, the finite-dimensional description of Schur–Horn orbitopes [35, Eq. (19)] reads

$$\mathrm{SH}(x_1)_n = \left\{\sum_{i=1}^q \lambda_i p_i : p_i \in \mathcal{K}_n,\ \sum_i p_i = 1,\ \tau(p_i) = m_i/m\right\}. \tag{C.2}$$

Our Theorem 2.3.6 now yields the following infinite-dimensional extension of these descriptions.

Proposition C.0.2. *Let $\lambda_1, \ldots, \lambda_q$ are distinct real numbers, and let $m_1, \ldots, m_q \in \mathbb{N}$ with sum $m = \sum_i m_i$. Then $\overline{\mathrm{SH}(x_1)_\infty}$ with $x_1 = \operatorname{diag}(\widetilde{\lambda})$ as above is*

$$\overline{\operatorname{conv}}\left\{x \in \overline{\mathbb{V}_\infty} : \mu_x^\tau = \sum_{i=1}^q \frac{m_i}{m}\delta_{\lambda_i}\right\} = \overline{\left\{\sum_{i=1}^q \lambda_i p_i : p_i \in \overline{\mathcal{K}_\infty},\ \sum_i p_i = 1,\ \tau(p_i) = m_i/m\right\}}. \tag{C.3}$$

Proof. First, the left-hand side of (C.3) is just $\overline{\mathrm{SH}(x_1)_\infty}$. Indeed, if $x_n \in \overline{\mathbb{V}_\infty}$ is a sequence converging to x with $\mu_{x_n}^\tau = \mu_{x_1}^\tau$ for all n, then for any continuous

$f\colon \mathbb{R} \to \mathbb{C}$ we have $\mu_x^\tau(f) = \tau(f(x)) = \lim_n \tau(f(x_n)) = \mu_{x_1}^\tau(f)$ so $\mu_x^\tau = \mu_{x_1}^\tau$. Conversely, if $\mu_x^\tau = \sum_i \frac{m_i}{m}\delta_{\lambda_i}$ then x admits a decomposition $x = \sum_i \lambda_i p_i$ with $\tau(p_i) = m_i/m$. Define $x_n = \mathcal{P}_n x = \sum_{i=1}^q \lambda_i(\mathcal{P}_n p_i)$ and note that $\mathcal{P}_n p_i \in \mathcal{K}_n$ for all i, that $\sum_i \mathcal{P}_n p_i = \mathcal{P}_n 1 = 1$, and that $\tau(\mathcal{P}_n p_i) = \tau(p_i) = m_i/m$. Thus, $x_n \in \mathrm{SH}(x_1)_n$ for all n and $x_n \to x$ by Lemma 2.2.22, hence $x \in \overline{\mathrm{SH}(x_1)_\infty}$.

Second, we appeal to Theorem 2.3.6 to obtain equality in (C.3). Note that the conic description (C.2) is of the form (ConicSeq) with $\mathscr{V} = \{\mathbb{V}_n\}, \mathscr{W} = \mathscr{V}^{\oplus q}, \mathscr{U} = \mathscr{W} \oplus \mathscr{V}^{\oplus 2} \oplus \mathbb{R}^q$, and

$$A_n x = (0, -x, 0, 0), \quad B_n(p_1, \dots, p_q) = \left(p_1, \dots, p_q, \sum\nolimits_i \lambda_i p_i, \sum\nolimits_i p_i, \tau(p_1), \dots, \tau(p_q)\right),$$

and $u_n = (0, 0, -I, -m_1/m, \dots, -m_q/m)$. Observing that $\{A_n\}, \{A_n^*\}, \{B_n\}, \{B_n^*\}$ are all morphisms, that $u_n = u_{n+1}$ (under our embeddings), the descriptions in (C.2) are dimension-free and satisfy the hypotheses of Proposition 2.3.2(a). Furthermore, putting the norm $\|(p_1, \dots, p_q)\| = \max_i \|p_i\|$ on W_∞ and $\|(p_1, \dots, p_q, y_1, y_2, v)\| \le \max\{\|p_i\|, \|y_j\|, \|v\|_\infty\}$ on $\mathbb{U}_\infty$, we have $\overline{W_\infty} = \overline{\mathbb{V}_\infty}^{\oplus q}$ and $\overline{\mathbb{U}_\infty} = \overline{W_\infty} \oplus \overline{\mathbb{V}_\infty}^{\oplus 2} \oplus \mathbb{R}^q$. Moreover, $\|A_n\|_{\mathrm{op}} = 1$, $\|B_n\|_{\mathrm{op}} \le \max\{\sum_i |\lambda_i|, q\}$ for all n. Thus, both A_n and B_n extend to the continuous limit and Theorem 2.3.6 indeed applies and yields (C.3). □

The Schur–Horn orbitopes $\mathrm{SH}(x_1)_n$ are precisely the Schur–Horn orbitopes $\mathrm{SH}(\lambda)_n$ in (2.15) from Section 2.4 if $\mathbb{A}_n = \mathbb{R}^{m2^n \times m2^n}$ with $\mathsf{G}_n = \mathrm{O}(m2^n)$. They also reduce to the permutahedra $\mathrm{SH}(x_1)_n = \mathrm{Perm}(\lambda)_n$ as in (2.12) from Section 2.4 if $\mathbb{A}_n$ is the algebra of diagonal matrices in $\mathbb{M}_n$ with $\mathsf{G}_n = \mathsf{S}_{m2^n}$.

Proposition C.0.2 yields a generalization of the Schur–Horn theorem to AF algebras. Indeed, let $\mathbb{D}_n = \{x \in \mathbb{V}_n : x \text{ diagonal}\}$ which is a unital subalgebra of $\mathbb{A}_n$ embedded in $\mathbb{D}_{n+1}$ under φ_n. For $x_1 \in \mathbb{D}_1$, define

$$\mathrm{Perm}(x_1)_n = \mathrm{SH}(x_1)_n \cap \mathbb{D}_n.$$

The sequence of linear maps $(\mathrm{diag}_n\colon \mathbb{V}_n \to \mathbb{D}_n)_n$ extracting the diagonal of a Hermitian matrix in $\mathbb{V}_n$ extend to a bounded linear map $\mathrm{diag}\colon \overline{\mathbb{V}_\infty} \to \overline{\mathbb{D}_\infty}$ since $\mathrm{diag}_{n+1} \circ \varphi_n = \varphi_n \circ \mathrm{diag}_n$ and $\|\mathrm{diag}_n\|_{\mathrm{op}} = 1$. The finite-dimensional Schur–Horn theorem states that $\mathrm{diag}_n(\mathrm{SH}(x_1)_n) = \mathrm{Perm}(x_1)_n$. Using Proposition C.0.2, we obtain the following infinite-dimensional extension.

Proposition C.0.3. *Let $x_1 \in \mathbb{D}_1$ and let $\mathrm{diag}\colon \overline{\mathbb{V}_\infty} \to \overline{\mathbb{D}_\infty}$ be the bounded linear map extending the diagonal maps. Then $\mathrm{diag}(\overline{\mathrm{SH}(x_1)_\infty}) = \overline{\mathrm{Perm}(x_1)_\infty}$.*

Proof. Since $\overline{\mathrm{Perm}(x_1)_\infty} \subseteq \overline{\mathrm{SH}(x_1)_\infty}$ and diag is a projection onto $\overline{\mathbb{D}_\infty}$, we get $\mathrm{diag}(\overline{\mathrm{SH}(x_1)_\infty}) \supseteq \overline{\mathrm{Perm}(x_1)_\infty}$. Conversely, Proposition C.0.2 and continuity of diag yields

$$\begin{aligned}
\mathrm{diag}(\overline{\mathrm{SH}(x_1)_\infty}) &\subseteq \overline{\left\{\sum\nolimits_{i=1}^{q} \lambda_i \mathrm{diag}(p_i) : p_i \in \overline{\mathcal{K}_\infty},\ \sum\nolimits_i p_i = 1,\ \tau(p_i) = m_i/m\right\}} \\
&= \overline{\left\{\sum\nolimits_{i=1}^{q} \lambda_i p_i : p_i \in \overline{\mathcal{K}_\infty} \cap \overline{\mathbb{D}_\infty},\ \sum\nolimits_i p_i = 1,\ \tau(p_i) = m_i/m\right\}} \\
&= \overline{\mathrm{Perm}(x_1)_\infty},
\end{aligned}$$

where the last equality follows from Proposition C.0.2 applied to $\overline{\mathbb{D}_\infty}$ instead of $\overline{\mathbb{V}_\infty}$. □

The finite-dimensional Schur–Horn theorem yields $\mathrm{Perm}(x_1)_\infty = \mathrm{diag}(\mathrm{SH}(x_1)_\infty)$, and taking closures yields a weaker statement since it involves the closure of the image of diag. Proposition C.0.3 shows that this closure can be removed.

There is a considerable literature on extending the Schur–Horn theorem to infinite-dimensional settings, including the extension in [26] for operators with a finite spectrum, and in [183] for von Neumann algebras. As explained in [183, §1], removing the closure over the image of diag has been a major challenge in these more general settings. Our descriptions of limiting convex sets from Theorem 2.3.6 can be seen as resolving this challenge in our simpler setting of AF algebras.