Superconcentration for Orthogonally Invariant Spin Glasses

by

AJ LaMotta

advised by

Mark Sellke and Laura DeMarco

A thesis presented to

Harvard University

in partial fulfillment of the requirements for the degree of

Bachelor of Arts with Honors
in Mathematics

Introduction

Under the supervision of Mark Sellke in the Harvard Department of Statistics and in collaboration with two colleagues abroad, Tobias Schmidt at the Technical University of Darmstadt and Sean Cotner at the University of Michigan, I have been working on proving one of Mark's conjectures: that the free energy of an orthogonally invariant SK model superconcentrates. We have made a lot of progress, and assuming that our efforts prove successful, we hope to publish a research paper on this topic in the near future. The goal of the present senior thesis is to detail the background necessary for understanding Mark's conjecture and outline the progress we have made towards proving it.

In probability theory, concentration of measure refers to the ubiquitous phenomenon where certain sequences of random variables *concentrate* closely around their mean in high dimensions, such that the order of fluctuations is often much less than the order of expectation. *Super*concentration refers to special cases where these fluctuations are of even lower order than classical techniques in concentration theory predict, so that the sequence *superconcentrates* around its mean. In his 2014 monograph [Cha14], Sourav Chatterjee more precisely defines and describes superconcentration, especially for functions of Gaussian processes, making many connections with "multiple valleys" and "chaos."

Some of Chatterjee's key applications of superconcentration are in spin glass theory. In particular, it is proven that the free energy of the Gaussian SK model superconcentrates [Cha14, Chapter 6]. Spin glasses are statistical models that have been used to describe the interaction of particles in magnetic systems, graphical models in machine learning, biological neural networks, and more. The Sherrington-Kirkpatrick (SK) model is one of the most famous examples of a spin glass, traditionally assumed to have Gaussian disorder. The main idea of this thesis is to consider extensions of [Cha14, Chapter 6] to variants of the SK model whose disorder is not Gaussian but uniform with respect to some geometric structure. In particular, we are interested in proving superconcentration in the case of an *orthogonally invariant* spin glass, that is, an SK model whose geometry is controlled by the (special) orthogonal group. We reduce the conjectured superconcentration to bounding iterated Laplacians of the free energy in Haar expectation.

In Chapter 1, we begin with an introduction to spin glasses in subchapter 1.1, a preview of the Gaussian SK model in subchapter 1.2, and finally a discussion of G-invariant SK models, including orthogonally invariant SK models, in subchapter 1.3. It is in this final subchapter that we formulate our main Conjecture 1.14, as well as a related Conjecture 1.16. We ensure that these conjectures are reasonable via simulation, and we also provide intuition by analogy with the Gaussian case, exemplified by Theorem 1.13.

Chapter 2 is dedicated to superconcentration and a proof of superconcentration for the Gaussian SK model, inspired by [Cha14, Chapter 6]. Subchapter 2.1 starts with an overview of classical concentration of measure, building up to the Gaussian Poincaré inequality in subchapter 2.2 based on Hermite expansions. Then, in subchapter 2.3 we turn to superconcentration, which is an informal discussion accompanied by several examples. We detail Chatterjee's spectral method for proving superconcentration in subchapter 2.4, where the main result is a proof of Corollary 1.9 promised in subchapter 1.2.

Starting in Chapter 3, we begin to shift gears from probability to geometry. This chapter is dedicated to establishing the necessary geometric foundations for our analysis of G-invariant SK models in the following chapter. We begin with a construction of the Haar measure on an arbitrary compact group in subchapter 3.1. Then, we review important notions from the representation theory of groups in subchapter 3.2, culminating in subchapter 3.3 with a proof of the Peter-Weyl theorem. Finally, we end the chapter with an overview of the classification of irreducible representations by highest weight for complex semisimple Lie algebras and connected compact Lie groups. We also state the Weyl character and Weyl dimension formulae.

We close with original results related to G-invariant SK models in Chapter 4. In subchapter 4.1, we discuss a spectral approach towards proving superconcentration for G-invariant SK models, where G is a connected compact Lie group, by analogy with the Gaussian case. In short, Peter-Weyl expansions replace Hermite expansions, while we formulate analogs to the improved Gaussian Poincaré inequality (Theorem 2.30) and Gaussian integration by parts (Lemma 2.15) via the quadratic Casimir operator on G. The main results of this subchapter are equation (4.15) and Lemma 4.4. This is followed by subchapter 4.2, where we discuss the quadratic Casimir operator in more detail. Finally, we conclude in subchapter 4.3 by applying the framework of subchapter 4.1 to the orthogonally invariant case. We make progress towards a proof of Conjecture 1.14, reducing the problem to bounding iterated Laplacians of the free energy in expectation (Theorem 4.12).

Following the main chapters, we include a short appendix of additional proofs absent in the main body of the thesis. Section (A) focuses on technical measure theoretic results, while section (B) is dedicated to filling in the details of certain approximation arguments used in Chapter 2. All other omitted proofs are in section (C). The final section (D) is simply a link to the GitHub repository containing all relevant code.

Acknowledgements

Firstly, I would like to thank Mark Sellke for introducing me to superconcentration and orthogonally invariant spin glasses. I spent much of my senior fall searching for open problems in probability that were accessible enough for an inexperienced undergraduate like myself to make headway on, yet also mathematically rich enough to inspire a senior thesis I could be proud of. Upon hearing Mark explain his conjecture on superconcentration for orthogonally invariant spin glasses, I knew I had found the right topic. Representations of Lie groups in probability theory? Count me in. I greatly appreciate the role Mark has played as my thesis advisor, and I owe much to his guidance and keen intuition.

Next, I would like to thank Harvard Professor of Mathematics Laura DeMarco, my thesis advisor in the mathematics department, for her feedback and guidance on exposition, organization, and general thesis trajectory.

I would also like to acknowledge Tobias Schmidt and Sean Cotner's collaboration as we have worked on resolving Mark's conjecture. Tobias and I in particular have had many great conversations about orthogonally invariant spin glasses, and I would like to thank him for his detailed feedback on my thesis. Recent Harvard alum Kevin Luo, class of '24, deserves recognition as well for his feedback on Chapters 1 and 2. Kevin is currently involved in spin glass research with Mark, and I thank Marvin Li for connecting me with Kevin.

Two courses, plus a third this semester, have been especially impactful on my understanding and presentation of the material in this thesis. On the one hand, many of the probabilistic topics in the first half of this thesis, such as concentration of measure, were covered in a course I took last semester on high-dimensional probability, Stat 216, taught by Subhabrata (Subha) Sen. Subha was incredibly helpful in my search for a thesis topic, as he introduced me to spin glasses and put me into contact with Mark Sellke.

On the other hand, much of the geometric content in the latter half of this thesis traces back to a course on differential geometry, Math 230, that I took under Professor Dan Freed last semester, as well as his course on Lie groups and Lie algebras, Math 222, this semester. I'm a big fan of Dan Freed's approach to geometry, which follows Felix Klein's *Erlangen program* by emphasizing the role of symmetry types and group actions. I would also like to thank Dan Freed for his feedback on my thesis and for advising a Math 91R my junior Spring, where my friend Lev and I were challenged to disavow any reference material for a semester and guide our own learning (a great primer for original research).

Of course, I would like to shout out some of my math "homies" Lev Kruglyak, Jarell Cheong, Marvin Li, Ignasi Segura Vicente, and Leonardo Recanati-Kaplan. Lev has been a close friend/blockmate, a huge help optimizing simulations (including his insistence on speeding things up with Rust), and a wonderful source of feedback on my thesis. Jarell has been great resource for formatting assistance. Marvin provided helpful feedback on Chapters 1 and 2, and we've had many memorable conversations on academia, industry, and life. And I'll never forget countless late nights eating dosa with Ignasi, Leo, and Lev.

Finally, many thanks go out to my family, including my mom, my dad, my grandmas, my brother Luke, and my favorite dog Lenny (yes, my dog), for their constant support.

Contents

1 Spin Glasses and the SK Model

Although the concept of a *spin glass* originates in statistical mechanics, a plethora of applications to neuroscience, machine learning, and mathematics, among other disciplines, has sparked significant independent interest. One of the most influential examples of a spin glass is the *Sherrington-Kirkpatrick* (SK) model, originally introduced in 1975 by physicists David Sherrington and Scott Kirkpatrick [SK75]. In this chapter, we hope to describe the basic setting of a spin glass, examine what is already known about these models, and finally introduce a variant of the SK model that will be the main focus of this thesis.

1.1 The Setup

No discussion of spin glasses is complete without mention of the original context. In the 1920s, physicist Wilhelm Lenz and his student Ernst Ising introduced a mathematical model of *ferromagnetism*, namely the 1-dimensional case of the famous *Ising model*. Over the following decades, several generalizations of the original Ising model have been proposed and studied, including a class of models known as *spin glasses*. We will begin with the general setup that unifies many of these Ising-like models. Much of this material (from different perspectives) can be found in the first chapter of David Tong's notes on statistical field theory [Ton17] and Michel Talagrand's *Mean Field Models for Spin Glasses* [Tal11].

Consider a collection of particles, each of which takes on one of two polar opposite states, represented as $+1$ or -1. This state is referred to as the particle's *Ising spin*, or simply its *spin*. The particles are arranged in a graph, often a lattice, where the vertices are called *sites* and the edges *bonds*. The set of sites will be denoted Λ, and the spin at each site $i \in \Lambda$ will be denoted $\sigma_i \in \{\pm 1\}$. We can combine these spins into a vector $\boldsymbol{\sigma} \in \{\pm 1\}^\Lambda$ representing the spin configuration of the entire system. Of course, the use of "spin" comes from particle physics. While we are primarily motivated by magnetic behavior, where an electron's magnetic moment (a vector quantity) rather than simply its spin (a discrete quantity) is most relevant, modeling only spin will be enough for air density, graphical models in machine learning, etc. There are models, *vector spin glasses*, that deal with "vector spins" rather than binary spins, but we will not concern ourselves with these.

For any two sites $i, j \in \Lambda$, there is a real-valued interaction J_{ij}, which we can store in a symmetric $\Lambda \times \Lambda$ matrix J, the *disorder* matrix. In the absence of an external field, we approximate the total energy of the system as a sum of the energies resulting from each pairwise interaction of particles. More generally, one could model interactions of three or more particles at a time, as well as an external field causing an additional interaction h_i at each site $i \in \Lambda$. However, we will restrict our attention to a closed system (no external field) with only pairwise interactions. We refer to this setup as a *0-field 2-spin model* with disorder J, or simply a *spin model*. Note that we do not need to specify Λ, since this can be recovered as the row or column index set for the matrix J. The total energy of a spin model, as a function of the spin configuration $\boldsymbol{\sigma}$, is given by the *Hamiltonian*

$$\mathcal{H}(\boldsymbol{\sigma}) = -\sum_{i,j \in \Lambda} J_{ij}\sigma_i\sigma_j = -\boldsymbol{\sigma}^\intercal J \boldsymbol{\sigma}. \tag{1.1}$$

A physical system always favors the lowest possible energy state, so that spin configurations minimizing the Hamiltonian are most desirable in equilibrium. It is therefore natural to interpret each interaction coefficient J_{ij} as a "signed weight" for the bond between two sites $i, j \in \Lambda$. The magnitude $|J_{ij}|$ corresponds to the strength of this interaction—its relative energy contribution. Meanwhile, since a more positive value of $J_{ij}\sigma_i\sigma_j$ will lower the total energy, the sign of J_{ij} encodes the preferred spin alignment among the particles at sites i and j. In particular, if $\sigma_i = \sigma_j$, then we say that these spins are *parallel*. Otherwise, they are *antiparallel*. The situation is summarized in Figure 1.

Sign of J_{ij}	**Favored Alignment**	**Terminology**
$J_{ij} > 0$	Parallel	*Ferromagnetic*
$J_{ij} = 0$	N.A.	No interaction
$J_{ij} < 0$	Antiparallel	*Anti*ferromagnetic

Figure 1: Spin alignment among sites i and j.

The terminology "ferromagnetic" and "antiferromagnetic" comes from the study of *ferromagnetism*, where it has been observed that materials such as iron (Latin *ferrum*) exhibit parallel alignment of electrons: all molecular magnetic dipoles point in the same direction. Alloys with this property are often called *ferromagnets*. In the context of a spin model, this means each J_{ij} is nonnegative.

If $J_{ij} = 0$ for a pair $i, j \in \Lambda$, the particles at sites i and j are said to be *non-interacting*. For simplicity, we view the graph of particles as fully-connected, with edge weights described by the disorder matrix J. Alternatively, we can account for non-interacting particles by only connecting interacting sites with edges, and then the sum in the Hamiltonian indexes over the edges of this graph. In the standard Ising model, for instance, only local interactions are allowed, so that most entries in the disorder matrix are zero. We also frequently disallow interactions between a particle and *itself*, which corresponds to setting $\operatorname{diag}(J) = 0$. Then the graph of particles is simple, i.e., there are no graph loops.

Example 1.1 [Nearest Neighbor Interactions]. In many cases, the particles in a spin model are assumed to lie in a d-dimensional lattice, with only neighboring lattice sites interacting. If two sites $i, j \in \Lambda$ are distinct neighbors, we write $\langle ij \rangle$, so the Hamiltonian becomes

$$\mathcal{H}(\boldsymbol{\sigma}) = -\sum_{\langle ij \rangle} J_{ij}\sigma_i\sigma_j. \tag{1.2}$$

Then we only have to define J_{ij} when $\langle ij \rangle$. If this lattice is infinite, say the standard lattice $\mathbb{Z}^d$ in $\mathbb{R}^d$, then the above sum is infinite. In order to avoid concerns about convergence, it is common to restrict our attention to a parallelepiped with N^d sites, where N is taken to be sufficiently large (i.e., we take $N \to \infty$).

There is one final parameter of significance for any spin model, namely the *inverse temperature* $\beta > 0$. Large values of β correspond to small temperatures, in which case we are said to be in a *low temperature regime*. Conversely, values of β close to 0 represent high temperatures. We would like to model our expectation that lower temperatures favor a more orderly system. This intuition is captured by the *Gibbs measure* (also referred to as the *Boltzmann distribution* by physicists).

The Gibbs measure p associated to a spin model with disorder J and inverse temperature β is a discrete probability measure on the hypercube $\{\pm 1\}^\Lambda$ of spin configurations. We interpret $p(\boldsymbol{\sigma})$ as the probability that a random system representative of the associated model has spin configuration $\boldsymbol{\sigma}$ at equilibrium.

Definition 1.2 [Gibbs Measure]. Consider a spin model with disorder J and inverse temperature β. The associated Gibbs measure p on $\{\pm 1\}^\Lambda$ is defined by the PMF

$$p(\boldsymbol{\sigma}) \propto e^{-\beta \mathcal{H}(\boldsymbol{\sigma})}. \tag{1.3}$$

The normalization constant for this measure is called the *partition function* $\mathcal{Z}_\beta$, which has sparked much independent investigation. Note that

$$\mathcal{Z}_\beta = \sum_{\boldsymbol{\sigma} \in \{\pm 1\}^\Lambda} e^{-\beta \mathcal{H}(\boldsymbol{\sigma})} \implies p(\boldsymbol{\sigma}) = \frac{e^{-\beta \mathcal{H}(\boldsymbol{\sigma})}}{\mathcal{Z}_\beta}. \tag{1.4}$$

Since $x \mapsto e^{-\beta x}$ is a decreasing positive function on $\mathbb{R}$, the Gibbs measure places more probability mass on spin configurations which minimize the Hamiltonian. The effect of the exponential transformation is to exaggerate the discrepancy between the highest probability spin configuration and the other possible configurations. As $\beta \to \infty$, the Gibbs measure p approaches a uniform measure on the minimizer(s) $\boldsymbol{\sigma}$ of $\mathcal{H}(\boldsymbol{\sigma})$. In other words, as the temperature drops, the system is more likely to be in its most orderly (i.e., energy-minimizing) spin configuration at equilibrium. As the temperature increases, so too does the probability of observing a suboptimal spin configuration at equilibrium.

Example 1.3 [Standard Ising Model]. The standard d-dimensional Ising model (with no external field) is a 0-field 2-spin model on a d-dimensional lattice with only nearest neighbor interactions. Moreover, the disorder J is deterministic, and for simplicity, we assume that all nonzero entries J_{ij} for $\langle ij \rangle$ are equal to a fixed constant J. The Hamiltonian is

$$\mathcal{H}(\boldsymbol{\sigma}) = -J \sum_{\langle ij \rangle} \sigma_i \sigma_j. \tag{1.5}$$

Again, to avoid infinite sums, we may restrict our attention to a finite, arbitrarily large portion of the lattice. If the constant J is positive, then this models a ferromagnet, so that the parallel configurations $\{+1\}^\Lambda$ and $\{-1\}^\Lambda$ are most favorable.

A *spin glass*, unlike the standard Ising model, is characterized by *random* disorder. In other words, the disorder matrix J is now a random variable. The associated Gibbs measure is therefore a random measure. Accordingly, rather than modeling systems in which all or a dominating majority of the bonds are (anti)ferromagnetic, in a spin glass we often prescribe a symmetric distribution for each nonzero J_{ij}, leading to relatively equal amounts of ferromagnetic and antiferromagnetic interactions. This creates significant disorder such that the interactions are called *frustrated*—see Figure 2. In low temperature regimes, the magnetic disorder of a spin glass is reminiscent of the positional disorder in materials like window glass, which is the origin of the term "spin glass." For a spin glass, one often refers to the low temperature regime as the *glassy phase*. From now on, by a *spin glass* we shall simply mean a 0-field 2-spin model with random disorder matrix.

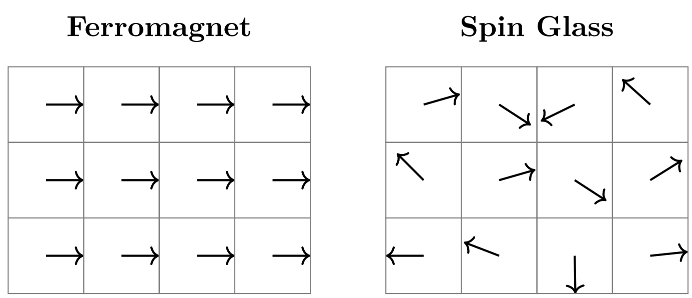

Figure 2: Parallel vs. frustrated magnetic moments at low temperatures.

Spin glasses have several applications outside the realm of statistical physics. In neuroscience, spin glasses are well-equipped to model *biological* neural networks, where the particles represent neurons and the Ising spin indicates whether or not a given neuron is activated. Analogously, spin glasses can also model *artificial* neural networks, making them relevant in machine learning and computer science. The most famous example is the Hopfield network, a recurrent neural network that functions as associative memory. This network is actually a spin glass, and the case of random initialization corresponds with the famous SK model (to be introduced shortly).

Example 1.4 [*Edwards-Anderson* (EA) Model]. The EA model is a spin glass with many similarities to the standard Ising model. In particular, the particles are arranged in a d-dimensional (partial) lattice with only nearest neighbor interactions. However, these interactions are random. Most commonly, the random variables J_{ij} for $\langle ij \rangle$ are i.i.d. Gaussian with mean J_μ and variance J_σ^2. We call this the *standard* or *Gaussian* EA model. If $J_\mu = 0$, then the disorder is symmetric about 0. In dimensions greater than 2, we observe a transition into a glassy phase when the inverse temperature β surpasses a critical threshold, an instance of *phase transition*. This critical temperature may be approximated by Monte Carlo simulation, but an exact solution is evasive. Even a mathematically rigorous proof that the EA model exhibits "glassy" behavior is relatively recent, with the first such proof, due to Chatterjee, only surfacing in 2023 [Cha23].

Both the standard Ising model and the EA model forbid long range interactions, creating a spatial correlation structure that complicates the mathematical analysis. However, the situation becomes much more tractable once the locality constraint is discarded. So long as we maintain random frustrated interactions, controlled by the disorder matrix, we can obtain macroscopic "glassy" behavior similar to the EA model. In effect, we replace every local interaction with a global *mean* interaction. If we properly scale the strength of these *mean fields* with the number N of sites, taking $N \to \infty$ is a good approximation to the EA model as the lattice dimension d tends to infinity. This is because in high dimensions, all pairwise Euclidean distances become large and relatively similar, so that a complete graph is a reasonable approximation to the lattice. Physicists often call $d \to \infty$ the *mean field limit*. Of course, in natural settings where the lattice is spatial, only dimensions $d \leq 3$ are realistic. Nonetheless, the mean field limit is a useful approximation even in low dimensions. Importantly, we still observe a glassy phase at low temperatures, so the term "spin glass" remains appropriate. This brings us to (arguably) the most important spin glass model.

Definition 1.5 [*Sherrington-Kirkpatrick* (SK) Model]. Let N be a positive integer. The N-dimensional *Gaussian* SK model is a spin glass with $N \times N$ disorder matrix J such that $\mathrm{diag}(J)$ is deterministically 0 and $(J_{ij})_{1\leq i<j\leq N}$ are i.i.d. normal with mean 0 and variance C/N, where $C > 0$ is independent of N. Since J is symmetric, this completely specifies its distribution. The Hamiltonian is

$$\mathcal{H}_N(\boldsymbol{\sigma}) = -2 \sum_{1\leq i<j\leq N} J_{ij}\sigma_i\sigma_j. \tag{1.6}$$

The scale factor C is dimensionless and hence insignificant for the limit, so we often take $C = 1/4$ in order to rewrite

$$\mathcal{H}_N(\boldsymbol{\sigma}) = -\frac{1}{\sqrt{N}} \sum_{i<j} g_{ij}\sigma_i\sigma_j, \tag{1.7}$$

where $(g_{ij})_{i<j}$ is a vector of i.i.d. standard Gaussians. Slightly more generally, we could fix an arbitrary (preferably symmetric) Borel measure μ on $\mathbb{R}$ with mean 0 and variance C/N, in which case the disorder J is defined as a random $N \times N$ matrix where

$$\mathrm{diag}(J) = 0 \quad \text{and} \quad (J_{ij})_{i<j} \overset{i.i.d.}{\sim} \mu. \tag{1.8}$$

We will refer to this spin glass as the *standard* SK Model associated to μ. As before, the constant C is unimportant, and one often takes $C = 1/4$.

The SK model is in some sense the "prototypical" spin glass, for it occurs as a (suitably interpreted) limit of many natural classes of spin glasses. Precise formulations of this heuristic can be found in [DMS17] and [Pan18], with algorithmic variants in [EMS23] and [Jon+23]. Amazingly, the equilibrium solution of the Gaussian SK model was discovered by Giorgio Parisi in 1979 via a technique known as the *replica method* [Par79]. Various tools were developed and employed over the following decades to mathematically verify Parisi's formula, the most notable of which is the so-called *cavity method*. Finally, in 2006, Michel Talagrand presented the first rigorous proof valid for all $\beta > 0$ [Tal06], building on previous work by Francesco Guerra [Gue03]. By "equilibrium solution," we mean an explicit formula, the *Parisi formula*, for the limit of the *quenched Gibbs free energy* of the system, divided by N, as $N \to \infty$. The free energy is a key quantity in statistical mechanics that represents a system's equilibrium state based on energy and entropy calculations. The term "quenched" just means that we do not allow our disorder (the randomness of J) to evolve over time. One often says that the disorder is *frozen* in the quenched setting.

Definition 1.6 [Free Energy]. For a spin glass with disorder J and inverse temperature β, the *quenched Gibbs free energy*, or simply the *free energy*, is

$$F_\beta := -\frac{1}{\beta}\mathbb{E}[\log \mathcal{Z}_\beta] = -\frac{1}{\beta}\mathbb{E}\Big[\log\Big(\sum_{\boldsymbol{\sigma}\in\{\pm 1\}^\Lambda} e^{-\beta\mathcal{H}(\boldsymbol{\sigma})}\Big)\Big], \tag{1.9}$$

where $\mathcal{Z}_\beta$ is the partition function (see Definition 1.2).

If we remove the expectation in Definition 1.6, we're left with a $\sigma(J)$-measurable random variable (i.e., a measurable function of the random matrix J). Somewhat confusingly, both this random variable and its expectation are often referred to as the free energy. The expected value is conventional in physics, but we will switch perspective momentarily.

1.2 Gaussian Disorder

We begin this subchapter by slightly tweaking some previous definitions to accord with mathematical rather than physical convention. This just amounts to removing irrelevant signs and, as discussed at the end of the previous subchapter, dropping the expectation in Definition 1.6. From now on, we work only with Definition 1.7.

Definition 1.7 [(Mathematician's) Hamiltonian and Free Energy]. For a spin glass with disorder J and inverse temperature β, we redefine the *Hamiltonian* and *free energy* as

$$\mathcal{H}(\boldsymbol{\sigma}) = \boldsymbol{\sigma}^\intercal J\boldsymbol{\sigma} \quad \text{and} \quad F_\beta = \frac{1}{\beta}\log\Big(\sum_{\boldsymbol{\sigma}\in\{\pm 1\}^\Lambda} e^{\beta\mathcal{H}(\boldsymbol{\sigma})}\Big). \tag{1.10}$$

With this new convention, the Hamilton associated to the N-dimensional Gaussian SK model in Definition 1.5 can be written

$$\mathcal{H}_N(\boldsymbol{\sigma}) = \frac{1}{\sqrt{N}}\sum_{1\le i<j\le N} g_{ij}\sigma_i\sigma_j, \tag{1.11}$$

where $(g_{ij})_{i<j}$ is a collection of i.i.d. standard Gaussians. A fundamental quantity of interest, then, is the associated free energy $F_{N,\beta}$, which is a random variable by our new convention (note the added subscript to indicate dependence on the dimension N).

As noted in the previous subchapter, it has been shown that the limit $\lim_{N\to\infty} \mathbb{E}[F_{N,\beta}]/N$ exists and is finite, and moreover, we have an explicit (albeit convoluted) formula for this limit, the Parisi formula. In particular, we have the asymptotic result

$$\mathbb{E}[F_{N,\beta}] = \Theta_\beta(N), \tag{1.12}$$

which is to say that for any $\beta > 0$, there exist constants c_β and C_β such that

$$c_\beta N \leq \mathbb{E}[F_{N,\beta}] \leq C_\beta N \tag{1.13}$$

for sufficiently large N. However, a full asymptotic picture is incomplete until we understand the limiting behavior of the *variance* $\mathrm{Var}(F_{N,\beta})$. As a first approximation, tools from the theory of *concentration of measure*, taken up in Chapter 2, show us that

$$\mathrm{Var}(F_{N,\beta}) = O_\beta(N). \tag{1.14}$$

This is already enough to conclude that for fixed $\beta > 0$, the fluctuations of the free energy $F_{N,\beta}$ are at most on the order of $\sqrt{N}$, while the expectation grows linearly with N. This is an example of *concentration*—in high dimensions, the free energy is closely concentrated around its mean. However, from both simulations and experimental observations, even the upper bound $O(N)$ on the variance appears potentially too conservative. In Figure 3 we provide a plot of $\mathrm{Var}(F_{N,\beta})/N$ simulated for $1 \leq N \leq 20$ and $\beta \in \{0.1, 1, 10\}$. Even for relatively small values of N, so that the computation of $F_{N,\beta}$ is tractable without Monte Carlo approximations, these graphs suggest an upper bound $o_\beta(N)$ might be available. Many physicists even expect a bound like $O_\beta(1)$, according to Chatterjee in his comments following [Cha14, Theorem 1.6].

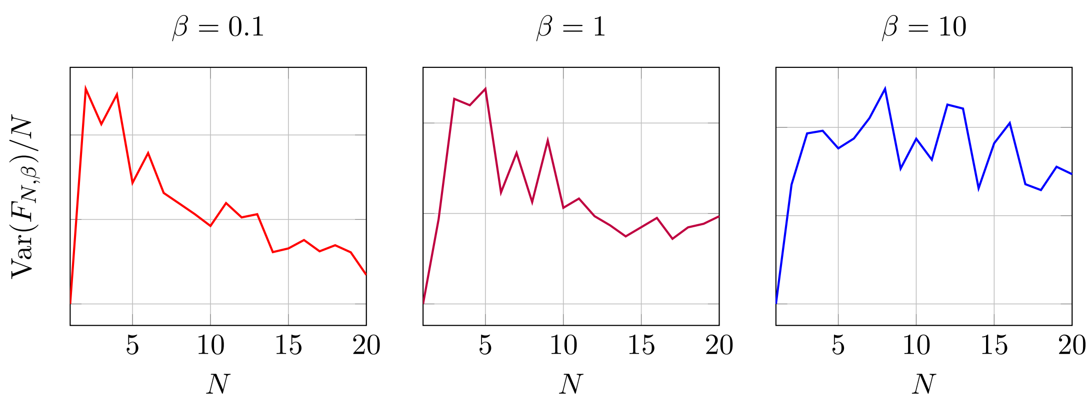

Figure 3: $\mathrm{Var}(F_{N,\beta})/N$ vs. N for different β.

It is in precisely this setting that Sourav Chatterjee's theory of *super*concentration becomes relevant. Chatterjee dedicates an entire chapter to the Gaussian SK model in his 2014 monograph, outlining what he calls the *spectral method* for superconcentration; using this technique, he arrives at the following theorem [Cha14, Theorem 6.3].

Theorem 1.8 [Chatterjee]. Let $F_{N,\beta}$ be the free energy for the N-dimensional Gaussian SK model. Then for each $\beta > 0$, there exists a constant C_β such that

$$\operatorname{Var}(F_{N,\beta}) \leq \frac{C_\beta N \log\log N}{\log N}. \tag{1.15}$$

Note that $\log\log N/\log N \to 0$ as $N \to \infty$, albeit extremely slowly. As a result, we obtain the following important corollary, which is a significant improvement over (1.14).

Corollary 1.9. For the N-dimensional Gaussian SK model, we have

$$\operatorname{Var}(F_{N,\beta}) = o_\beta(N). \tag{1.16}$$

Our main result is motivated in large part by Chatterjee's proof of Theorem 1.8. Accordingly, we devote the next chapter to developing this proof strategy in a way that highlights the latent symmetry of the problem. For clarity, we will simplify Chatterjee's argument, arriving only at Corollary 1.9 (sufficient for our purposes) rather than Theorem 1.8 or an even sharper bound also due to Chatterjee [Cha09, Theorem 1.5].

1.3 Orthogonally Invariant Disorder

The main goal of this thesis is to understand a class of spin glasses whose disorder is uniformly random with respect to some fixed geometric structure. The geometry is controlled by a compact (Hausdorff) group G, and the uniform randomness is attained by sampling from the *Haar* measure on G (constructed in subchapter 3.1). Formally,

Definition 1.10 [G-Invariant SK Model]. Let N be a positive integer and G a compact group acting continuously on the space $\operatorname{Sym}_N\mathbb{R}$ of symmetric $N \times N$ matrices. Fix a matrix $\Lambda \in \operatorname{Sym}_N\mathbb{R}$. The N-dimensional *G-invariant SK model* with *structure* Λ and inverse temperature β is a spin glass with disorder

$$J \sim \operatorname{Haar}(G) \cdot \Lambda. \tag{1.17}$$

By (1.17) we mean that J has the distribution of $g \cdot \Lambda$, where $g \sim \operatorname{Haar}(G)$, so that J is in the most natural sense uniformly sampled on the G-orbit of Λ. Such a distribution is often referred to as an *orbital measure* on the group G.

Definition 1.11. [Orthogonally Invariant SK Model] Let N be a positive integer. Then we call an O_N-invariant SK model with structure Λ, where Λ is a diagonal matrix and O_N acts on $\operatorname{Sym}_N\mathbb{R}$ by conjugation $O \cdot A := O^\intercal A O$, an *orthogonally invariant* SK model.

Remark 1.12. This is equivalently an SO_N-invariant SK model with structure Λ (with SO_N acting by conjugation) because $f(O) \sim \mathrm{Haar}(\mathrm{SO}_N)$ if $O \sim \mathrm{Haar}(\mathrm{O}_N)$, where $f : \mathrm{O}_N \to \mathrm{SO}_N$ maps $A \mapsto A$ if $\det A = 1$ and $A \mapsto \mathrm{diag}(-1, 1, \ldots, 1)A$ otherwise. But certainly $O^\intercal \Lambda O = f(O)^\intercal \Lambda f(O)$, so the distribution of the disorder doesn't change. Working with the connected group SO_N rather than O_N can be more convenient, as we will see in Chapter 3.

By the spectral theorem, any symmetric $N \times N$ matrix is diagonalizable with respect to an orthonormal basis of eigenvectors, all of whose eigenvalues are real. Thus, an orthogonally invariant SK model is one for which the disorder has fixed eigenvalues and uniformly random eigenvectors. Our main objective is to establish superconcentration for the free energy of an orthogonally invariant spin glass. The more general framework of G-invariant SK models provides a natural source for extensions, which we'll revisit in Chapter 4. Of course, in order to study these asymptotics, we need to specify a sequence of structure matrices $(\Lambda_N)_{N \geq 1}$. Therefore, we ask: what conditions on (Λ_N) lead to a superconcentration result like $\mathrm{Var}(F_{N,\beta}) = o_\beta(N)$? To do this, we compare with the Gaussian SK model, where we know superconcentration holds. This is motivated by the following relationship between $\mathrm{Haar}(\mathrm{O}_N)$ and the Gaussian distribution.

Theorem 1.13. Let G be an $N \times N$ matrix with i.i.d. standard Gaussian entries. Then G is almost surely invertible, and if we use Gram-Schmidt to write $G = QR$ for $Q \in \mathrm{O}_N$ and R upper-triangular, then $Q \sim \mathrm{Haar}(\mathrm{O}_N)$.

Proof. See the Appendix (C). □

Theorem 1.13 makes precise the heuristic that a standardized Gaussian matrix inherits randomness from two sources: random eigenvalues and uniformly random eigenvectors. Intuitively, an orthogonally invariant SK model should superconcentrate even better than the Gaussian SK model, so long as Λ_N grows no worse than the typical spectrum of the disorder matrix in the Gaussian case. After all, an orthogonally invariant disorder consists solely of random eigenvectors, with deterministic eigenvalues.

If we let G be as in Theorem 1.13, then this isn't quite the disorder matrix J of the Gaussian SK model. Up to a constant independent of N, however, we can realize the disorder J as the matrix $A_N = (G + G^\intercal)/\sqrt{2N}$ with its diagonal set to 0. Ignoring discrepancies with the diagonal, the matrix A_N follows a famous distribution: the N-dimensional *Gaussian orthogonal ensemble* $\mathrm{GOE}(N)$. Note that the strictly upper triangular entries of A_N are i.i.d. $\mathcal{N}(0, 1/n)$, the diagonal entries are i.i.d. $\mathcal{N}(0, 2/n)$ and independent of the upper triangular ones, and the lower triangular entries are chosen to make A_N symmetric. Conveniently, the asymptotic spectrum of a $\mathrm{GOE}(N)$ matrix is well-known. Given a random $A \in \mathrm{Sym}_N(\mathbb{R})$ with eigenvalues $\lambda_1(A), \ldots, \lambda_N(A)$, the *spectral density* is the random measure

$$\mu_A := \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i(A)}. \tag{1.18}$$

Wigner's famous *semicircular law* states that the spectral density of a GOE(N) matrix converges (in a suitable sense) to the *semicircular distribution* on $[-2, 2]$ as $N \to \infty$. This is the distribution whose PDF $f(x)$ is proportional to a semicircle of radius 2, i.e.,

$$f(x) = \frac{1}{2\pi}\sqrt{4 - x^2} \quad \text{for } x \in [-2, 2]. \tag{1.19}$$

See Figure 4 for a visualization of this semicircular law, where we simulate a random matrix A from GOE(1000) and plot a histogram of its eigenvalues against the PDF $f(x)$. Note how the histogram, which represents the spectral density μ_A of A, closely approximates the semicircular density function $f(x)$.

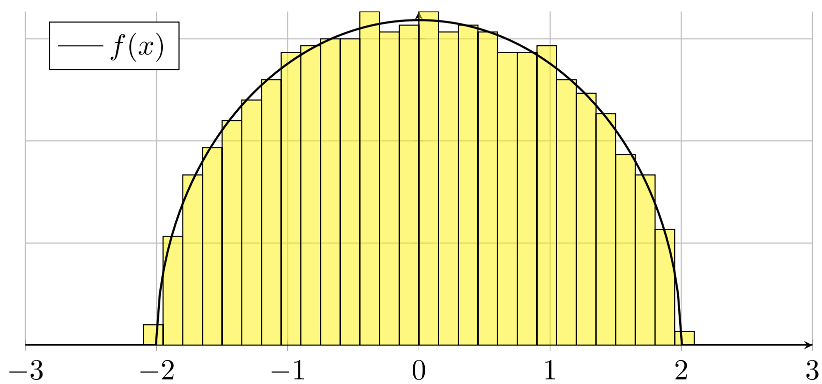

Figure 4: Spectrum of a matrix $A \sim$ GOE(1000) (simulated with Python).

Accordingly, we should impose a condition on Λ_N mimicking the behavior of the semicircular density. Namely, we require that the entries of Λ_N be uniformly bounded, i.e., $\|\Lambda_N\| = O(1)$. We now examine by simulation whether or not this is a reasonable condition for superconcentration. First, we can use Theorem 1.13 to sample from the Haar measure on O_N; one simply generates the matrix G, performs Gram-Schmidt to obtain a QR decomposition, and then extracts Q. (Since most computer algebra systems do not compute QR by Gram-Schmidt for numerical stability reasons, we may need to tweak Q if it is obtained via a different algorithm.) For each N and a fixed $\beta > 0$ and Λ_N, we can use this technique to sample 1000 values from $F_{N,\beta}$ and take the sample variance.

For $\beta \in \{0.1, 1, 10\}$ we desire a plot of the sample variance, divided by N, against N, much as we did in Figure 3. To account for different possible values of Λ_N, for each N we sample 100 Λ_N with diagonal entries chosen uniformly and independently from the interval $[-1, 1]$ (so that $\sup_N \|\Lambda_N\| \leq 1$), and for each Λ_N we apply this procedure to obtain a separate sample variance. For each N and β, we choose to plot the maximum of these sample variances (divided by N). Overall, this algorithm is $O(2^N)$, so we will have to restrict ourselves to sufficiently small N, say $1 \leq N \leq 15$. We plot the resulting graphs in Figure 5. Indeed, our intuition that the variance should be $o_\beta(N)$ appears reasonable if we compare with Figure 3. We are therefore led to the following conjecture.

Conjecture 1.14. [Main] Fix $\beta > 0$. Let (Λ_N) be a sequence of real $N \times N$ diagonal matrices with $\|\Lambda_N\| = O_\beta(1)$. Then if $F_{N,\beta}$ is the free energy of an N-dimensional orthogonally invariant SK model with structure Λ_N, we have

$$\mathrm{Var}(F_{N,\beta}) = o_\beta(N). \tag{1.20}$$

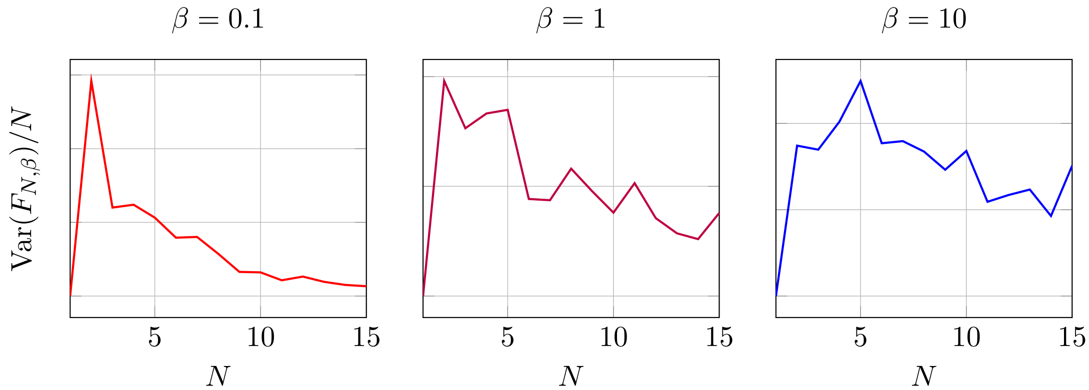

Figure 5: $\mathrm{Var}(F_{N,\beta})/N$ vs. N in the orthogonally invariant case.

Conjecture 1.14 is the motivation for this thesis, and we present progress towards a proof in Chapter 4. This is, of course, completely analogous to Corollary 1.9. However, for a full comparison with the Gaussian SK model, we should also investigate the expectation of $F_{N,\beta}$ in the orthogonally invariant case. Assume the setup of Conjecture 1.14. We claim that $\mathbb{E}[F_{N,\beta}] = O_\beta(N)$, and conjecture that also $\mathbb{E}[F_{N,\beta}] = \Omega_\beta(N)$, so the expectation, as in the Gaussian case, is $\Theta_\beta(N)$. The upper bound follows readily from the following theorem, which also implies, as a preliminary naive upper bound, that $\mathrm{Var}(F_{N,\beta}) = O_\beta(N^2)$.

Theorem 1.15. Let $F_{N,\beta}$ be the free energy of an N-dimensional orthogonally invariant SK model with structure Λ, viewed as a function $\mathrm{O}_N \to \mathbb{R}$. Then for all $O \in \mathrm{O}_N$, we have

$$\left| F_{N,\beta}(O) - N\frac{\log 2}{\beta} \right| \le N\|\Lambda\|. \tag{1.21}$$

Proof. Let $\boldsymbol{\sigma} \in \{\pm 1\}^N$. First we bound the Hamiltonian $\mathcal{H}_N(\boldsymbol{\sigma})$. Since O is orthogonal and $\|\boldsymbol{\sigma}\|^2 = N$, we have

$$|\mathcal{H}_N(\boldsymbol{\sigma})| = |(O\boldsymbol{\sigma})^\intercal \Lambda (O\boldsymbol{\sigma})| \le \|\Lambda\|\|O\boldsymbol{\sigma}\|^2 = \|\Lambda\|\|\boldsymbol{\sigma}\|^2 = N\|\Lambda\|. \tag{1.22}$$

Since $x \mapsto e^{\beta x}$ is an increasing function of $x \in \mathbb{R}$, from (1.22) we deduce that the partition function $\mathcal{Z}_\beta$ can be bounded

$$2^N e^{-N\beta\|\Lambda\|} \le \mathcal{Z}_\beta \le 2^N e^{N\beta\|\Lambda\|}. \tag{1.23}$$

Taking the logarithm of (1.23) and dividing by β implies (1.21), as desired. $\square$

Now, as for the lower bound on the expectation, we run simulations identical to those for the variance, expect with the sample mean of $F_{N,\beta}$, as opposed to the sample variance divided by N, plotted against N. The graphs are displayed in Figure 6. It appears that the expectation does in fact grow linearly with N, motivating the following conjecture (which we do not attempt to make progress towards proving).

Conjecture 1.16. Let $F_{N,\beta}$ be as in Conjecture 1.14. Then $\mathbb{E}[F_{N,\beta}] = \Theta_\beta(N)$.

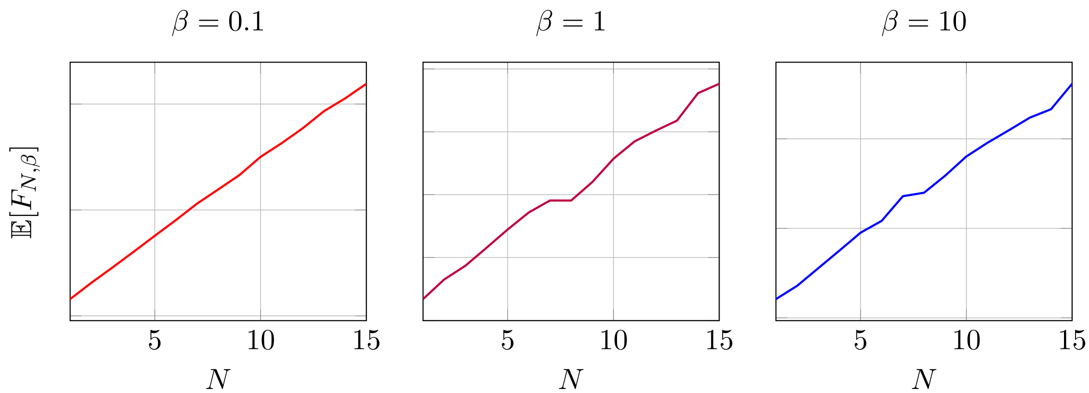

Figure 6: $\mathbb{E}[F_{N,\beta}]$ vs. N in the orthogonally invariant case.

Assuming Conjecture 1.16, we deduce that the asymptotics of an orthogonally invariant SK model mirror those of the Gaussian SK model. While the expectation of $F_{N,\beta}$ grows like N, the fluctuations grow much slower than even $\sqrt{N}$, so that the free energy concentrates very tightly. In Chapter 4, we will also demonstrate that the upper bound on $\mathrm{Var}(F_{N,\beta})$ given by classical concentration theory is $O_\beta(N)$, so that the concentration implies by Conjecture 1.14 is indeed an example of *super*concentration.

We end this chapter by noting that recent results on orthogonally invariant spin glasses at high temperatures can be found in [BS15], [FLS22], and [FW24]. In the low temperature regime, very little is known. This motivates proving superconcentration for all $\beta > 0$, since other basic properties of the model are not understood.

2 Superconcentration for the Gaussian SK Model

2.1 Concentration of Measure

It is no understatement that one of the most foundational and significant phenomena in all of high-dimensional probability is *concentration of measure*. Interestingly, this behavior was first observed by geometers studying the closely related problem of *isoperimetry*. Let $(\mathcal{X}, d)$ be a metric space with Borel probability measure μ. Geometrically, we view μ as a density, so that $\mu(A)$ represents the fraction of the mass of $\mathcal{X}$ contained in A. For $\varepsilon > 0$ we define the ε-*fattening* of a nonempty Borel set $A \subseteq \mathcal{X}$ as

$$A^\varepsilon := \{x \in \mathcal{X} \mid d(x, A) \leq \varepsilon\}. \tag{2.1}$$

Intuitively, A^ε is the Borel set obtained by stretching out the boundary/*perimeter* of A by a distance ε in all directions. Note that A^ε is a closed set, for it is the preimage of $[0, \varepsilon] \subseteq \mathbb{R}$ under the continuous (in fact 1-Lipschitz) map $x \mapsto d(x, A)$.

Heuristically, we say that μ exhibits *concentration* or *isoperimetry* when the following holds: if a subset of $\mathcal{X}$ already contains at least half of the mass, then an arbitrarily small fattening of the set is enough to cover most of the mass. Formally, μ satisfies an *isoperimetric inequality* whenever there exist positive universal constants C and σ^2 such that

$$\mu(A^\varepsilon) \geq 1 - C \exp\Big(-\frac{\varepsilon^2}{2\sigma^2}\Big) \tag{2.2}$$

for all $\varepsilon > 0$ and Borel sets A such that $\mu(A) \geq 1/2$.

Example 2.1 [Spherical Isoperimetry]. In the early 20th century, Paul Lévy uncovered the first example of what one might retroactively dub *isoperimetry*. Let S^n be the unit n-dimensional sphere embedded in $\mathbb{R}^{n+1}$, and σ_n the standard geodesic probability measure on S^n from Riemannian geometry. This is the measure which agrees with our intuitive notion of "surface area," normalized so that $\sigma_n(S^n) = 1$. Lévy was interested in regions of the sphere with minimal "perimeter" for fixed surface area, determining that these regions are precisely *spherical caps*, i.e., regions of the form

$$\rho(x, R) = \{y \in S^n \mid \|x - y\| \leq R\} \tag{2.3}$$

for $x \in S^n$ and $R > 0$ (here $\|\cdot\|$ is Euclidean distance in $\mathbb{R}^{n+1}$); see Figure 7. More precisely, if A is any Borel subset of S^n with $\sigma_n(A) = \sigma_n(\rho(x, R))$, then for all $\varepsilon > 0$,

$$\sigma_n(A^\varepsilon) \geq \sigma_n(\rho(x, R)^\varepsilon). \tag{2.4}$$

From this fact (which may be deduced from the Brunn-Minkowski Theorem), one can derive the *spherical isoperimetric inequality*: there exist universal $C > 0$ and $c > 0$ such that

$$\sigma_n(A^\varepsilon) \geq 1 - C \exp(-cn\varepsilon^2) \tag{2.5}$$

for all dimensions n and all Borel sets $A \subseteq S^n$ with $\sigma_n(A) \geq 1/2$ [Led01, Section 2].

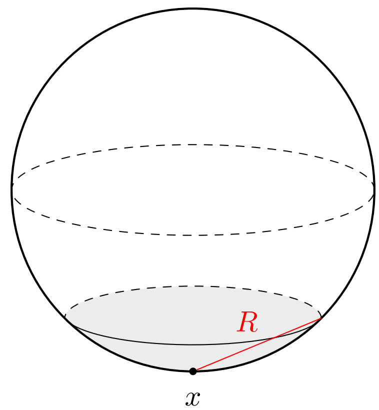

Figure 7: A spherical cap $\rho(x, R)$ (shaded gray).

The theory of concentration of measure, as we understand it today, can be traced back to the work of Vitali Milman in the 1970s. The key idea comes from the following functional reformulation of isoperimetry. Recall that a *median* of a real-valued function $f : \mathcal{X} \to \mathbb{R}$ is any $m \in \mathbb{R}$ such that $\mu(\{f \geq m\}) \geq 1/2$ and $\mu(\{f \leq m\}) \geq 1/2$. Such an m always exists, although it may not be unique. We denote the set of medians of f as $\mathrm{med}(f)$.

Theorem 2.2 [Functional Isoperimetry]. Let $(\mathcal{X}, d)$ be a metric space with Borel probability measure μ. Then μ satisfies the isoperimetric inequality (2.2) if and only if for all 1-Lipschitz $f : \mathcal{X} \to \mathbb{R}$, $m \in \mathrm{med}(f)$, and $\varepsilon > 0$, we have

$$\mu(\{f - m \geq \varepsilon\}) \leq C \exp\Big(-\frac{\varepsilon^2}{2\sigma^2}\Big). \tag{2.6}$$

Proof. See the Appendix (C). □

Importantly, if $(\mathcal{X}_n, d_n, \mu_n)$ is a sequence of metric measure spaces, and an isoperimetric inequality holds for each measure μ_n with respect to a *universal* constant C (i.e., independent of n) and $\sigma_n^2 = O(n^{-1})$, then the sequence $(\mathcal{X}_n, d_n, \mu_n)$ is said to form a *normal Lévy family*. We view n as a dimension proxy, so that intuitively a normal Lévy family is one such that concentration around the median becomes even more pronounced as the dimension increases.

Example 2.3 [Spherical Isoperimetry (*Revisited*)]. Returning to Example 2.1, notice that $(S^n, \|\cdot\|, \sigma_n)$ is a normal Lévy family. This implies a potentially surprising fact about the geometry of high-dimensional spheres. If $H \subseteq S^n$ is a high-dimensional hemisphere, so that $\sigma_n(H) = 1/2$, then $\sigma_n(H^\varepsilon) \approx 1$ for $\varepsilon \approx 0$. Consequently, if we fix any equator on a high-dimensional sphere, then most of the sphere's surface area lies "near" the equator. This is purely a high-dimensional phenomenon—see Figure 8.

Density $\sigma_n(E^\varepsilon)$	$\varepsilon = 0.5$	$\varepsilon = 0.05$	$\varepsilon = 0.005$
$n = 2$	0.484123	0.0499844	0.00499998
$n = 20$	0.977574	0.17483	0.0176183
$n = 200$	≈ 1	0.520088	0.0563019
$n = 2000$	≈ 1	0.97468	0.176916
$n = 20000$	≈ 1	≈ 1	0.520496
$n = 200000$	≈ 1	≈ 1	0.974653

Figure 8: Surface density of S^n within an equatorial band E^ε, where E is an equator (computed with *Wolfram Mathematica* 14.1).

In many cases, it is more convenient to work with the mean of a random variable rather than its median, and these two quantities will of course be similar when concentration is present, so for probabilists the most common sort of concentration inequalities are those of the form

$$\mathbb{P}(f(X) - \mathbb{E}[f(X)] \geq \varepsilon) \leq B_\varepsilon, \tag{2.7}$$

where X is a random vector of arbitrarily large dimension, f is a measurable, real-valued function that is often implicitly defined (so as to be difficult to work with directly), and B_ε is some bound which decays very quickly with ε. (2.7) is an example of a *one-sided tail bound*. We can of course consider bounds in the other direction, bounding instead the probability that $f(X) \leq \mathbb{E}[X] - \varepsilon$, or even *two-sided* tail bounds of the form

$$\mathbb{P}(|f(X) - \mathbb{E}[f(X)]| \geq \varepsilon) \leq B_\varepsilon. \tag{2.8}$$

Given tail bounds in both directions, we can always apply the union bound to obtain a two-sided tail bound.

Example 2.4 [Exponential Tail Bound]. The most useful tail bounds, in connection with isoperimetry, have the form $B_\varepsilon = C\exp(-c\varepsilon^2)$, where C and c are positive constants. Such a bound is called an *exponential* tail bound. This tells us that the tails of $f(X)$ are no "fatter" than those of a Gaussian. Countless *concentration inequalities*, such as the famous *Azuma-Hoeffding* and *McDiarmid* inequalities, are exponential tail bounds. The spherical concentration inequality (2.5) is another example. Using the Gaussian Poincaré inequality in the following subchapter, along with another trick called *Gaussian interpolation*, one may derive a useful exponential tail bound for Lipschitz functions of i.i.d. Gaussians [BLM13, Theorem 5.6].

In high-dimensional probability, we care especially about tail bounds in some suitably defined infinite dimensional limit. For example, perhaps we have a sequence of Borel measurable functions $f_n : \mathbb{R}^n \to \mathbb{R}$ and random vectors $X_n \in \mathbb{R}^n$, in which case our tail bounds B_ε are actually dependent on n. The dependence of this bound on n is crucial.

If B_ε does not depend on n at all and is in that sense *universal* or *dimensionless*, we can justifiably say that X_n concentrates tightly around its expectation in high dimensions. Such universal bounds are rare in practice, but so long as the decay of B_ε with ε overpowers the dependence on the dimension n, we remain in the realm of high-dimensional concentration.

Tail bounds give a lot of information about the deviation of a random variable from its mean, but in many cases an understanding of the "average" deviation, namely the *variance*, suffices for illustrating concentration phenomena. An invaluable variance bound in the theory of concentration is the *Efron-Stein* inequality, which we now state and prove.

Theorem 2.5 [Efron-Stein]. Let $\mathcal{X}_1, \dots, \mathcal{X}_n$ be measurable spaces, and $X_1, \dots, X_n$ independent random variables with X_i valued in $\mathcal{X}_i$. Let $X = (X_1, \dots, X_n)$ and $X' = (X_1', \dots, X_n')$ an independent copy of X. For any measurable $f : \mathcal{X}_1 \times \cdots \times \mathcal{X}_n \to \mathbb{R}$ with $f(X) \in L^2$, define the variance proxy

$$\nu_f := \sum_{i=1}^n \mathbb{E}[\mathrm{Var}(f(X)|X_{-i})], \tag{2.9}$$

where X_{-i} is the random vector X with the ith entry removed. Then we have

$$\mathrm{Var}(f(X)) \le \nu_f = \frac{1}{2}\sum_{i=1}^n \mathbb{E}[(f(X) - f(X^{(i)}))^2] = \sum_{i=1}^n \mathbb{E}[(f(X) - f(X^{(i)}))_\pm^2], \tag{2.10}$$

where $X^{(i)}$ is X with its ith entry replaced by the independent copy X_i', and $(\cdot)_\pm$ denotes the positive/negative part of a real number.

Proof. The standard proof of this inequality makes use of a so-called *martingale-difference* sequence. Namely, for each integer $0 \le i \le n$ we define

$$\mathbb{E}_i[f(X)] := \mathbb{E}[f(X)|X_1, \dots, X_i], \tag{2.11}$$

so that in particular $\mathbb{E}_0[f(X)] = \mathbb{E}[f(X)]$ and $\mathbb{E}_n[f(X)] = f(X)$. If we define the differences $\Delta_i = \mathbb{E}_i[f(Z)] - \mathbb{E}_{i-1}[f(Z)]$ for $1 \le i \le n$, note that we have the telescoping sum

$$f(X) - \mathbb{E}[f(X)] = \sum_{i=1}^n \Delta_i. \tag{2.12}$$

The key point is that the (Δ_i) form a martingale-difference sequence with respect to the standard filtration on X. This is to say that each Δ_i is $\sigma(X_1, \dots, X_i)$-measurable, has expectation 0, and satisfies $\mathbb{E}_i[\Delta_j] = 0$ for all $i < j$. In particular, for $i < j$ we have

$$\mathrm{Cov}(\Delta_i, \Delta_j) = \mathbb{E}[\Delta_i \Delta_j] = \mathbb{E}[\mathbb{E}_i[\Delta_i \Delta_j]] = \mathbb{E}[\Delta_i \mathbb{E}_i[\Delta_j]] = 0, \tag{2.13}$$

so that $\mathrm{Var}(f(X)) = \sum_{i=1}^n \mathbb{E}[\Delta_i^2]$. So far, we have not used the independence assumption. The key observation is that by independence, $\mathbb{E}_i[\mathbb{E}[f(X)|X_{-i}]] = \mathbb{E}_{i-1}[f(X)]$, so

$$\Delta_i = \mathbb{E}_i[f(X)] - \mathbb{E}_{i-1}[f(X)] = \mathbb{E}_i[f(X) - \mathbb{E}[f(X)|X_{-i}]]. \tag{2.14}$$

By conditional Jensen's inequality, $\Delta_i^2 \leq \mathbb{E}_i[(f(X) - \mathbb{E}[f(X)|X_{-i}])^2]$, yielding

$$\mathrm{Var}(f(X)) = \sum_{i=1}^{n} \mathbb{E}[\Delta_i^2] \leq \sum_{i=1}^{n} \mathbb{E}[(f(X) - \mathbb{E}[f(X)|X_{-i}])^2] = \nu_f. \tag{2.15}$$

This gives the inequality in (2.10). The first equality follows from a conditional variant of the basic identity $\mathrm{Var}(X) = \frac{1}{2}\mathbb{E}[(X-Y)^2]$ for i.i.d. L^2 random variables X and Y. The second equality in (2.10) is a simple symmetrization argument based on the fact that

$$(f(X) - f(X^{(i)}))_+ \stackrel{d}{=} (f(X) - f(X^{(i)}))_- \tag{2.16}$$

because $f(X)$ and $f(X^{(i)})$ are i.i.d. □

One of the most useful corollaries of Theorem 2.5 applies to functions with the so-called *bounded difference* property. The guiding principle is that we should expect concentration of a function of several random variables whenever these random variables do not depend much on each other and the function doesn't depend too much on any individual input.

Definition 2.6. Let $\mathcal{X}_1, \ldots, \mathcal{X}_n$ be sets and $f : \mathcal{X}_1 \times \cdots \times \mathcal{X}_n \to \mathbb{R}$ a function. Then if there exist real numbers $c_1, \ldots, c_n \geq 0$ such that for all $x_1 \in \mathcal{X}_1, \ldots, x_n \in \mathcal{X}_n$ and all $1 \leq i \leq n$,

$$\sup_{x_i' \in \mathcal{X}_i} |f(x_1, \ldots, x_i, \ldots, x_n) - f(x_1, \ldots, x_i', \ldots, x_n)| \leq c_i, \tag{2.17}$$

we say that f satisfies the *bounded difference property* with parameters $c_1, \ldots, c_n$.

Corollary 2.7 [Bounded Difference Inequality]. Let $\mathcal{X}_1, \ldots, \mathcal{X}_n$ be measurable spaces, where $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$ is given the product σ-algebra. If $X = (X_1, \ldots, X_n)$ is a random variable valued in $\mathcal{X}$ with independent entries, and $f : \mathcal{X} \to \mathbb{R}$ has the bounded difference property with parameters $c_1, \ldots, c_n$, then

$$\mathrm{Var}(f(X)) \leq \frac{1}{4} \sum_{i=1}^{n} c_i^2. \tag{2.18}$$

Proof. By the independence and bounded difference assumptions, for each $1 \leq i \leq n$ the random variable $\mathrm{Var}(f(X)|X_{-i})$ can be expressed in the form $v_i(X_{-i})$, where for any possible input $\boldsymbol{x}_{-i} \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_{i-1} \times \mathcal{X}_{i+1} \times \cdots \times \mathcal{X}_n$, $v_i(\boldsymbol{x}_{-i})$ is the variance of a random variable whose support has diameter at most c_i. Therefore $|v_i| \leq c_i^2/4$, so that (2.18) follows from Efron-Stein, as desired. □

Our bounded difference inequality is merely a variance bound, but with extra work one can obtain an exponential tail bound for functions satisfying the bounded difference property. This result is known as *McDiarmid's inequality*, mentioned in Example 2.4.

We end this subchapter with two applications of Theorem 2.5 and Corollary 2.7. Both of these examples are meant to illustrate the power of concentration theory for studying *implicitly defined* functions of random variables.

Example 2.8 [Bin Packing]. Suppose that a delivery company receives n boxes, each of whose weight W_i (in pounds) is independently drawn from some unknown distribution supported on $[0,1]$. We do not necessarily assume that $W_1, \ldots, W_n$ are all drawn from the *same* distribution. The delivery company can pack as many boxes as they want into a bin, so long as the total weight in the bin does not exceed 1 pound. Then a quantity of interest to the delivery company is the minimum number B of bins required to pack the n boxes. Clearly we can write $B = f(W_1, \ldots, W_n)$ for a suitable measurable function $f : [0,1]^n \to \mathbb{Z}_{\geq 0}$. Amazingly, one can derive the bound

$$\mathrm{Var}(B) \leq \frac{n}{4} \tag{2.19}$$

from Corollary 2.7 without any assumptions on the distributions of W_i. The key point is that while a nice, closed-form expression for f is evasive, we can still deduce that f satisfies the bounded difference property with parameters $c_i = 1$ for all $1 \leq i \leq n$.

As an illustration of Example 2.8, we simulate $\mathrm{Var}(B)$ vs. n, as well was the theoretical upper bound $n/4$, in the case where $W_i \sim \mathrm{Beta}(i, 1)$ for $1 \leq i \leq n$ independently. We do this for all $1 \leq n \leq 20$. See Figure 9 below.

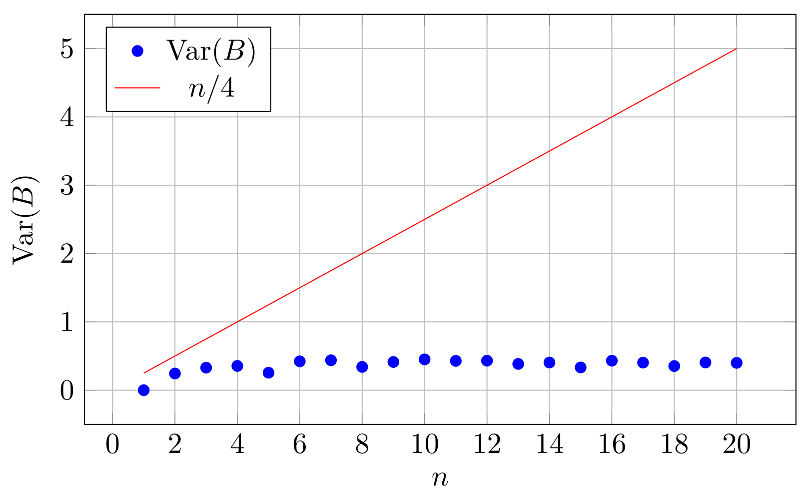

Figure 9: $\mathrm{Var}(B)$ vs. n, where $W_i \sim \mathrm{Beta}(i, 1)$.

Our final example is an application of concentration of measure in random matrix theory (RMT). Of considerable interest in RMT is the spectrum of a random matrix. Using only Efron-Stein, we can deduce a powerful universal result about the largest eigenvalue of a symmetric random matrix with uniformly bounded, independent entries.

Example 2.9 [Largest Eigenvalue]. Let A be a random $n \times n$ symmetric matrix with entries $A_{ij} \in [-1, 1]$, where $(A_{ij})_{1 \le i \le j \le n}$ are independent. Then the largest eigenvalue $Z = \lambda_{\max}(A)$, which is also the spectral norm when A is positive semidefinite, has variance bounded by 4. Notably, this is independent of n. To show this, write

$$Z = \sup_{\boldsymbol{u} \in S^{n-1}} \boldsymbol{u}^\intercal A \boldsymbol{u} = \boldsymbol{v}^\intercal A \boldsymbol{v}, \tag{2.20}$$

where $\boldsymbol{v}$ is a unit vector at which this supremum is achieved. Let $(A'_{ij})_{1 \le i \le j \le n}$ be an independent copy of $(A_{ij})_{1 \le i \le j \le n}$. For each $1 \le i \le j \le n$, we define a new matrix $A^{(i,j)}$ as the random symmetric matrix obtained from A by replacing the (i, j)th (and (j, i)th) entry with A'_{ij}. Then if $Z'_{ij} := \lambda_{\max}(A^{(i,j)})$, certainly $Z'_{ij} \ge \boldsymbol{v}^\intercal A^{(i,j)} \boldsymbol{v}$, so

$$Z - Z'_{ij} \le \boldsymbol{v}^\intercal A \boldsymbol{v} - \boldsymbol{v}^\intercal A^{(i,j)} \boldsymbol{v} = 2 v_i v_j (A_{ij} - A'_{ij}) \le 4 |v_i v_j| \tag{2.21}$$

for $i \ne j$. If $i = j$, we similarly get the bound $Z - Z'_{ii} \le 2 v_i^2$. Hence

$$(Z - Z'_{ij})_+^2 \le \begin{cases} 16 v_i^2 v_j^2 \mathbb{I}(Z > Z'_{ij}) & \text{if } i < j, \\ 4 v_i^4 \mathbb{I}(Z > Z'_{ii}) & \text{if } i = j. \end{cases} \tag{2.22}$$

By symmetry, $\mathbb{P}(Z > Z'_{ij}) \le 1/2$ for all $i \le j$. Now, note that $\lambda_{\max}$ is a measurable (in fact Lipschitz) function on the space of $n \times n$ symmetric matrices, which we can represent by their $n(n+1)/2$ upper triangular entries. Writing $Z = f((A_{ij})_{i \le j})$ in this way, by Efron-Stein and the fact that $\boldsymbol{v}$ is a unit vector we get

$$\operatorname{Var}(Z) \le \sum_{i \le j} \mathbb{E}[(Z - Z'_{ij})_+^2] \le \frac{1}{2} \left(\sum_{i=1}^n 4 v_i^4 + \sum_{i<j} 16 v_i^2 v_j^2 \right) \le 4 \Big(\sum_{i=1}^n v_i^2 \Big)^2 = 4. \tag{2.23}$$

As in Example 2.8, it's worth pointing out that in Example 2.9, no assumptions are made on the entries of A aside from their support and independence. In particular, they need not even follow the same distribution. Of course, nothing is special about the support $[-1, 1]$; we could have just as easily obtained a similar result by replacing $[-1, 1]$ with any other bounded interval. The same is true in Example 2.8, where one could work with W_i supported on $[0, M]$ and a maximum total weight of M pounds per bin for any $M > 0$.

2.2 The Gaussian Poincaré Inequality

Having established the basics of concentration theory in the previous subchapter, we hone our focus on a tool which is invaluable in proving concentration inequalities for functions of *Gaussian* random variables. We will denote the multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ by $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Recall that $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the continuous distribution on $\mathbb{R}^N$ with PDF

$$\varphi(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^N \det \boldsymbol{\Sigma}}} \exp\left\{ -\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu})^\intercal \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \right\}. \tag{2.24}$$

Theorem 2.10 [Gaussian Poincaré Inequality]. Let $Z \sim \mathcal{N}(0, I_N)$ and $f \in C^1(\mathbb{R}^n)$. Then if $f(Z)$ has finite variance,

$$\operatorname{Var}(f(Z)) \leq \mathbb{E}[\|\nabla f(Z)\|^2], \tag{2.25}$$

where $\|\cdot\|$ denotes the Euclidean norm of the $1 \times N$ random vector $\nabla f(Z)$.

Theorem 2.10 is one of many so-called *Poincaré inequalities.* There are several ways to arrive at (2.25), such as a clever CLT argument paired with Efron-Stein [BLM13, Theorem 3.20]. However, for our purposes, it will be most convenient to provide a *spectral* proof.

We will assume basic familiarity with Hilbert spaces, content which can be learned from, e.g., [Rud87, Chapter 4]. In particular, recall that for any (positive) measure μ on a measurable space $(\mathcal{X}, \Sigma)$, the space $L^2(\mu)$ consisting of square integrable functions $f : \mathcal{X} \to \mathbb{R}$, defined up to equality almost everywhere, is a (real) Hilbert space with respect to the inner product

$$\langle f, g\rangle_\mu := \int_{\mathcal{X}} fg\, d\mu. \tag{2.26}$$

A Hilbert basis for a Hilbert space H is a set of nonzero orthogonal vectors in H whose span is dense in H. The space H is separable if and only if it has a countable Hilbert basis. We shall call a measure space $(\mathcal{X}, \Sigma, \mu)$ *separable* if its underlying Hilbert space $L^2(\mu)$ is separable. For example, if Σ is countably generated and μ is σ-finite, then $L^2(\mu)$ is separable (A.1). This holds when μ is a σ-finite Borel measure on a second-countable topological space, for instance (A.2). We denote the norm induced by $\langle -, -\rangle_\mu$ as $\|\cdot\|_\mu$. Then if γ_N is the distribution $\mathcal{N}(0, I_N)$ on $\mathbb{R}^N$, we can restate Theorem 2.10 in slightly different language. The finite expectation condition is $f \in L^1(\gamma)$, while (2.25) becomes

$$\|f\|_{\gamma_N}^2 - \left(\int f\, d\gamma_N\right)^2 \leq \int \|\nabla f\|^2\, d\gamma_N. \tag{2.27}$$

This change of perspective is useful because $L^2(\gamma_N)$ comes with a convenient Hilbert basis, the set of *Hermite* polynomials, which we now introduce. To begin, we restrict our attention to the 1-dimensional case $N = 1$, writing $\gamma := \gamma_1$ (so that $\gamma_N = \gamma^{\otimes N}$).

Definition 2.11 [Hermite Polynomial]. The kth *Hermite* polynomial for $k \in \mathbb{Z}_{\geq 0}$ is

$$H_k(x) := (-1)^k e^{x^2/2} \frac{d^k}{dx^k} e^{-x^2/2}. \tag{2.28}$$

We can compute $H_0 \equiv 1$, $H_1(x) = x$, $H_2(x) = x^2 - 1$, and so on. Notice that each H_k is a monic polynomial of degree k. This follows easily from an important recurrence relation for the Hermite polynomials, which is also a useful computational tool.

Lemma 2.12 [Hermite Recurrence]. If $k \geq 0$, then we have the recurrence relation

$$H_{k+1}(x) = xH_k(x) - H_k'(x). \tag{2.29}$$

Proof. Let $k \geq 0$. Then by the product rule,

$$H_k'(x) = (-1)^k \frac{d}{dx}\left(e^{x^2/2} \cdot \frac{d^k}{dx^k} e^{-x^2/2}\right) = xH_k(x) - H_{k+1}(x). \tag{2.30}$$

Rearranging yields (2.29). $\square$

Corollary 2.13. For each $k \geq 0$, $H_k(x)$ is a monic polynomial of degree k.

Proof. This follows easily by induction on k. The base case $k = 0$ is trivial, as $H_0 \equiv 1$. For the inductive step, just use (2.29). $\square$

Example 2.14. Using Lemma 2.12, we can more efficiently determine the first several Hermite polynomials. Starting from $H_0 \equiv 1$, we can recursively compute

$$H_1(x) = x \cdot 1 - 0 = x, \tag{2.31}$$
$$H_2(x) = x \cdot x - 1 = x^2 - 1, \tag{2.32}$$
$$H_3(x) = x \cdot (x^2 - 1) - 2x = x^3 - 3x, \tag{2.33}$$
$$H_4(x) = x \cdot (x^3 - 3x) - (3x^2 - 3) = x^4 - 6x^2 + 3, \tag{2.34}$$

and so forth.

Since all moments of γ are finite (in fact, its moment generating function exists everywhere), each Hermite polynomial is in $L^2(\gamma)$. Before we show in Theorem 2.20 that the Hermite polynomials form a Hilbert basis for the space $L^2(\gamma)$, we first state and prove a lemma which demonstrates the beauty of these polynomials. This lemma will also be used to prove Theorem 2.20 and, eventually, the Gaussian Poincaré Inequality.

Lemma 2.15 [Gaussian Integration by Parts]. Let $f : \mathbb{R} \to \mathbb{R}$ be differentiable. Then for any $k \geq 1$, we have

$$\langle f, H_k \rangle_\gamma = \langle f', H_{k-1} \rangle_\gamma, \tag{2.35}$$

so long as both sides of (2.35) are defined and finite.

Remark 2.16. The case $k = 1$ is often referred to as *Stein's lemma*.

Proof. Let f be differentiable such that fH_k and $f'H_{k-1}$ are in $L^1(\gamma)$ for some $k \geq 1$. Applying integration by parts yields

$$\begin{aligned}
\langle f, H_k \rangle_\gamma &= \frac{(-1)^k}{\sqrt{2\pi}} \int f(x) \frac{d^k}{dx^k} e^{-x^2/2} \, dx && (2.36)\\
&= \frac{(-1)^k}{\sqrt{2\pi}} \left[f(x) \frac{d^k}{dx^k} e^{-x^2/2} \Big|_{-\infty}^{\infty} - \int f'(x) \frac{d^{k-1}}{dx^{k-1}} e^{-x^2/2} \, dx \right] && (2.37)\\
&= f(x) H_k(x) \frac{e^{-x^2/2}}{\sqrt{2\pi}} \Big|_{-\infty}^{\infty} + \langle f', H_{k-1} \rangle && (2.38)\\
&= \langle f', H_{k-1} \rangle_\gamma. && (2.39)
\end{aligned}$$

The final line (2.39) follows because $fH_k \in L^1(\gamma)$ implies $\lim_{x \to \pm\infty} f(x)H_k(x)e^{-x^2/2} = 0$, which completes the proof. $\square$

Remark 2.17. It is not true that the existence of the left-hand side is equivalent to that of the right-hand side in (2.35), so we really do need to assume both sides exist and are finite. Consider, for example, $f(x) = x \sin(e^{x^2})$ and $k = 1$.

Corollary 2.18. Let $k \geq 0$ and $f : \mathbb{R} \to \mathbb{R}$ k-times differentiable with $f^{(\ell)} H_{k-\ell} \in L^1(\gamma)$ for all $0 \leq \ell \leq k$. Then

$$\langle f, H_k \rangle_\gamma = \langle f^{(k)}, H_0 \rangle_\gamma = \int f^{(k)}(x) \, d\gamma(x). \tag{2.40}$$

Proof. This follows immediately from Lemma 2.15 by induction on k. $\square$

Remark 2.19. The condition $f^{(\ell)} H_{k-\ell} \in L^1(\gamma)$ for all $0 \leq \ell \leq k$ appears unwieldy, so we note a stronger, sufficient condition that is easier to verify. Namely, it is sufficient that $f^{(\ell)} \in L^2(\gamma)$ for all $0 \leq \ell \leq k$. For this, it is further sufficient that f is C^k and supported on a set of finite Lebesgue measure (e.g., compact support).

Theorem 2.20. Hermite polynomials form a Hilbert basis for $L^2(\gamma)$.

Proof. First we verify that $(H_k)_{k \geq 0}$ is orthogonal. If $0 \leq \ell < k$, then by Corollary 2.18,

$$\langle H_k, H_\ell \rangle_\gamma = \int H_\ell^{(k)}(x) \, d\gamma(x) = \int 0 \, d\gamma(x) = 0, \tag{2.41}$$

since H_ℓ is a polynomial of degree $\ell < k$. Note that the conditions of Corollary 2.18 are satisfied because all derivatives of H_ℓ, being polynomials, are in $L^2(\gamma)$ (cf. Remark 2.19).

Each H_k is, of course, nonzero, but $(H_k)_{k\geq 0}$ is not ortho*normal*. For this reason, it is useful to compute $\langle H_k, H_k\rangle_\gamma$. Accordingly, for any $k \geq 0$, note that

$$H_k^{(k)} \equiv k! \implies \langle H_k, H_k\rangle_\gamma = \int k!\, d\gamma = k!, \tag{2.42}$$

because H_k is a monic degree k polynomial.

It remains to prove that the span of $(H_k)_{k\geq 0}$ is dense in $L^2(\gamma)$. By Corollary 2.13,

$$\operatorname{span}(\{H_k \mid k \geq 0\}) = \mathbb{R}[x], \tag{2.43}$$

so we just need to show that $\mathbb{R}[x]$ is dense in $L^2(\gamma)$. By [Rud87, Theorem 4.18(i)], it suffices to prove that if $f \in L^2(\gamma)$ satisfies $\langle f, p\rangle_\mu = 0$ for all polynomials $p \in \mathbb{R}[x]$, then f is 0 almost everywhere $[\gamma]$. To that end, suppose $f \in \mathbb{R}[x]^\perp$, so that in particular $\langle f, x^n\rangle_\gamma = 0$ for all $n \geq 0$. There are many ways to reach the desired conclusion about f, the most common of which is via a digression on *Hermite functions*, but we opt instead for a "quick and dirty" proof using the Fourier Inversion Theorem.

Consider the complex function $F : \mathbb{C} \to \mathbb{C}$ defined by

$$F(z) := \int e^{zx} f(x)\, d\gamma(x) = \frac{1}{\sqrt{2\pi}} \int f(x) e^{-x^2/2+zx}\, dx. \tag{2.44}$$

Fixing $z \in \mathbb{C}$, note that for any nonnegative integer N, we have

$$\left|\sum_{n=0}^{N} \frac{(zx)^n}{n!} f(x) e^{-x^2/2}\right| \leq |f(x)| e^{-x^2/2} \sum_{n=0}^{N} \frac{|zx|^n}{n!} \leq |f(x)| e^{-x^2/2+|z||x|}. \tag{2.45}$$

The right-hand side of (2.45) is integrable, since by Cauchy-Schwarz

$$\langle |f|, e^{|z||x|}\rangle_\gamma \leq \|f\|_\gamma \|e^{|z||x|}\|_\gamma < \infty. \tag{2.46}$$

Therefore, by the dominated convergence theorem, we can swap a summation and integral to obtain

$$F(z) = \frac{1}{\sqrt{2\pi}} \int \sum_{n=0}^{\infty} \frac{z^n x^n}{n!} f(x)\, d\gamma(x) = \sum_{n=0}^{\infty} \frac{\langle f, x^n\rangle_\gamma}{n!} z^n \equiv 0. \tag{2.47}$$

This shows that F is a well-defined entire function on $\mathbb{C}$, and in fact it is identically 0. In particular, $F \equiv 0$ on the imaginary axis, which shows that the Fourier transform of $f(x)e^{-x^2/2}$ is identically 0. By a standard corollary of the Fourier Inversion Theorem [Rud87, Theorem 9.12], the function $f(x)e^{-x^2/2}$, and hence also f, must be 0 almost everywhere, with respect to the Lebesgue measure λ. Since $\gamma \ll \lambda$, $f = 0$ a.e. $[\gamma]$. $\square$

We now drop the assumption that $N = 1$ and generalize the Hermite polynomials to arbitrary dimensions N. We can then state and prove higher dimensional generalizations of Lemma 2.15 and Corollary 2.18. We will denote the set $\mathbb{Z}_{\geq 0}$ by $\mathbb{Z}_+$ and refer to vectors in $\mathbb{Z}_+^N$ as N-dimensional *multi-indices*.

Given a multi-index $\boldsymbol{k} = (k_1, k_2, \ldots, k_N) \in \mathbb{Z}_+^N$, we denote by $|\boldsymbol{k}| := k_1 + \cdots + k_N$ the *order* of $\boldsymbol{k}$, and by $\boldsymbol{k}! := k_1! \cdots k_N!$ the *factorial* of $\boldsymbol{k}$. If $f : \mathbb{R}^N \to \mathbb{R}$ is a function, the $\boldsymbol{k}$th *partial* of f is

$$\partial^{\boldsymbol{k}} f := \frac{\partial^{|\boldsymbol{k}|} f}{\partial x_1^{k_1} \cdots \partial x_N^{k_N}}. \tag{2.48}$$

Definition 2.21 [General Hermite Polynomial]. The $\boldsymbol{k}$th N-dimensional *Hermite* polynomial for a multi-index $\boldsymbol{k} = (k_1, k_2, \ldots, k_N) \in \mathbb{Z}_+^N$ is

$$H_{\boldsymbol{k}}(\boldsymbol{x}) := H_{k_1}(x_1) H_{k_2}(x_2) \cdots H_{k_N}(x_N). \tag{2.49}$$

Note that this is the tensor product $H_{\boldsymbol{k}} := \otimes_{1 \le i \le N} H_{k_i}$.

Remark 2.22. By Corollary 2.13, each $H_{\boldsymbol{k}}(\boldsymbol{x})$ is a monic degree $\boldsymbol{k}$ polynomial in n variables. By this we mean to say that the leading monomial in $H_{\boldsymbol{k}}(\boldsymbol{x})$ with respect to any monomial ordering on $\mathbb{R}[x_1, \ldots, x_N]$ is $\boldsymbol{x}^{\boldsymbol{k}} := x_1^{k_1} \cdots x_N^{k_N}$. By (2.42),

$$\langle H_{\boldsymbol{k}}, H_{\boldsymbol{k}} \rangle_{\gamma_N} = \boldsymbol{k}!. \tag{2.50}$$

Viewing $\mathbb{Z}_+^N$ as a subset of the vector space $\mathbb{R}^N$, we can add two multi-indices to obtain another multi-index, and if $\boldsymbol{k} \le \boldsymbol{\ell}$, meaning that $k_i \le \ell_i$ for all $1 \le i \le N$, then $\boldsymbol{\ell} - \boldsymbol{k}$ is also a multi-index. For $1 \le i \le N$, the ith standard basis vector $\boldsymbol{e}_i \in \mathbb{R}^N$ is a multi-index as well. Then $\partial^i f := \partial^{\boldsymbol{e}_i} f$ is the ith partial derivative of f. With this language, we are ready to state and prove the following multivariable generalization of Lemma 2.15.

Lemma 2.23 [Multivariable Gaussian Integration by Parts]. Let $f : \mathbb{R}^N \to \mathbb{R}$ be such that the ith partial $\partial^i f$ exists. Then for any $1 \le i \le N$ and $\boldsymbol{k} \ge \boldsymbol{e}_i$, we have

$$\langle f, H_{\boldsymbol{k}} \rangle_{\gamma_N} = \langle \partial^i f, H_{\boldsymbol{k} - \boldsymbol{e}_i} \rangle_{\gamma_N}, \tag{2.51}$$

so long as both sides of (2.51) are defined and finite.

Proof. Since $\gamma_N = \gamma^{\otimes N}$, integration with respect to γ_N can be expressed as N nested integrals with respect to γ, by Fubini's Theorem. Integrating $f H_{\boldsymbol{k}}$ first with respect to the ith coordinate and applying Lemma 2.15 yields (2.51). $\square$

Corollary 2.24. Let $\boldsymbol{k} \in \mathbb{Z}_+^N$, and suppose $f : \mathbb{R}^N \to \mathbb{R}$ has the property that $\partial^{\boldsymbol{\ell}} f$ exists and is in $L^2(\gamma_N)$ for all $\boldsymbol{\ell} \le \boldsymbol{k}$. Then

$$\langle f, H_{\boldsymbol{k}} \rangle_{\gamma_N} = \langle \partial^{\boldsymbol{k}} f, H_{\boldsymbol{0}} \rangle_{\gamma_N} = \int \partial^{\boldsymbol{k}} f(\boldsymbol{x})\, d\gamma_N(\boldsymbol{x}). \tag{2.52}$$

Remark 2.25. Of course, in view of Corollary 2.18 the stated condition on f is strictly stronger than necessary, but it is sufficient for our purposes. Often, we can approximate f by compactly supported smooth functions.

We are just about ready to prove Theorem 2.10 by verifying (2.27), although we will first need to establish that the N-dimensional Hermite polynomials form a Hilbert basis for $L^2(\gamma_N)$ in arbitrary dimensions N. Since $\gamma_N = \gamma^{\otimes N}$, this follows readily from (2.49), Theorem 2.20, and a standard indicator argument using Dynkin's π-λ theorem. We therefore declare the following theorem.

Theorem 2.26. For any positive integer N, the set of N-dimensional Hermite polynomials is a Hilbert basis for $L^2(\gamma_N)$.

In view of Theorem 2.26, we can express any $f \in L^2(\gamma_N)$ in terms of its *Hermite coefficients* $a_{\boldsymbol{k}} := \langle f, H_{\boldsymbol{k}} \rangle_{\gamma_N}$ as

$$f = \sum_{\boldsymbol{k} \in \mathbb{Z}_+^N} \frac{a_{\boldsymbol{k}}}{\boldsymbol{k}!} H_{\boldsymbol{k}}. \tag{2.53}$$

The $\boldsymbol{k}!$ factors are the correct normalization constants, by (2.50). If f is smooth and all of its partials lie in $L^2(\gamma_N)$, then $a_{\boldsymbol{k}}$ is simply the integral of $\partial^{\boldsymbol{k}} f$ with respect to γ_N, by Corollary 2.24. In this case, (2.27) is a result of Parseval's identity applied to both f and its first-order partials $\partial^i f$. Concretely,

$$\|f\|_{\gamma_N}^2 - \left(\int f \, d\gamma_N\right)^2 = \sum_{\boldsymbol{k} \in \mathbb{Z}_+^N - \{\boldsymbol{0}\}} \frac{(\int \partial^{\boldsymbol{k}} f \, d\gamma_N)^2}{\boldsymbol{k}!^2} \leq \sum_{i=1}^N \sum_{\boldsymbol{\ell} \in \mathbb{Z}_+^N} \frac{(\int \partial^{\boldsymbol{\ell}} \partial^i f \, d\gamma_N)^2}{(\boldsymbol{\ell} + \boldsymbol{e}_i)!^2}. \tag{2.54}$$

The inequality in (2.54) is *not* an equality because some of the higher partial derivative terms are overcounted. In subchapter 2.4, we will obtain a better bound by more carefully accounting for this overcounting. For proving Theorem 2.10, however, (2.54) is sufficient—by the trivial inequality $\boldsymbol{\ell}! \leq (\boldsymbol{\ell} + \boldsymbol{e}_i)!$ we get

$$\|f\|_{\gamma_N}^2 - \left(\int f \, d\gamma_N\right)^2 \leq \sum_{i=1}^N \sum_{\boldsymbol{\ell} \in \mathbb{Z}_+^N} \frac{(\int \partial^{\boldsymbol{\ell}} \partial^i f \, d\gamma_N)^2}{\boldsymbol{\ell}!^2} = \sum_{i=1}^N \|\partial^i f\|_{\gamma_N}^2. \tag{2.55}$$

Note that the right-hand side of (2.55) is simply $\int \|\nabla f\|^2 \, d\gamma_N$, yielding (2.27).

For general $f \in C^1(\mathbb{R}^N) \cap L^2(\gamma_N)$, we can deduce (2.27) by approximating f with suitable compactly supported smooth functions. There are a few ways to do this. An elegant, modern approach would be via the Ornstein-Uhlenbeck semigroup, which is the framework Chatterjee opts for [Cha14, Chapter 2.2]. However, one can also accomplish this with basic tools of functional analysis, such as bump functions, mollifiers, and convolution. We take the latter approach. First, note that both (2.25) and (2.27) are trivial if the first-order partials of f are not in $L^2(\gamma_N)$, so we may assume that $\partial^i f \in L^2(\gamma_N)$ for all

$1 \le i \le N$. The idea is to first reduce to the case where f is smooth via convolution with smooth mollifiers, and then further reduce to the case where f has compact support via an appropriate cutoff function argument. We leave the details to the Appendix (B) and consider Theorem 2.10 proved.

Finally, we complete this subchapter with a proof of (1.14). Recall that the N-dimensional Gaussian SK model at inverse temperature β has Hamiltonian $\mathcal{H}_N(\sigma)$ given by (1.11). The disorder is controlled by an upper triangular array of $N(N-1)/2$ i.i.d. standard Gaussians $G = (g_{ij})_{1\le i<j\le N}$, in which case we can view $F_{N,\beta}$ as a measurable function

$$F_{N,\beta} : \mathbb{R}^{\frac{N(N-1)}{2}} \to \mathbb{R}. \tag{2.56}$$

Let $p_{N,\beta}$ and $\mathcal{Z}_{N,\beta}$ be the associated Gibbs measure and partition function, as defined in Definition 1.2. For a function $f(\boldsymbol{\sigma}, G)$, we use the notation

$$\langle f\rangle := \mathbb{E}_{p_{N,\beta}}[f(\sigma, G)|G] = \frac{\sum_{\boldsymbol{\sigma}\in\{\pm1\}^N} f(\boldsymbol{\sigma}, G)e^{\beta\mathcal{H}_N(\boldsymbol{\sigma})}}{\sum_{\boldsymbol{\sigma}\in\{\pm1\}^N} e^{\beta\mathcal{H}_N(\sigma)}}. \tag{2.57}$$

In other words, $\langle\cdot\rangle$ denotes the average with respect to $\boldsymbol{\sigma} \sim p_N$, conditional on the disorder. We call this operation the *Gibbs average*.

Lemma 2.27. Let $F_{N,\beta}$ be as in (2.56), with input variables denoted $(g_{ij})_{i<j}$ (in this context, the g_{ij} are *not* random variables—we define the free energy by plugging i.i.d. Gaussians into these input slots). Then $F_{N,\beta}$ is C^1 with partials

$$\frac{\partial F_{N,\beta}}{\partial g_{ij}} = \frac{1}{\sqrt{N}}\langle\sigma_i\sigma_j\rangle. \tag{2.58}$$

Proof. This is a standard computation. We have

$$\frac{\partial F_{N,\beta}}{\partial g_{ij}} = \frac{1}{\beta\mathcal{Z}_{N,\beta}}\sum_{\boldsymbol{\sigma}\in\{\pm1\}^N}\frac{\partial}{\partial g_{ij}}e^{\beta\mathcal{H}_N(\boldsymbol{\sigma})} = \frac{1}{\sqrt{N}\mathcal{Z}_{N,\beta}}\sum_{\boldsymbol{\sigma}\in\{\pm1\}^N}\sigma_i\sigma_j e^{\beta\mathcal{H}_N(\boldsymbol{\sigma})}. \tag{2.59}$$

We simply notice that the right-hand side of (2.59) is $\frac{1}{\sqrt{N}}\langle\sigma_i\sigma_j\rangle$. □

By Lemma 2.27, the partials of $F_{N,\beta}$ must be bounded in absolute value by $1/\sqrt{N}$, since $\sigma_i\sigma_j$ is bounded in absolute value by 1. Combining this with the mean value inequality, we can show $\mathrm{Var}(F_{N,\beta}) < \infty$, and then by Theorem 2.10 we get

$$\mathrm{Var}(F_{N,\beta}) \le \frac{N(N-1)}{2N} = \frac{N-1}{2} < \frac{1}{2}N. \tag{2.60}$$

This implies (1.14) with the constant $C_\beta = 1/2$ for all $\beta > 0$. From here, there are two natural directions in which to head. On the one hand, from a simple variance bound like the Gaussian Poincaré inequality, one can develop useful exponential tail bounds [BLM13, Theorem 5.6]. On the other hand, we demonstrated in Chapter 1 that (1.14) is a loose bound, a perfect segue into *superconcentration*.

2.3 An Introduction to Superconcentration

We finally arrive at Chatterjee's recent theory of *superconcentration*, as detailed in his monograph [Cha14]. Roughly, if the order of fluctuations of a stochastic process is less than the upper bounded predicted by classical concentration of measure, we say that the process *superconcentrates*. A more precise definition, using the unifying framework of *Markov semigroups*, can be found in [Cha14, Definition 3.1], but we will not need this level of abstraction. It's worth noting, however, that much of the next subchapter could be phrased in that language, specifically in terms of the famous *Ornstein-Uhlenbeck* (OU) semigroup. Even the Efron-Stein inequality can be viewed as a result concerning the *independent flips* semigroup [Cha14, Theorem 7.1]. Regardless, Chatterjee's monograph is not an introduction to semigroup analysis, but rather is dedicated to illustrating several examples of superconcentration as well as cataloging techniques for proving such results. Since the next subchapter is dedicated to one long, detailed proof showcasing one of these techniques, we will omit proofs in this subchapter. We aim to demonstrate by two diverse examples that superconcentration is not a phenomenon unique to spin glasses.

We begin, as Chatterjee does, with a discussion of *First-Passage Percolation* (FPP). The general setup is as follows. Let $\Gamma = (V, E)$ be a graph, and associate to each edge $e \in E$ a nonnegative random variable ω_e representing the time it takes for a particle to travel along the edge e. Given a path π on Γ, i.e., a collection of consecutive incident edges $e_1, \ldots, e_n$ in E, we define the *passage time* along π, denoted $T(\pi)$, as the sum of all passage times ω_{e_k} for each edge in π. Then the *first-passage time* between two vertices $v, w \in V$ is the minimum passage time $T(v, w)$ among all paths π from v to w. Denoting the set of such paths as $\Pi(v, w)$, this means

$$T(v, w) := \inf_{\pi \in \Pi(v,w)} T(\pi). \tag{2.61}$$

See Figure 10 for a visualization of FPP on a finite graph; the fastest path is colored blue.

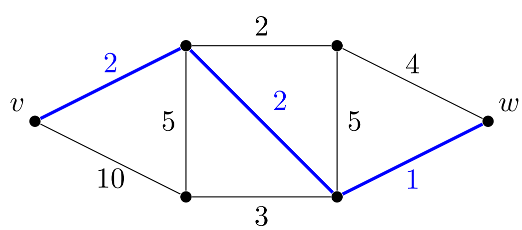

Figure 10: An example of FPP with $T(v, w) = 5$.

The most common instance of FPP, introduced in 1965 by Hammersley and Welsh [HW65], occurs when Γ is a full d-dimensional lattice $\mathbb{Z}^d$, with all neighboring sites connected by an edge. We also assume that $(\omega_e)_{e \in E}$ are i.i.d. for simplicity. Given a point $x \in \mathbb{R}^d$ and an integer $n \in \mathbb{Z}$, we write $[nx]$ for the site $i \in \mathbb{Z}^d$ which minimizes the Euclidean distance $\|i - nx\|$. If there are multiple such sites, we choose the one which is lexicographically minimal. Then if $T_n(x) := T(0, [nx])$, it turns out that the limit $T_n(x)/n$ as $n \to \infty$ exists and is deterministic. In particular, $T_n(x) = \Theta(n)$. A proof using Kingman's subadditive ergodic theorem can be found in [Kes86].

Set $T_n = T_n(e_1)$, where e_1 is the unit vector $(1, 0, \ldots, 0) \in \mathbb{R}^d$. Since 1993, it has been known that $\mathrm{Var}(T_n) = O(n)$ [Kes93]. In other words, we have a standard concentration phenomenon, where T_n is roughly on the order of n with fluctuations on the order of $\sqrt{n}$. However, in certain cases, we have superconcentration $\mathrm{Var}(T_n) = o(n)$. One instance of this is the Benjamini-Kalai-Schramm (BKS) theorem [BKS03].

Theorem 2.28 [BKS]. Suppose that the edge weights ω_e are i.i.d. Bernoulli, and that $d \geq 2$. Then there exists a constant $C > 0$, depending on $\mathbb{P}(\omega_e = 1)$ and d but *not* on n, such that

$$\mathrm{Var}(T_n) \leq \frac{Cn}{\log n}. \tag{2.62}$$

Remark 2.29. Chatterjee proves a nice variant of the BKS theorem in Chapter 5 of his superconcentration monograph [Cha14, Theorem 1.1].

Discussion of superconcentration in the context of *Gaussian polymer models* and *extremal Gaussian fields* can also be found in [Cha14]. Rather than detail these examples, we refer the interested reader to Chatterjee's exposition and instead introduce another instance of the superconcentration phenomenon—in fact, an entire class of seemingly unrelated instances, the *Kardar-Parisi-Zhang* (KPZ) class. What unites all of these models is that, while the classical theory predicts fluctuations on the order $O(n^{1/2})$, in reality fluctuations are (provably or by simulation) on the order $n^{1/3}$. We focus here on one model from this proposed "universality" class, which comes from RMT.

Recall from subchapter 1.3 the distribution $\mathrm{GOE}(n)$. In the $n \to \infty$ limit, we saw that the spectra of such matrices follow the semicircle law, but what if we specialize to only the largest eigenvalue? Intuitively, this behavior is somewhat akin to an SK model at 0 temperature. It is not difficult to show that the top eigenvalue is a Lipschitz function of the matrix, which we can leverage with the Gaussian Poincaré inequality to prove a variance bound $O(n)$. But one can do much more. Amazingly, the entire limiting distribution, along with a CLT towards this law, can be determined by careful analysis. This limiting CDF is often denoted F_1, belonging to the *Tracy-Widom* (TW) family. If $A_n \sim \mathrm{GOE}(n)$, and $\lambda_{\max}(A)$ denotes the top eigenvalue of a matrix A, then we have the CLT

$$\lim_{n\to\infty} \mathbb{P}(n^{2/3}(\lambda_{\max}(A_n) - 2) \leq s) = F_1(s) \quad \text{for all } s \in \mathbb{R}. \tag{2.63}$$

From this CLT we glean that $\lambda_{\max}(A_n)$ concentrates around 2, as predicted by Wigner's semicircular law, and that the fluctuations of the top eigenvalue behave like $n^{-2/3}$. Under appropriate renormalization, considering nA_n rather than A_n so that the expectation of its top eigenvalue is $\Theta(n)$ rather than $\Theta(1)$, these fluctuations are $n^{1/3}$. Unfortunately, a closed formula for F_1 is unavailable, but there are many nice functional expressions relating F_1 to the solution of a clean second-order differential equation, the Painlevé II ODE [TW09]. In Figure 11 below, we attempt to illustrate the CLT (2.63).

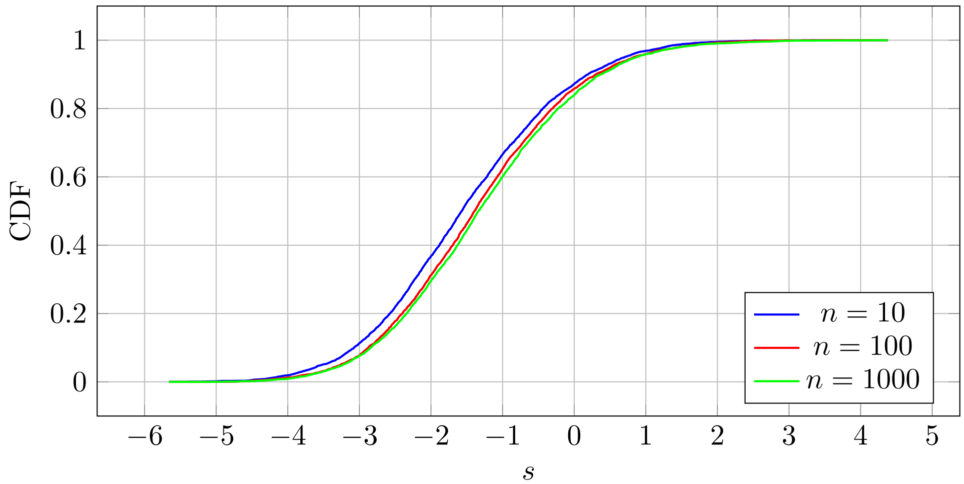

Figure 11: Empirical CDFs for $n^{2/3}(\lambda_{\max}(A_n) - 2)$ where $n = 10, 100, 1000$, which converge to the TW distribution F_1.

2.4 The Spectral Method and the Gaussian SK Model

The goal of this final subchapter is to demonstrate one of Chatterjee's techniques for deriving superconcentration results, which he refers to as the *spectral method*, by proving superconcentration for the Gaussian SK Model, as promised in subchapter 1.2. The main idea of the "spectral method" is to obtain improved Poincaré inequalities that, in cases of sufficient symmetry, lead to better order bounds than the traditional theory provides. In our case, we observe a basic $(\mathbb{Z}/2)$-symmetry in the Gaussian SK model, which shows that several Hermite coefficients become 0 (a sort of "evenness" reduction, to be made precise). Combined with an "improved" Gaussian Poincaré inequality that allows us to leverage this special extra knowledge about the Hermite coefficients, at least when they are of sufficiently small order, and better bound the higher order terms, we are led to superconcentration. We begin with a rephrasing of [Cha14, Equation 6.3].

Theorem 2.30 [Improved Gaussian Poincaré Inequality]. Let N and M be positive integers, and suppose $f \in C^M(\mathbb{R}^N)$ is such that $\partial^{\boldsymbol{k}} f \in L^2(\gamma_N)$ for all $|\boldsymbol{k}| \leq M$. Then

$$\mathrm{Var}_{\gamma_N}(f) \leq \frac{\mathbb{E}_{\gamma_N}[\|\nabla f\|^2]}{M} + \sum_{0<|\boldsymbol{k}|<M} \frac{(\mathbb{E}_{\gamma_N}[\partial^{\boldsymbol{k}} f])^2}{\boldsymbol{k}!^2}. \tag{2.64}$$

Proof. As in the proof of the standard Gaussian Poincaré Inequality, it suffices to prove that (2.64) holds for any $f \in C_c^\infty(\mathbb{R}^N)$, and then we can extend to more general f by a standard approximation argument. Thus, we restrict ourselves to the compactly supported smooth case and once again relegate the general case to the Appendix (B).

Let $f \in C_c^\infty(\mathbb{R}^N)$. First, we apply Parseval's identity to write

$$\operatorname{Var}_{\gamma_N}(f) = \sum_{0<|\boldsymbol{k}|<M} \frac{(\mathbb{E}_{\gamma_N}[\partial^{\boldsymbol{k}} f])^2}{\boldsymbol{k}!^2} + \sum_{|\boldsymbol{k}|\geq M} \frac{(\mathbb{E}_{\gamma_N}[\partial^{\boldsymbol{k}} f])^2}{\boldsymbol{k}!^2}, \tag{2.65}$$

so that it suffices to bound

$$\sum_{|\boldsymbol{k}|\geq M} \frac{(\mathbb{E}_{\gamma_N}[\partial^{\boldsymbol{k}} f])^2}{\boldsymbol{k}!^2} \leq \frac{\mathbb{E}_{\gamma_N}[\|\nabla f\|^2]}{M}. \tag{2.66}$$

The idea behind (2.66) is similar to (2.55), except this time we are more conservative about overcounting. Let $|\boldsymbol{k}| \geq M$. Then we can rewrite $\partial^{\boldsymbol{k}} f$ as $\partial^{\ell}\partial^i f$ for a multi-index ℓ in $n_{\boldsymbol{k}}$ distinct ways, where $n_{\boldsymbol{k}}$ is the number of nonzero entries in $\boldsymbol{k}$. If we let $\boldsymbol{n}_{\boldsymbol{k}}$ denote the set of indices $1 \leq i \leq N$ for which $k_i \neq 0$ (so that $|\boldsymbol{n}_{\boldsymbol{k}}| = n_{\boldsymbol{k}}$), then

$$\frac{M}{\boldsymbol{k}!^2} = \frac{\sum_{i\in \boldsymbol{n}_{\boldsymbol{k}}} k_i}{\prod_{j\in \boldsymbol{n}_{\boldsymbol{k}}} k_j^2} \leq \sum_{i\in \boldsymbol{n}_{\boldsymbol{k}}} \frac{k_i^2}{\prod_{j\in \boldsymbol{n}_{\boldsymbol{k}}} k_j^2} = \sum_{i\in \boldsymbol{n}_{\boldsymbol{k}}} \frac{1}{(\boldsymbol{k}-\boldsymbol{e}_i)!^2}. \tag{2.67}$$

Dividing by M and applying to (2.66), we conclude

$$\sum_{|\boldsymbol{k}|\geq M} \frac{(\mathbb{E}_{\gamma_{E_N}}[\partial^{\boldsymbol{k}} f])^2}{\boldsymbol{k}!^2} \leq \frac{1}{M}\sum_{i=1}^{N} \sum_{|\boldsymbol{\ell}|\geq M-1} \frac{(\mathbb{E}_{\gamma_{E_N}}[\partial^{\boldsymbol{\ell}}\partial^i f])^2}{\boldsymbol{\ell}!^2} \leq \frac{\mathbb{E}_{\gamma_{E_N}}[\|\nabla f\|^2]}{M}. \tag{2.68}$$

This completes the proof. □

As a result of Theorem 2.30 and the bound $\|\nabla F_{N,\beta}\|^2 < N/2$ obtained from Lemma 2.27, for any positive integer M we can write

$$\operatorname{Var}_{\gamma_{E_N}}(F_{N,\beta}) < \sum_{0<|\boldsymbol{k}|<M} \frac{(\mathbb{E}_{\gamma_{E_N}}[\partial^{\boldsymbol{k}} F_{N,\beta}])^2}{\boldsymbol{k}!^2} + \frac{N}{2M}. \tag{2.69}$$

In order to obtain superconcentration, we aim to show

$$\sum_{0<|\boldsymbol{k}|<M} \frac{(\mathbb{E}_{\gamma_{E_N}}[\partial^{\boldsymbol{k}} F_{N,\beta}])^2}{\boldsymbol{k}!^2} = O_M(1). \tag{2.70}$$

Then, we can establish $\operatorname{Var}_{\gamma_{E_N}}(F_{N,\beta})/N \to 0$ by first taking $M \to \infty$ and then $N \to \infty$.

Given a positive integer N, let K_N denote the complete graph with N vertices. Identify the vertex set as $[N] := \{1, 2, \ldots, N\}$ and the edge set as $E_N := \{(i,j) \in [N]^2 \mid i < j\}$. A *weighted subgraph* of K_N is a nonzero integer weighting on each of the edges, or equivalently a multi-index $\boldsymbol{k} \in \mathbb{Z}_+^{E_N}$. The *total weight* of $\boldsymbol{k}$ is the sum of all weights, i.e., $|\boldsymbol{k}|$. The *degree* of a vertex $i \in [N]$, denoted $\deg_{\boldsymbol{k}}(i)$ or simply $\deg(i)$ if $\boldsymbol{k}$ is understood, is the sum of all weights among edges incident to i. Notice that

$$\sum_{i=1}^{N} \deg_{\boldsymbol{k}}(i) = 2|\boldsymbol{k}|. \tag{2.71}$$

A weighted subgraph is called *even* if the degree of every vertex is even. Otherwise, it is called *odd*. We have the following combinatorial lemma.

Lemma 2.31. Let M be a positive integer. Then there exists a positive constant C_M such that, for any positive integer N, there are at most $C_M N^M$ even weighted subgraphs of K_N with total weight M.

Proof. Suppose that $\boldsymbol{k}$ is an even weighted subgraph of K_N such that $|\boldsymbol{k}| = M$. Let p be the total number of vertices $i \in [N]$ with $\deg_{\boldsymbol{k}}(i) \neq 0$. For all such i, we must have $\deg(i) \geq 2$ by the evenness of $\boldsymbol{k}$. Hence, it follows from (2.71) that $2p \leq 2M$, so $p \leq M$. In particular, if $M \leq N$ then $\boldsymbol{k}$ is an embedding of a weighted subgraph of K_M with total weight M into K_N. The number of distinct graph embeddings $K_M \hookrightarrow K_N$ is

$$|\mathrm{Hom}(K_M, K_N)| = \binom{N}{M} \leq N^M. \tag{2.72}$$

Letting C_M be the number of weighted subgraphs of K_M with total weight M, which is a constant independent of N, it follows that there are at most $C_M N^M$ even weighted subgraphs of K_N with total weight M, whenever $M \leq N$.

If $M > N$, then any weighted subgraph of K_N with total weight M can be mapped to a unique weighted subgraph of K_M, also with total weight M, via any fixed embedding $K_N \hookrightarrow K_M$. Therefore the number of even $\boldsymbol{k} \in \mathbb{Z}_+^{E_N}$ with $|\boldsymbol{k}| = M$ is bounded above by C_M. Since $1 \leq N^M$, the upper bound $C_M N^M$ is valid for all N. $\square$

Remark 2.32. As shown in the proof of Lemma 2.31, we can take C_M to be the total number of $\boldsymbol{k} \in \mathbb{Z}_+^{E_M}$ such that $|\boldsymbol{k}| = M$. By a standard combinatorial argument,

$$C_M = \binom{M + \binom{M}{2} - 1}{M} = O(M^{2M}). \tag{2.73}$$

We glean from Lemma 2.31 that among all weighted subgraphs of K_N with fixed total weight M, the portion of even subgraphs is small. Indeed, the number of subgraphs with total weight M is

$$\binom{M + \binom{N}{2} - 1}{M} = O(N^{2M}), \tag{2.74}$$

which is of a larger order than $O(N^M)$. A key indication of superconcentration is the following, which shows that a majority of the Hermite coefficients of $F_{N,\beta}$ are in fact 0.

Lemma 2.33. Let N be a positive integer, $\beta > 0$, and $\boldsymbol{k}$ an odd weighted subgraph of K_N. Then

$$\mathbb{E}_{\gamma_{E_N}}[\partial^{\boldsymbol{k}} F_{N,\beta}] = 0. \tag{2.75}$$

Proof. The crux of this proof is noticing and leveraging a latent symmetry in the model. For each $\boldsymbol{\alpha} \in \{\pm 1\}^N$, define the *gauge transformation*

$$T_{\boldsymbol{\alpha}} : \mathbb{R}^{\frac{N(N-1)}{2}} \to \mathbb{R}^{\frac{N(N-1)}{2}}, \quad (g_{ij})_{i<j} \mapsto (\alpha_i \alpha_j g_{ij})_{i<j}. \tag{2.76}$$

By the symmetry of γ, if $G = (g_{ij})_{1 \le i < j \le N}$ is the disorder of the N-dimensional Gaussian SK model, i.e., an upper triangular array of i.i.d. standard Gaussians, then

$$T_{\boldsymbol{\alpha}}(G) \stackrel{d}{=} G \quad \text{for all } \boldsymbol{\alpha} \in \{\pm 1\}^N. \tag{2.77}$$

Moreover, $F_{N,\beta} = F_{N,\beta} \circ T_{\boldsymbol{\alpha}}$ for all $\boldsymbol{\alpha}$, as can be seen by reindexing the summation in the partition function according to $\boldsymbol{\sigma} \mapsto \boldsymbol{\sigma} \odot \boldsymbol{\alpha}$, where $(\boldsymbol{\sigma} \odot \boldsymbol{\alpha})_i := \sigma_i \alpha_i$.

For each $1 \le i \le N$, let $\boldsymbol{\alpha}^{(i)} \in \{\pm 1\}^N$ be the *flip* which is -1 in the ith coordinate and $+1$ elsewhere. Then by the chain rule and a simple counting argument, if $\boldsymbol{k} \in \mathbb{Z}_+^{E_N}$ is such that $\deg_{\boldsymbol{k}}(i)$ is odd, then

$$\partial^{\boldsymbol{k}} F_{N,\beta} = \partial^{\boldsymbol{k}}(F_{N,\beta} \circ T_{\boldsymbol{\alpha}^{(i)}}) = -(\partial^{\boldsymbol{k}} F_{N,\beta}) \circ T_{\boldsymbol{\alpha}^{(i)}}. \tag{2.78}$$

Combining (2.77) and (2.78), we deduce

$$\mathbb{E}_{\gamma_{E_N}}[\partial^{\boldsymbol{k}} F_{N,\beta}] = \mathbb{E}_{\gamma_{E_N}}[(\partial^{\boldsymbol{k}} F_{N,\beta}) \circ T_{\boldsymbol{\alpha}^{(i)}}] = -\mathbb{E}[\partial^{\boldsymbol{k}} F_{N,\beta}], \tag{2.79}$$

which implies (2.75). $\square$

In order to deduce superconcentration from (2.69), we still need to bound the nonzero Hermite coefficients. We do this by naively bounding the partials $\partial^{\boldsymbol{k}} F_{N,\beta}$ for $|\boldsymbol{k}| < M$. The key tool is the following lemma.

Lemma 2.34. Let N be a positive integer and recall from (2.57) the Gibbs average notation $\langle \cdot \rangle$ induced by the disorder of the N-dimensional Gaussian SK model at inverse temperature β. Let $f : \{\pm 1\}^N \to \mathbb{R}$. Then for all $1 \le i < j \le N$,

$$\frac{\partial}{\partial g_{ij}} \langle f(\boldsymbol{\sigma}) \rangle = \frac{\beta}{\sqrt{N}} \Big(\langle \sigma_i \sigma_j f(\boldsymbol{\sigma}) \rangle - \langle \sigma_i \sigma_j \rangle \langle f(\boldsymbol{\sigma}) \rangle \Big). \tag{2.80}$$

Proof. Throughout, all variables $\boldsymbol{\sigma}$ are assumed to index over the hypercube $\{\pm 1\}^N$ unless stated otherwise. By the quotient rule, the desired derivative may be written

$$\frac{\mathcal{Z}_{N,\beta} \sum_{\boldsymbol{\sigma}} \frac{\beta}{\sqrt{N}} \sigma_i \sigma_j f(\boldsymbol{\sigma}) e^{\beta \mathcal{H}_N(\boldsymbol{\sigma})}}{\mathcal{Z}_{N,\beta}^2} - \left(\frac{\sum_{\boldsymbol{\sigma}} f(\boldsymbol{\sigma}) e^{\beta \mathcal{H}_N(\boldsymbol{\sigma})}}{\mathcal{Z}_{N,\beta}} \right) \left(\frac{\partial \log \mathcal{Z}_{N,\beta}}{\partial g_{ij}} \right), \tag{2.81}$$

where have used that $d(\log g(x))/dx = g(x)^{-1} \cdot dg/dx$ for any positive function g to rewrite the second summand. Canceling the $\mathcal{Z}_{N,\beta}$ terms, the first summand in (2.81) is easily seen to be $(\beta/\sqrt{N}) \langle \sigma_i \sigma_j f(\boldsymbol{\sigma}) \rangle$. The second summand is a product of two terms, the first of which is $\langle f(\boldsymbol{\sigma}) \rangle$. The other term in this product is $(\beta/\sqrt{N}) \langle \sigma_i \sigma_j \rangle$ by recalling that $F_{N,\beta} = \frac{1}{\beta} \log \mathcal{Z}_{N,\beta}$ and applying Lemma 2.27. Altogether, we conclude (2.80) from (2.81). $\square$

Equipped with Lemma 2.34, we can obtain the desired bounds on the partials of $F_{N,\beta}$, encoded in the following result.

Lemma 2.35. There exists a sequence of positive numbers $b_1, b_2, b_3, \ldots$ such that for all positive integers N, nonzero multi-indices $\boldsymbol{k} \in \mathbb{Z}_+^N$, and $\beta > 0$, we have

$$|\partial^{\boldsymbol{k}} F_{N,\beta}| \leq b_{|\boldsymbol{k}|} \frac{\beta^{|\boldsymbol{k}|-1}}{N^{|\boldsymbol{k}|/2}}. \tag{2.82}$$

Proof. Let ℓ^∞ denote the infinite-dimensional real vector space of functions $s : \mathbb{Z}_{>0} \to \mathbb{R}$ with finite support, i.e., finite real-valued sequences. We call $s \in \ell^\infty$ *natural* if its image is a subset of $\mathbb{Z}_{\geq 0}$. Consider the linear endomorphism $L : \ell^\infty \to \ell^\infty$ defined by

$$L(f)(n) := nf(n) + (n-1)f(n-1), \tag{2.83}$$

where the second summand is omitted if $n = 1$. Note that L maps natural sequences to natural sequences. Given a natural sequence $s \in \ell^\infty$, we say that a function F is of *type* s if for each $n \in \mathbb{Z}_{>0}$, there exist $s(n)$ functions $f_{n,s(1)}, \ldots, f_{n,s(n)}$ defined on the hypercube $\{\pm 1\}^N$ and taking values in $[-1, 1]$ such that

$$F = \sum_{n=1}^{\infty} \prod_{m=1}^{s(n)} \langle f_{n,m}(\boldsymbol{\sigma}) \rangle. \tag{2.84}$$

Note that the sum in (2.84) is actually finite, since s has finite support. It follows readily from the product rule, induction, and Lemma 2.34 that if F is a function of type $s \in \ell^\infty$, for any $1 \leq i < j \leq N$ the partial $\partial F / \partial g_{i,j}$ is $\beta/\sqrt{N}$ times a function of type $L(s)$.

Let s_1 be the natural sequence such that $s_1(1) = 1$ and $s_1(n) = 0$ for all $n > 1$. Then the partials of $F_{N,\beta}$ can be written as $1/\sqrt{N}$ times a function of type s_1, by Lemma 2.27. Letting $s_k := L^{\circ(k-1)}(s_1)$, so that $s_2 = L(s_1)$, $s_3 = L(s_2)$, etc., it follows by induction that for any nonzero multi-index $\boldsymbol{k}$, we can write

$$\partial^{\boldsymbol{k}} F_{N,\beta} = \frac{\beta^{|\boldsymbol{k}|-1}}{N^{|\boldsymbol{k}|/2}} G_{N,\beta,\boldsymbol{k}}, \tag{2.85}$$

where $G_{N,\beta,\boldsymbol{k}}$ is a function of type $s_{|\boldsymbol{k}|}$. Then if we define the sequence $b_k := \sum_{n \geq 1} s_k(n) < \infty$ for all positive integers k, we conclude (2.82) from (2.85). $\square$

Remark 2.36. The first several values of b_n are $b_1 = 1$, $b_2 = 2$, $b_3 = 6$, $b_4 = 26$, $b_5 = 150$, and $b_6 = 1082$. See OEIS A000629 for a connection with a famous necklace counting problem. The sequence b_n grows *rapidly*, but for our purposes, all that matters is that the sequence does not depend on N.

We are finally in the homestretch. Combining Lemma 2.31, Lemma 2.33, Lemma 2.35, and the basic bound $\boldsymbol{k}! \le |\boldsymbol{k}|!$, we conclude that for any positive integer M,

$$\sum_{|\boldsymbol{k}|=M} \frac{(\mathbb{E}_{\gamma_{E_N}}[\partial^{\boldsymbol{k}} F_{N,\beta}])^2}{\boldsymbol{k}!^2} \le C_M N^M \cdot \frac{b_M^2 \beta^{2M-2}}{N^M M!^2} = \frac{C_M b_M^2 \beta^{2M-2}}{M!^2}. \tag{2.86}$$

Let $A_{1,\beta} = 0$ and $A_{M,\beta} = \sum_{m=1}^{M-1} C_m b_m^2 \beta^{2m-2}/m!^2$ for $M \in \mathbb{Z}_{>1}$. Then by (2.86),

$$\sum_{0<|\boldsymbol{k}|<M} \frac{(\mathbb{E}_{\gamma_{E_N}}[\partial^{\boldsymbol{k}} F_{N,\beta}])^2}{\boldsymbol{k}!^2} \le A_{M,\beta} = O_M(1). \tag{2.87}$$

Dividing both sides of (2.69) by N and taking $M \to \infty$ followed by $N \to \infty$, we conclude that $\mathrm{Var}_{\gamma_N}(F_{N,\beta}) = o_\beta(N)$.

More precisely, combining (2.87) with (2.69) yields

$$\frac{\mathrm{Var}_{\gamma_{E_N}}(F_{N,\beta})}{N} < \frac{A_{M,\beta}}{N} + \frac{1}{2M} \tag{2.88}$$

for all positive integers M. Thus, if $\varepsilon > 0$, then fixing any positive integer $M > 1/\varepsilon$, for all $N \ge 2A_{M,\beta}/\varepsilon$ we have $\mathrm{Var}(F_{N,\beta})/N < \varepsilon$. This proves Corollary 1.9. In order to obtain the slightly stronger Theorem 1.8, one can more carefully choose M in the above argument by paying closer attention to the growth rate of $A_{M,\beta}$.

3 Matrix Coefficients and the Peter-Weyl Theorem

As we have seen, one of the key tools in proving superconcentration for the Gaussian SK model was a convenient choice of Hilbert basis for $L^2(\gamma_N)$—the Hermite polynomials. Given a compact group G equipped with its standard Haar probability measure μ, the most natural Hilbert basis for $L^2(\mu)$ is a set of *matrix coefficients*, made precise by the famous *Peter-Weyl Theorem*. The goal of this chapter, motivated by the study of G-invariant SK models, is to rigorously develop this general setup for compact groups. We assume basic familiarity with Lie groups and smooth manifolds, at the level of [Lee13]. Some knowledge of representation theory is also important, although we will do our best to redevelop this theory from scratch, at least initially. [FH91] and [Kna86] are suggested references.

3.1 Construction of the Haar Measure

Let G be a group. Denote the multiplication map $(g, h) \mapsto gh$ by $m : G \times G \to G$ and the inversion map $g \mapsto g^{-1}$ by $i : G \to G$. If G is given a topology making m and i continuous (where $G \times G$ is endowed with the product topology), then we say that G is a *topological* group. If G is also given a smooth structure making m and i smooth, it is a *Lie* group. In other words, a topological group is a group object in the category of topological spaces, while a Lie group is a group object in the category of smooth manifolds. We assume that all smooth manifolds are Hausdorff and second-countable.

Every suitably nice topological group comes equipped with a natural measure, the *Haar* measure. In particular, every locally compact Hausdorff group (including all Lie groups) has both a left-invariant and right-invariant Haar measure, unique up to a scalar multiple. For example, the Lebesgue measure on $(\mathbb{R}^n, +)$ is both a left and right Haar measure. In general, the left and right Haar measures need not coincide, but when they do we call the group *unimodular*. Importantly, all compact Hausdorff groups are unimodular. As a convention, by *compact group* we mean a topological group that is *both* compact and Hausdorff. In this subchapter, we will carefully construct the Haar measure on an arbitrary compact group. In the special case that G is a compact Lie group, one can do this by choosing an invariant density. However, there is a nice story in the more general setting of compact groups, based loosely on the law of large numbers, that we instead opt to detail. Much of our exposition takes inspiration from lecture notes due to [Mil].

Fix a topological group G. A Borel measure μ on G is said to be *right-invariant* (resp. *left-invariant*) if right (resp. left) multiplication by any element of G preserves μ. If μ is both right and left-invariant, it is called *bi-invariant*. A Borel measure μ is a *right* (resp. *left*) *Haar measure* if it is right (resp. left)-invariant, *locally finite* (finite on compacts), *outer regular* ($\mu(B) = \inf\{\mu(U) \mid B \subseteq U \text{ open}\}$ for Borel B), and *tight/inner regular* ($\mu(U) = \sup\{\mu(K) \mid U \supseteq K \text{ compact}\}$ for open U). We refer to a Borel measure satisfying the latter three regularity conditions as *Riesz*. On a Polish space, such as a Lie group, every finite Borel measure is Riesz [Gaa, Chapter 2]. First-countable compact groups are always Polish by the Birkhoff-Kakutani theorem [MZ55, Section 1.22]. It's clear that if μ is a right (resp. left) Haar measure, so is $c\mu$ for $c \geq 0$. We aim to prove:

Theorem 3.1 [Haar's Theorem (Compact)]. Let G be a compact group. Then there exists a unique right and a unique left Haar measure on G normalized so that $\mu(G) = 1$. Furthermore, these two measures agree. One calls this common bi-invariant measure the *Haar probability measure* on G.

Fix a compact group G, and let $C(G)$ denote the Banach space of continuous functions $G \to \mathbb{C}$ equipped with the supremum norm $\|\varphi\| := \sup_{g\in G} |\varphi(g)|$. Note that G acts on $C(G)$ both on the left and on the right by defining

$$(g \cdot \varphi)(h) := \varphi(g^{-1}h) \quad \text{and} \quad (\varphi \cdot g)(h) := \varphi(hg^{-1}), \tag{3.1}$$

and moreover these actions commute. From this we obtain a natural $\mathbb{C}[G]$-bimodule structure on $C(G)$, where $\mathbb{C}[G]$ is the group algebra of G over $\mathbb{C}$. For each $g \in G$, we denote by e_g the corresponding element of $\mathbb{C}[G]$. Then given a finite sequence $\boldsymbol{a} = (a_1, \dots, a_n)$ of elements in G, we can definite the *right* (resp. *left*) *average* of a function $\varphi \in C(G)$ with respect to $\boldsymbol{a}$ by

$$\boldsymbol{a} \cdot \varphi := \left(\frac{1}{n}\sum_{i=1}^{n} e_{a_i}\right) \cdot \varphi \quad \text{and} \quad \varphi \cdot \boldsymbol{a} := \varphi \cdot \left(\frac{1}{n}\sum_{i=1}^{n} e_{a_i}\right). \tag{3.2}$$

We obtain commuting linear endomorphisms $\rho_{\boldsymbol{a}}$ and $\ell_{\boldsymbol{b}}$ of $C(G)$ by defining $\rho_{\boldsymbol{a}}(\varphi) := \varphi \cdot \boldsymbol{a}$ and $\ell_{\boldsymbol{b}}(\varphi) := \boldsymbol{b} \cdot \varphi$. For a fixed $\varphi \in C(G)$, we let $\mathcal{R}_\varphi$ be the set of right averages $\varphi \cdot \boldsymbol{a}$ of φ as $\boldsymbol{a}$ ranges over all finite sequences, and similarly the set of left averages is denoted $\mathcal{L}_\varphi$. Intuitively, the law of large numbers suggests that if μ is a bi-invariant probability measure on G, then the constant function equal to $\int \varphi \, d\mu$ should lie in both $\overline{\mathcal{R}_\varphi}$ and $\overline{\mathcal{L}_\varphi}$. In fact, we will show that for any $\varphi \in C(G)$, there is always a *unique* constant function lying in either $\overline{\mathcal{R}_\varphi}$ or $\overline{\mathcal{L}_\varphi}$, and moreover this constant function lies in both closures.

Lemma 3.2. Let $\varphi \in C(G)$. Then there exists a unique constant function lying in either $\overline{\mathcal{R}_\varphi}$ or $\overline{\mathcal{L}_\varphi}$, and this constant function is an element of both spaces.

For the moment, suppose that we have proven Lemma 3.2. Then we can define a function $\mu : C(G) \to \mathbb{C}$ such that $\mu(\varphi)$ is the unique constant $c \in \mathbb{C}$ for which $g \mapsto c$ lies in $\overline{\mathcal{R}_\varphi}$ (or equivalently $\overline{\mathcal{L}_\varphi}$). If we can further show that μ is a positive linear functional, by the Riesz-Markov-Kakutani representation theorem [Rud87, Theorem 2.14] we will obtain an induced measure μ on G for which $\mu(\varphi) = \int \varphi \, d\mu$ $(\varphi \in C(G))$. This measure turns out to be a bi-invariant probability measure on G, and *any* Riesz probability measure ν on G that is either right or left-invariant is easily seen to equal μ. We will more carefully justify these details, beginning with a proof of Lemma 3.2.

First, for a function $f : G \to \mathbb{C}$ and a subset $A \subseteq G$ we can define the *right* (resp. *left*) *oscillation* of f on A, denoted $\omega_R(f, A)$ (resp. $\omega_L(f, A)$), by

$$\omega_R(f, A) := \sup_{gh^{-1}\in A} |f(g) - f(h)| \quad \text{and} \quad \omega_L(f, A) := \sup_{g^{-1}h\in A} |f(g) - f(h)|. \tag{3.3}$$

In the case that $A = G$, these two numbers are both equal to

$$\omega(f) = \sup_{g,h \in G} |f(g) - f(h)|, \tag{3.4}$$

called the (*total*) *oscillation* of f. Note that $\omega(\varphi) \leq 2\|\varphi\| < \infty$ if $\varphi \in C(G)$, and for any $B \subseteq A$ we have $\omega_R(f, B) \leq \omega_R(f, A)$ and $\omega_L(f, B) \leq \omega_L(f, A)$. It's also evident that $\omega_R(\cdot, A)$ and $\omega_L(\cdot, A)$ are seminorms on $C(G)$, for each $A \subseteq G$.

The motivation for defining right and left oscillation is essentially two-fold. On the one hand, the total oscillation $\omega(f)$ gives us a way of quantifying how much a given $f : G \to \mathbb{C}$ deviates from being a constant function. Indeed, $\omega(f) = 0$ if and only if f is constant. On the other hand, the language of oscillation on a neighborhood of the identity element $e \in G$ allows us to formulate a satisfactory definition of *uniform continuity* on G.

Definition 3.3. Let $f : G \to \mathbb{C}$. Then we say that f is *right* (resp. *left*) *uniformly continuous* if

$$\inf_U \omega_R(f, U) = 0 \quad (\text{resp. } \inf_U \omega_L(f, U) = 0), \tag{3.5}$$

where U ranges over all neighborhoods of the identity $e \in G$.

We know that every continuous function on a compact subset of $\mathbb{R}^n$ is uniformly continuous, so it is reasonable to expect that every $\varphi \in C(G)$ is (both right and left) uniformly continuous. Conveniently, this happens to be the case.

Lemma 3.4. Let $\varphi \in C(G)$. Then φ is both right and left uniformly continuous.

Proof. It suffices to prove that φ is right uniformly continuous, for otherwise we can consider the group G^{opp} whose underlying space is G but with reversed multiplication $g \cdot_{\text{opp}} h := hg$. In particular, note that for any $A \subseteq G$ and $f : G \to \mathbb{C}$ we have $\omega_L(f, A) = \omega_R(f^{\text{opp}}, A)$ and vice versa, where f^{opp} is the same as f but with the underlying group G^{opp}. It is then clear from Definition 3.3 that the left uniform continuity of φ on G is equivalent to the right uniform continuity of φ on G^{opp}.

Now fix $\varphi \in C(G)$ and $\varepsilon > 0$. We need to construct a neighborhood U of the identity $e \in G$ for which $\omega_R(\varphi, U) \leq \varepsilon$. Let

$$V_\varepsilon := \{(g, h) \in G \mid |\varphi(g) - \varphi(h^{-1}g)| < \varepsilon\}, \tag{3.6}$$

which is evidently an open subset of $G \times G$ containing $G \times \{e\}$. By compactness, we can find a neighborhood $U \ni e$ with $G \times U \subseteq V_\varepsilon$. Then if $gh^{-1} \in U$, we have

$$|\varphi(g) - \varphi(h)| = |\varphi(g) - \varphi((gh^{-1})^{-1}g)| < \varepsilon \tag{3.7}$$

because $(g, gh^{-1}) \in G \times U \subseteq V_\varepsilon$. This implies that $\omega_R(\varphi, U) \leq \varepsilon$, completing the proof. $\square$

Next, we examine how the supremum norm and oscillation behave with respect to taking right and left averages. Both the standard right and left actions of G on $C(G)$ are easily seen to be isometric (i.e., all elements act by isometries with respect to $\|\cdot\|$), so clearly $\|\varphi \cdot \boldsymbol{a}\| \leq \|\varphi\|$ and $\|\boldsymbol{a} \cdot \varphi\| \leq \|\varphi\|$ for all $\varphi \in C(G)$ and finite sequences $\boldsymbol{a}$. In other words, the linear endomorphisms $\rho_{\boldsymbol{a}}$ and $\ell_{\boldsymbol{a}}$ are continuous, with operator norm at worst 1.

As for oscillation, fix a subset $A \subseteq G$ and a function $f : G \to \mathbb{C}$. If $g_0 \in G$, then note that $gh^{-1} = (gg_0^{-1})(hg_0^{-1})^{-1}$ and $g^{-1}h = (g_0^{-1}g)^{-1}(g_0^{-1}h)$ for all $g, h \in G$, so that by a simple change of variables $\omega_R(f \cdot g_0, A) = \omega_R(f, A)$ and $\omega_L(g_0 \cdot f, A) = \omega_L(f, A)$. Using that $\omega_R(\cdot, A)$ and $\omega_L(\cdot, A)$ are both seminorms, for all $\varphi \in C(G)$ finite sequences $\boldsymbol{a}$ we see that

$$\omega_R(\varphi \cdot \boldsymbol{a}, A) \leq \omega_R(\varphi, A) \quad \text{and} \quad \omega_L(\boldsymbol{a} \cdot \varphi, A) \leq \omega_L(\varphi, A) \tag{3.8}$$

Lemma 3.5. Let $\varphi \in C(G)$. Then $\mathcal{R}_\varphi$ and $\mathcal{L}_\varphi$ are relatively compact in $C(G)$.

Proof. By Arzelà-Ascoli, it suffices to prove that $\mathcal{R}_\varphi$ and $\mathcal{L}_\varphi$ are both pointwise bounded and equicontinuous. If $\psi \in \mathcal{R}_\varphi \cup \mathcal{L}_\varphi$, then $\|\psi\| \leq \|\varphi\|$, which proves pointwise boundedness. As for equicontinuity, we make a stronger claim: $\mathcal{R}_\varphi$ is right uniformly equicontinuous and $\mathcal{L}_\varphi$ is left uniformly equicontinuous. By this we mean that for any $\varepsilon > 0$, we can find a neighborhood $U \ni e$ such that $\omega_R(\psi, U) < \varepsilon$ and $\omega_L(\theta, U) < \varepsilon$ for all $\psi \in \mathcal{R}_\varphi$ and $\theta \in \mathcal{L}_\varphi$. Both claims are immediate by Lemma 3.4 and (3.8). $\square$

Lemma 3.6. Let $\varphi \in C(G)$ be nonconstant. Then there exists a finite sequence $\boldsymbol{a}$ in G such that $\omega(\varphi \cdot \boldsymbol{a}) < \omega(\varphi)$. Similarly, there exists $\boldsymbol{b}$ for which $\omega(\boldsymbol{b} \cdot \varphi) < \omega(\varphi)$.

Proof. It suffices to prove this for right averages, since as in the proof of Lemma 3.4 we can otherwise replace G with G^{opp}. Indeed, replacing G with G^{opp} flips the left and right actions of $\mathbb{C}[G]$ on $C(G)$, so that left averages become right averages and vice versa. In particular, $\omega(\boldsymbol{b} \cdot \varphi) = \omega(\varphi^{\text{opp}} \cdot \boldsymbol{b})$ for any finite sequence $\boldsymbol{b}$.

Now let $\varphi \in C(G)$ be nonconstant, so $\omega(\varphi) > 0$. Choose any $0 < \varepsilon < \omega(\varphi)$ and let $U = \varphi^{-1}(B_{\varepsilon/2}(\varphi(e)))$, where $B_r(z)$ is the open ball of radius r centered at z in $\mathbb{C}$. By continuity, U is a neighborhood of the identity $e \in G$, and since the right translates of U cover G, by compactness we can find a finite sequence $\boldsymbol{a} = (a_1, \ldots, a_n)$ in G such that the open sets $U_i = Ua_i$ cover G ($1 \leq i \leq n$). Let $g, h \in G$ be arbitrary, and choose indices i, j for which $g \in U_i$ and $h \in U_j$. Let σ be the permutation of $\{1, 2, \ldots, n\}$ swapping i and j and leaving all other indices fixed. Then

$$|(\varphi \cdot \boldsymbol{a})(g) - (\varphi \cdot \boldsymbol{a})(h)| \leq \frac{1}{n}|\varphi(ga_i^{-1}) - \varphi(ha_j^{-1})| + \frac{1}{n}\sum_{k \neq i}|\varphi(ga_k^{-1}) - \varphi(ha_{\sigma(k)}^{-1})| \tag{3.9}$$

$$< \frac{\varepsilon}{n} + \frac{n-1}{n}\omega(\varphi), \tag{3.10}$$

where the second inequality uses that $ga_i^{-1}, ha_j^{-1} \in U$ and that $|\varphi(g_1) - \varphi(g_2)| \leq \omega(\varphi)$ for all $g_1, g_2 \in G$. In particular, $\omega(\varphi \cdot \boldsymbol{a}) \leq (\varepsilon + (n-1)\omega(\varphi))/n < \omega(\varphi)$, as desired. $\square$

Note that the total oscillation $\omega : C(G) \to \mathbb{R}_{\geq 0}$ is continuous, since

$$|\omega(\varphi) - \omega(\psi)| \leq \omega(\varphi - \psi) \leq 2\|\varphi - \psi\| \quad \text{for all } \varphi, \psi \in C(G). \tag{3.11}$$

Then by Lemma 3.5, ω must achieve its minimum on both $\overline{\mathcal{R}_\varphi}$ and $\overline{\mathcal{L}_\varphi}$. Note also that $\mathcal{R}_\varphi$ is closed under right averages and $\mathcal{L}_\varphi$ is closed under left averages, since if $\boldsymbol{a} = (a_1, \ldots, a_n)$ and $\boldsymbol{b} = (b_1, \ldots, b_m)$ are finite sequences in G, then

$$(\varphi \cdot \boldsymbol{a}) \cdot \boldsymbol{b} = \varphi \cdot \boldsymbol{c} \quad \text{and} \quad \boldsymbol{a} \cdot (\boldsymbol{b} \cdot \varphi) = \boldsymbol{c} \cdot \varphi \tag{3.12}$$

for a finite sequence $\boldsymbol{c}$. In particular, we can take $\boldsymbol{c}$ to be the sequence of nm elements, indexed by pairs of positive integers (i, j) with $i \leq n$ and $j \leq m$, given by $c_{i,j} = a_i b_j$. We are finally equipped to prove Lemma 3.2.

Proof. (Lemma 3.2) First we claim that $\overline{\mathcal{R}_\varphi}$ and $\overline{\mathcal{L}_\varphi}$ both contain constant functions. As usual, it suffices to prove this for $\overline{\mathcal{R}_\varphi}$, since we can otherwise consider G^{opp}, as $\mathcal{L}_\varphi = \mathcal{R}_{\varphi^{\text{opp}}}$.

Let $\psi \in \overline{\mathcal{R}_\varphi}$ be such that $\omega(\psi)$ is the minimum of ω on $\overline{\mathcal{R}_\varphi}$. Suppose, for the sake of contradiction, that ψ is not constant. By Lemma 3.6, we can find $\boldsymbol{a}$ for which $\omega(\psi \cdot \boldsymbol{a}) < \omega(\psi)$. But by the continuity of $\rho_{\boldsymbol{a}}$ and the invariance of $\mathcal{R}_\varphi$ under right averaging,

$$\rho_{\boldsymbol{a}}(\overline{\mathcal{R}_\varphi}) \subseteq \overline{\rho_{\boldsymbol{a}}(\mathcal{R}_\varphi)} \subseteq \overline{\mathcal{R}_\varphi}. \tag{3.13}$$

In particular, $\psi \cdot \boldsymbol{a} \in \overline{\mathcal{R}_\varphi}$, which contradicts the minimality of $\omega(\psi)$.

It remains to verify uniqueness. To that end, suppose that c and d are constant functions such that $c \in \overline{\mathcal{R}_\varphi}$ and $d \in \overline{\mathcal{L}_\varphi}$. We wish to show $c = d$. Let $\varepsilon > 0$. Then we can find sequences $\boldsymbol{a}$ and $\boldsymbol{b}$ for which

$$\|\varphi \cdot \boldsymbol{a} - c\| < \frac{\varepsilon}{2} \quad \text{and} \quad \|\boldsymbol{b} \cdot \varphi - d\| < \frac{\varepsilon}{2}. \tag{3.14}$$

Note that constant functions are invariant under both right and left averaging, so that

$$\|\boldsymbol{b} \cdot \varphi \cdot \boldsymbol{a} - c\| = \|\boldsymbol{b} \cdot \varphi \cdot \boldsymbol{a} - \boldsymbol{b} \cdot c\| = \|\boldsymbol{b} \cdot (\varphi \cdot \boldsymbol{a} - c)\| \leq \|\varphi \cdot \boldsymbol{a} - c\| < \frac{\varepsilon}{2}, \tag{3.15}$$

and analogously $\|\boldsymbol{b} \cdot \varphi \cdot \boldsymbol{a} - d\| < \varepsilon/2$. Therefore by the triangle inequality,

$$\|c - d\| \leq \|c - \boldsymbol{b} \cdot \varphi \cdot \boldsymbol{a}\| + \|\boldsymbol{b} \cdot \varphi \cdot \boldsymbol{a} - d\| < \varepsilon. \tag{3.16}$$

Taking $\varepsilon \to 0$, we conclude that $c = d$, as desired. $\square$

Now we can let $\mu : C(G) \to \mathbb{C}$ be the function described in the discussion following the statement of Lemma 3.2. We claim that μ is a positive linear functional. Invariance under scalar multiplication is clear, since for all $\alpha \in \mathbb{C}$, $\varphi \in C(G)$, and finite sequences $\boldsymbol{a}$ we have $(\alpha\varphi) \cdot \boldsymbol{a} = \alpha(\varphi \cdot \boldsymbol{a})$; this implies that $\mathcal{R}_{\alpha\varphi} = \alpha\mathcal{R}_\varphi$, so in particular $\overline{\mathcal{R}_{\alpha\varphi}} = \alpha\overline{\mathcal{R}_\varphi}$. Next we verify additivity. Let $\varphi, \psi \in C(G)$, so that we wish to prove

$$c_{\mu(\varphi)+\mu(\psi)} = c_{\mu(\varphi)} + c_{\mu(\psi)} \in \overline{\mathcal{R}_{\varphi+\psi}}, \tag{3.17}$$

where c_α is the constant map on G valued $\alpha \in \mathbb{C}$. To this end, fix $\varepsilon > 0$. Choose finite sequences $\boldsymbol{a}$ and $\boldsymbol{b}$ with $\|\varphi \cdot \boldsymbol{a} - c_{\mu(\varphi)}\| < \varepsilon/2$ and $\|\psi \cdot \boldsymbol{b} - c_{\mu(\psi)}\| < \varepsilon/2$. If $\boldsymbol{c}$ is as in (3.12),

$$\|\varphi \cdot \boldsymbol{c} - c_{\mu(\varphi)}\| = \|(\varphi \cdot \boldsymbol{a} - c_{\mu(\varphi)}) \cdot \boldsymbol{b}\| \leq \|\varphi \cdot \boldsymbol{a} - c_{\mu(\varphi)}\| < \frac{\varepsilon}{2}, \tag{3.18}$$

and similarly $\|\psi \cdot \boldsymbol{c} - c_{\mu(\psi)}\| < \varepsilon/2$. Thus, by the triangle inequality,

$$\|(\varphi + \psi) \cdot \boldsymbol{c} - (c_{\mu(\varphi)} + c_{\mu(\psi)})\| \leq \|\varphi \cdot \boldsymbol{c} - c_{\mu(\varphi)}\| + \|\psi \cdot \boldsymbol{c} - c_{\mu(\psi)}\| < \varepsilon. \tag{3.19}$$

It remains to prove positivity. Since $[0, \infty)$ is a closed subset of $\mathbb{C}$, the set $C_+(G)$ of continuous maps $G \to [0, \infty)$ is closed in $C(G)$ (as uniform convergence implies pointwise convergence). If $\varphi \in C_+(G)$, it is clear that $\mathcal{R}_\varphi \subseteq C_+(G)$, and therefore

$$c_{\mu(\varphi)} \in \overline{\mathcal{R}_\varphi} \subseteq \overline{C_+(G)} = C_+(G). \tag{3.20}$$

Hence $\mu(\varphi) \geq 0$, which proves our claim that μ is a positive linear functional. Note that $\mu(1) = 1$, where $1 \in C(G)$ is shorthand for the constant function c_1. Therefore, by the Riesz-Markov-Kakutani representation theorem, μ arises from a Riesz probability measure on G, which we also denote by μ.

Proof. (Theorem 3.1) We first claim that μ is bi-invariant, and hence both a right and left Haar probability measure on G. Let $\varphi \in G$ and $g \in G$. Then for any finite sequence $\boldsymbol{a} = (a_1, \ldots, a_n)$, we have

$$(\varphi \cdot g) \cdot \boldsymbol{a} = \varphi \cdot (g\boldsymbol{a}) \quad \text{and} \quad \boldsymbol{a} \cdot (g \cdot \varphi) = (\boldsymbol{a}g) \cdot \varphi, \tag{3.21}$$

where $g\boldsymbol{a} = (ga_1, \ldots, ga_n)$ and $\boldsymbol{a}g = (a_1 g, \ldots, a_n g)$. In particular,

$$\mathcal{R}_{\varphi \cdot g} \subseteq \mathcal{R}_\varphi = \mathcal{R}_{(\varphi \cdot g) \cdot g^{-1}} \subseteq \mathcal{R}_{\varphi \cdot g} \implies \mathcal{R}_{\varphi \cdot g} = \mathcal{R}_\varphi, \tag{3.22}$$

and similarly $\mathcal{L}_{g \cdot \varphi} = \mathcal{L}_\varphi$. Thus $\mu(\varphi \cdot g) = \mu(\varphi) = \mu(g \cdot \varphi)$. Letting R_g and L_g denote right and left multiplication by g on G, this means that the probability measures μ, $(R_g)_*\mu$, and $(L_g)_*\mu$ all agree as functionals on $C(G)$. Since R_g and L_g are both homeomorphisms of G, it's evident that $(R_g)_*\mu$ and $(L_g)_*\mu$ inherit the regularity properties of μ so as to ensure that the uniqueness portion of the Riesz-Markov-Kakutani representation theorem applies. In particular, all three measures μ, $(R_g)_*\mu$, and $(L_g)_*\mu$ agree, proving bi-invariance.

Now we handle uniqueness. Suppose that ν is a right Haar probability measure on G. Certainly $\int \varphi \cdot \boldsymbol{a} \, d\nu = \int \varphi \, d\nu$ for all $\varphi \in C(G)$ and finite sequences $\boldsymbol{a}$. Of course, the linear operator $\psi \mapsto \int \psi \, d\nu$ is continuous (of operator norm at worst 1) as a map $C(G) \to \mathbb{C}$, and so being constant on $\mathcal{R}_\varphi$, it must also be constant on its closure. Therefore

$$\mu(\varphi) = \int \mu(\varphi) \, d\nu = \int \varphi \, d\nu. \tag{3.23}$$

Then by the uniqueness portion of Riesz-Markov-Kakutani, we conclude $\nu = \mu$. An analogous proof allows us to reach the same conclusion if ν is instead assumed to be a *left* Haar probability measure on G. □

We have already seen how the Haar measure on O_N relates to the Gaussian distribution and Gram-Schmidt in Theorem 1.13. We can also devise quite explicit descriptions of the Haar measure for other compact groups.

Example 3.7 [Finite Groups]. Any finite group G is in particular a compact Lie group, and the Haar probability measure is simply the uniform measure on the set G, i.e., the counting measure on G divided by its cardinality $|G|$.

Example 3.8 [Circle Group]. Let $\mathbb{T}$ denote the circle group, that is the unit circle $S^1 \subseteq \mathbb{C}$ equipped with (complex) multiplication. This is a compact Lie group and hence has a bi-invariant probability measure μ. Consider the measurable bijection $f : [0, 1) \to \mathbb{T}$ sending $\theta \mapsto e^{2\pi i\theta}$. Then $\mu = f_*\lambda$, where λ is the Lebesgue measure restricted to $[0, 1)$. In other words, μ is uniform in the angular coordinate θ.

We end this subchapter with two general results about the Haar measure.

Lemma 3.9. Let G be a compact group with Haar probability measure μ. Then $\mu(U) > 0$ for every nonempty open subset $U \subseteq G$.

Proof. Let U be an arbitrary nonempty open subset of G. By compactness, we can find $g_1, \ldots, g_n \in G$ such that $Ug_1, \ldots, Ug_n$ cover G. Hence

$$1 = \mu(G) \leq \sum_{i=1}^{n} \mu(Ug_i) = n\mu(U) \implies \frac{1}{n} \leq \mu(U). \tag{3.24}$$

Thus $\mu(U) > 0$, as desired. $\square$

Lemma 3.10. Let G be a compact group with Haar probability measure μ, and let $i : G \to G$ be the inversion map $i(g) = g^{-1}$. Then $i_*\mu = \mu$.

Proof. Since i is a homeomorphism, $i_*\mu$ is a Riesz probability measure on G. Thus, it suffices to show $\int \varphi \cdot g \, di_*\mu = \int \varphi \, di_*\mu$ for all $\varphi \in C(G)$ and $g \in G$. But this is immediate by the observation that $(\varphi \cdot g) \circ i = g^{-1} \cdot (\varphi \circ i)$, combined with the bi-invariance of μ. $\square$

3.2 Linear Representations

Given a compact group G with Haar probability measure μ, we write $L^p(G) := L^p(\mu)$ for each $0 < p \leq \infty$. Since μ is a probability measure, $\| \cdot \|_p \leq \| \cdot \|_q$ whenever $p < q$. The goal of the next two subchapters is to develop a useful setting for Fourier analysis on a compact Lie group G, such as O_N or SO_N, by constructing a convenient Hilbert basis for $L^2(G)$.

The Peter-Weyl theorem answers this question in terms of the representation theory of G. Thus, we devote this subchapter to recalling basic notions from representation theory.

Let G be a group and k a base field. A *left* (resp. *right*) (*linear*) *representation* of G, or *G-rep* for short, is a left (resp. right) action of G on a k-vector space V, called the *representation space*, by linear maps. In other words, a G-rep is a group homomorphism $\rho : G \to \mathrm{GL}(V)$, where $\mathrm{GL}(V)$ is the group of linear automorphisms of V. Unless otherwise specified, all of our representations will be left by default. As an abuse of notation, we may denote a G-rep merely by its associated representation space V, with the corresponding action left implicit. The *dimension* of this representation is just $\dim V$, allowing for a notion of *finite-dimensional* (FD) and *infinite-dimensional* G-reps. Associated to every FD G-rep $\rho : G \to \mathrm{GL}(V)$ is its *character* $\chi_V : G \to k$ defined as the composition $\mathrm{tr} \circ \rho$, where $\mathrm{tr} : \mathrm{End}(V) \to k$ is the trace operator. If $\rho(g) = \mathrm{id}_V$, notice $\chi_V(g) = \dim V$.

Example 3.11 [Trivial Representation]. For every group G and vector space V, we have the *trivial representation* on V, where each $g \in G$ acts as the identity on V. Such a representation is referred to as *a* trivial G-rep, while *the* trivial G-rep is reserved for the 1-dimensional trivial representation of G on k.

A subspace $W \subseteq V$ of a G-rep is called *G-invariant*, or simply invariant if G is implied, whenever $G{\cdot}W \subseteq W$. Then V naturally restricts to a G-rep on W, called a *subrepresentation* of V. Arbitrary intersections and sums of invariant subspaces are invariant. G-reps over k form a category $G\text{-}\mathsf{Rep}_k$ whose morphisms are the G-equivariant linear maps $V \to W$, also called *G-linear*. Notice that the character is invariant under G-isomorphism. The set of G-linear maps $V \to W$ is denoted $\mathrm{Hom}_G(V, W)$. Importantly, given $\varphi \in \mathrm{Hom}_G(V, W)$, the spaces $\ker \varphi$ and $\mathrm{im}\, \varphi$ are subrepresentations of V and W, respectively. If $U \subseteq W$ is invariant, the preimage $\varphi^{-1}(U) \subseteq V$ is also invariant.

Standard operations on vector spaces, such as direct sums, tensor products, Hom-spaces, quotients, and duals give rise to operations on representations. We assume familiarity with these operations on representations, which are detailed in, e.g., [FH91]. If V and W are G-reps, we equip $V \oplus W$, $V \otimes W$, $\mathrm{Hom}(V, W)$, V/W, and V^* with their induced representations by default. Writing V^G for the invariant subspace of vectors $v \in V$ fixed by all $g \in G$, notice $\mathrm{Hom}_G(V, W) = \mathrm{Hom}(V, W)^G$ is a subrepresentation of $\mathrm{Hom}(V, W)$. If $\rho : G \to \mathrm{GL}(V)$ is a G-rep and $\varphi : H \to G$ is a group homomorphism, then $\rho \circ \varphi$ is an H-rep on V, called the *restriction* of ρ along φ. In particular, this applies whenever H is a subgroup of G with φ the inclusion map. A G-rep V is called *irreducible* if it contains no nontrivial invariant subspaces (i.e., $W \subseteq V$ is invariant only if $W = \{0\}$ or $W = V$) and is otherwise called *reducible*. It is common to abbreviate “irreducible representation” as *irrep*.

Lemma 3.12 [Schur's Lemma]. Let $\varphi : V \to W$ be a G-linear map over a field k, with V irreducible. Then either $\varphi = 0$ or φ is injective. If W is also irreducible and φ is injective, then φ is an isomorphism. If k is algebraically closed and $V = W$ is finite-dimensional, then $\varphi = \lambda \cdot \mathrm{id}_V$ for some $\lambda \in k$.

Proof. Since $\ker\varphi$ is a subrepresentation of V, and V is irreducible, either $\ker\varphi = V$ (so that $\varphi = 0$) or $\ker\varphi = \{0\}$ (so that φ is injective). If W is also irreducible and $\varphi \neq 0$, then since $\operatorname{im}\varphi$ is invariant necessarily $\operatorname{im}\varphi = W$, making φ an isomorphism. Finally, suppose that $V = W$ is FD and k is algebraically closed. Then φ has an eigenvalue $\lambda \in k$, and its eigenspace E_λ is a nonzero invariant subspace of V. Hence $E_\lambda = V$, so $\varphi = \lambda \cdot \mathrm{id}_V$. ☐

Corollary 3.13. *If V and W are FD G-reps over an algebraically closed field k, then $\dim \operatorname{Hom}_G(V, W)$ is 1 when $V \cong W$ as G-reps and 0 otherwise.*

Corollary 3.14. *If G is abelian and k is algebraically closed, then every G-irrep over k is 1-dimensional.*

Proof. Let $\rho : G \to \mathrm{GL}(V)$ be irreducible. For each $g \in G$, $\rho(g) : V \to V$ is G-linear because

$$\rho(g)(hv) = \rho(gh)v = \rho(hg)v = h\rho(g)(v). \tag{3.25}$$

By Schur's lemma, $\rho(g)$ is a scalar multiple of the identity. But then every line in V is invariant, so by irreducibility $\dim V = 1$. ☐

If G is a group, we write $\mathcal{I}_k(G)$ for the isomorphism classes of FD G-irreps over k. Note that $\mathcal{I}_k(G)$ can actually be realized as a set, for if V and W are isomorphic vector spaces, any G-rep on V is isomorphic to one on W, in which case any n-dimensional G-rep is isomorphic to a G-rep on k^n. For an isomorphism class π of G-irreps, let W_π be a fixed representative of π. In many cases, a representation can be decomposed as a direct sum of irreps, and then the representation is said to be *semisimple*. The decomposition of any semisimple FD G-rep into irreps is unique, in the following sense.

Theorem 3.15. *Let V be a semisimple FD G-rep, and suppose $V = U_1 \oplus \cdots \oplus U_n$ and $V = U_1' \oplus \cdots \oplus U_m'$ are two decompositions into irreps. Then $n = m$, and there exists a permutation $\sigma \in S_n$ such that $U_i \cong U_{\sigma(i)}'$ as G-reps for all $i \leq n$.*

The proof of Theorem 3.15 is an application of Schur's lemma. Given a G-rep V and an isomorphism class π of G-irreps, we define the π-*isotype* V_π of V as the sum of all subrepresentations of V lying in the class π (i.e., isomorphic to W_π). By Schur's lemma, this is the sum of all images of W_π under G-linear maps $W_\pi \to V$. Notably, V_π is itself an invariant subspace of V, and in fact, it can be realized as the image of the G-linear map

$$\Phi_\pi : W_\pi \otimes \operatorname{Hom}_G(W_\pi, V) \to V, \quad w \otimes \varphi \mapsto \varphi(w). \tag{3.26}$$

Suppose now that V is semisimple. Then V is certainly the span of its isotypes. Suppose further that V is FD and fix an irreducible decomposition $V = U_1 \oplus \cdots \oplus U_n$. The key insight is that, for each $\pi \in \mathcal{I}_k(G)$, the π-isotype V_π is the sum of all U_i for which $U_i \in \pi$.

Since the isotypes of V are canonical (independent of an irreducible decomposition), for each $\pi \in \mathcal{I}_k(G)$ we can define $a_\pi := \dim V_\pi / \dim W_\pi$, so that there are precisely a_π irreps $U_i \in \pi$. This immediately implies Theorem 3.15. Indeed, we get a canonical, finitely supported *multiplicity map* $a : \mathcal{I}_k(G) \to \mathbb{Z}_{\geq 0}$ by defining $\pi \mapsto a_\pi$, and then

$$V = \bigoplus_{\pi \in \mathcal{I}_k(G)} V_\pi \cong \bigoplus_{\pi \in \mathcal{I}_k(G)} W_\pi^{\oplus a_\pi}. \tag{3.27}$$

The internal direct sum in (3.27) is called the *isotypic decomposition* of V. Each a_π is referred to as the *multiplicity* of π in the semisimple G-rep V. If k is algebraically closed, then by Schur's lemma $a_\pi = \dim \operatorname{Hom}_G(W_\pi, V)$.

Proof. (Theorem 3.15) Let $I_\pi = \{i \in [n] \mid U_i \in \pi\}$. By the preceding discussion, it suffices to prove that $V_\pi = \oplus_{i \in I_\pi} U_i$. By definition $U_i \subseteq V_\pi$ for each $i \in I_\pi$, so that $\oplus_{i \in I_\pi} U_i \subseteq V_\pi$. Conversely, suppose that $U \in \pi$ is a subrepresentation of V and fix $i \notin I_\pi$. Then the composition $\operatorname{pr} \circ \iota : U \to U_i$ of $\iota : U \hookrightarrow V$ and the projection map $\operatorname{pr} : V \to U_i$ (with respect to the direct sum $V = U_1 \oplus \cdots \oplus U_n$) is 0 by Schur's lemma. Hence $U \subseteq \oplus_{i \in I_\pi} U_i$. But the span of all such U is V_π, so that altogether $V_\pi = \oplus_{i \in I_\pi} U_i$. □

More generally, if V is a possibly infinite-dimensional semisimple G-rep and $V = \oplus_{i \in I} U_i$ is an irreducible decomposition, our proof of Theorem 3.15 can still be used to conclude that the direct sum of U_i in a fixed isomorphism class of G-reps is equal to the corresponding isotype of V. In that case, we again obtain an isotypic decomposition of V. We can also restrict our attention to the invariant subspace V_{fin} of V formed by the union of all FD invariant subspaces. This is the set of *finite type* vectors in V. For every $\pi \in \mathcal{I}_k(G)$, the π-isotype V_π of V is also the π-isotype of V_{fin}, and the function Φ_π in (3.26) maps into V_{fin}. Then the Φ_π induce a G-linear map

$$\Phi : \bigoplus_{\pi \in \mathcal{I}_k(G)} (W_\pi \otimes \operatorname{Hom}_G(W_\pi, V)) \to V_{\text{fin}}. \tag{3.28}$$

A notion very closely related to semisimplicity is *complete reducibility.* A representation V is called *completely reducible* if given any invariant subspace W of V, there exists an invariant complement W', i.e., $V = W \oplus W'$. Many authors make no distinction between semisimplicity and complete reducibility, as they are ultimately equivalent (assuming the Axiom of Choice for the infinite-dimensional case).

Lemma 3.16. Fix a base field k. Let G be a group and V a G-rep. Then V is semisimple if and only if it is completely reducible.

Proof. See the Appendix (C). □

Since intersections of invariant subspaces are invariant, every subrepresentation W of a completely reducible G-rep V is completely reducible. Indeed, for $U \subseteq W$ invariant, if U' is an invariant complement to U in V, then $U' \cap W$ is an invariant complement to U in W.

If V is a real or complex G-rep equipped with a G-invariant inner product $\langle -, - \rangle$, meaning that G acts on V by orthogonal/unitary maps, then V is completely reducible. In fact, if W is an invariant subspace of V, then $W^\perp$ is also invariant. When V is FD, this means that a decomposition into orthogonal irreps is possible. We say that a group G is completely reducible (over a field k) if every FD G-rep (over k) is completely reducible.

Lemma 3.17. Let G be completely reducible over an algebraically closed field k and V a G-rep over k. Assume that the isotypes of V_{fin} are FD. Then for every $\pi \in \mathcal{I}_k(G)$, the map Φ_π in (3.26) is an isomorphism onto V_π. Moreover, V is semisimple with isotypic decomposition $V_{\text{fin}} = \oplus_{\pi \in \mathcal{I}_k(G)} V_\pi$, so that Φ in (3.28) is an isomorphism.

Proof. Let $\pi \in \mathcal{I}_k(G)$. We already know that Φ_π is G-linear and maps onto V_π. By complete reducibility and finite-dimensionality, we can decompose V_π as a direct sum of G-irreps, and the multiplicity $a_\pi = \operatorname{Hom}_G(W_\pi, V_\pi)$ because k is algebraically closed. But evidently $\operatorname{Hom}_G(W_\pi, V) \cong \operatorname{Hom}_G(W_\pi, V_\pi)$, so that $\dim(W_\pi \otimes \operatorname{Hom}_G(W_\pi, V)) = a_\pi \dim W_\pi$. Hence by the rank-nullity theorem, Φ_π must be an isomorphism and $V_\pi \cong W_\pi^{\oplus a_\pi}$. By complete reducibility and the definition of V_{fin}, Φ must be surjective. It suffices to prove that $V_\pi \cap V_{\pi'} = \{0\}$ for $\pi \neq \pi' \in \mathcal{I}_k(G)$. Indeed, the multiplicity of any G-irrep in the decomposition of the FD subrepresentation $V_\pi \cap V_{\pi'}$ must be 0, as desired. $\square$

Remark 3.18. The condition that the π-isotype V_π is FD for every $\pi \in \mathcal{I}_k(G)$ holds, for example, whenever $\dim \operatorname{Hom}_G(W_\pi, V) < \infty$. Under the conditions of Lemma 3.17, an analog to Theorem 3.15 also holds: given two irreducible decompositions $V_{\text{fin}} = \oplus_{i \in I} U_i$ and $V_{\text{fin}} = \oplus_{j \in J} U'_j$, there is a bijection $\sigma : I \to J$ such that $U_i \cong U'_{\sigma(i)}$ as G-reps for all $i \in I$ (an easy verification).

If G is a topological group, we can also consider *continuous* representations, meaning that the representation space V is a real/complex topological vector space (TVS) and that the action $G \times V \to V$ is continuous. By convention, we shall assume every TVS is Hausdorff, so that any FD real/complex vector space has a unique vector topology. All subrepresentations of a continuous representation are continuous with respect to the subspace topology. When V is FD, a continuous G-rep is equivalently a continuous homomorphism $G \to \mathrm{GL}(V)$. If V and W are FD continuous G-reps, so are $V \oplus W$, $V \otimes W$, $\operatorname{Hom}(V, W)$, V/W, and V^*. Restrictions of continuous representations along continuous homomorphisms are also continuous.

When G is a Lie group and V is FD, $\mathrm{GL}(V)$ is a Lie group, and continuous G-reps on V are automatically Lie group homomorphisms. From now on, if G is a topological group, then we assume G-reps are continuous unless stated otherwise. We redefine $\operatorname{Hom}_G(V, W)$, $\mathcal{I}_{\mathbb{R}}(G)$, and $\mathcal{I}_{\mathbb{C}}(G)$ in the continuous case accordingly. Effectively, we work in the category $G\text{-}\mathsf{CRep}_k$ of continuous G-reps over $k \in \{\mathbb{R}, \mathbb{C}\}$, whose morphisms are continuous G-linear maps, rather than the larger category $G\text{-}\mathsf{Rep}_k$ defined previously.

Example 3.19 [Standard Representation]. Let G be a *matrix* (Lie) group, i.e., a Lie subgroup of $\mathrm{GL}(V)$, where V is an FD TVS. Then the action $G \circlearrowright V$ by $A \cdot v = Av$ is called the *defining* or *standard representation* of G. This is a continuous action (and hence also smooth).

Checking that a given G-rep V is continuous can be a hassle directly. Two necessary conditions are that G acts on V by continuous linear operators, and that for every $v \in V$ the map $G \to V$, $g \mapsto g \cdot v$ is continuous. It turns out that when G is a compact group and V is a Banach space, these two conditions are together sufficient. This observation makes verifying the continuity of a G-rep much more tractable.

Theorem 3.20. Let V be a, not necessarily continuous, representation of a compact group G on a Banach space V by continuous operators. Then if $g \mapsto g \cdot v$ is continuous as a map $G \to V$ for each $v \in V$, the representation of G on V is continuous. For right representations, this is also true if we replace $g \cdot v$ by $v \cdot g$.

Proof. It suffices to prove this for left representations. An analogous proof applies in the right case, or we can note that inversion $i(g) = g^{-1}$ yields a bijection between left and right G-actions $G \to \mathrm{Perm}(S)$ on an arbitrary set S (where $\mathrm{Perm}(S)$ is the permutation group of S) by precomposing with i. This bijection also respects continuity of G-actions, since i is a homeomorphism of G with itself.

Let V be a (left, not necessarily continuous) representation of a compact group G on a Banach space. Then the family of linear maps $\{\rho(g) \mid g \in G\} \subseteq \mathrm{End}(V)$ has bounded orbits, so by the Banach-Steinhaus theorem there exists $M > 0$ such that $\|g \cdot v\| \leq M\|v\|$ for all $g \in G$—see the discussion following Theorem 2.6 in [Rud91]. Let $\varepsilon > 0$, and fix $(g, v) \in G \times V$. Choose a neighborhood $U \ni g$ such that $h \in U$ implies $\|g \cdot v - h \cdot v\| < \varepsilon/2$. Then if $(w, h) \in B_{\varepsilon/2M}(v) \times U$, we have

$$\begin{aligned} \|g \cdot v - h \cdot w\| &\leq \|g \cdot v - h \cdot v\| + \|h \cdot v - h \cdot w\| && (3.29) \\ &\leq \|g \cdot v - h \cdot v\| + M\|v - w\| < \varepsilon. && (3.30) \end{aligned}$$

This proves that the action $G \times V \to V$ is continuous, as desired. $\square$

In general, if G is a topological group and V is a normed space, then a representation ρ of G on V is dubbed *isometric* whenever $\rho(g)$ is isometric for all $g \in G$. In the case that V is a pre-Hilbert space, this is equivalent to requiring each $\rho(g)$ be unitary by the polarization identity, in which case the representation is called *unitary*. When G is a compact group, the space $L^2(G)$ is a Hilbert space with the usual inner product

$$\langle f_1, f_2 \rangle := \int_G f_1(g)\overline{f_2(g)}\, dg. \tag{3.31}$$

Here dg denotes integration with respect to the Haar probability measure on G, a convention which we will employ from this point on.

The standard right and left isometric representations of G on $C(G)$, which we call ρ and ℓ, respectively, extend to representations on $L^2(G)$, also denoted ρ and ℓ. By the invariance of the Haar measure, ρ and ℓ are unitary representations on $L^2(G)$. Technically, we have not yet verified that ρ and ℓ on $C(G)$ and $L^2(G)$ are continuous, although we know each $\rho(g)$ and $\ell(g)$ is continuous ($g \in G$). By Theorem 3.20, a necessary condition for the continuity of ρ and ℓ is the continuity of $g \mapsto g \cdot f$ and $g \mapsto f \cdot g$ as maps $G \to C(G)$ (resp. $G \to L^2(G)$) for each $f \in C(G)$ (resp. $f \in L^2(G)$). In the $C(G)$ case,

$$\|g \cdot f - h \cdot f\| = \|f - (g^{-1}h) \cdot f\| \tag{3.32}$$

for all $g, h \in G$, so that the left uniform continuity of f implies the continuity of $g \mapsto g \cdot f$. Similarly, the map $g \mapsto f \cdot g$ is continuous. We can then extend to the $L^2(G)$ case by using that $C(G)$ continuously embeds as a dense subspace of $L^2(G)$ [Rud87, Theorem 3.14].

A key fact about compact groups G (including all finite groups) is that *every* FD real/complex G-rep can be endowed with a G-invariant inner product. In particular, every compact group is completely reducible over $\mathbb{R}$ and $\mathbb{C}$, and every FD complex G-rep is *unitarizable*, being isomorphic to a continuous homomorphism $G \to \mathrm{U}_N$, where N is the dimension of the representation and U_N is the unitary group in N dimensions. This will be our first example of a Haar "averaging" argument, also referred to as Weyl's *unitary trick*.

Theorem 3.21. Let V be a real or complex FD representation of a compact group G. Then there exists a G-invariant inner product on V.

Proof. Let μ be the Haar probability measure on G, and fix *any* inner product $\langle -, - \rangle$ on V. Since V is an FD TVS, this inner product is compatible with the topology of V. Then for $v, w \in V$, the map $g \mapsto \langle g \cdot v, g \cdot w \rangle$ is continuous, and we can define

$$\langle v, w \rangle_G := \int_G \langle g \cdot v, g \cdot w \rangle \, d\mu(g). \tag{3.33}$$

One easily verifies that $\langle -, - \rangle_G$ defines a G-invariant inner product on V. Effectively, we have "averaged" an arbitrary inner product into a G-invariant one. □

3.3 The Peter-Weyl Theorem

In order to state any version of the Peter-Weyl Theorem, we need to discuss the space (in fact algebra) of *matrix coefficients* associated to a compact group G. Sometimes, these are called *representative functions* (hence the notation $R(G)$). As with subchapter 3.1, we base much of this discussion on notes due to [Mil].

Definition 3.22 [Matrix Coefficients]. For a compact group G, the set $R(G) \subseteq C(G)$ of *matrix coefficients* consists of all $\varphi : G \to \mathbb{C}$ which can be expressed in the form $\varphi(g) = \langle \alpha, g \cdot v \rangle$ for some $v \in V$ and $\alpha \in V^*$, where V is an FD complex G-rep.

Remark 3.23. Since all FD representations of a topological group are assumed continuous (unless stated otherwise), and every linear functional on an FD TVS is automatically continuous, each $\varphi \in R(G)$ is indeed continuous.

Note that $C(G)$ forms a Banach *algebra* when equipped with pointwise multiplication of operators. For each $\varphi \in C(G)$, we also have the functions $\bar{\varphi}$ and $\hat{\varphi}$ in $C(G)$ given by $\bar{\varphi}(g) := \overline{\varphi(g)}$ and $\hat{\varphi}(g) := \varphi(g^{-1})$ for all $g \in G$. The operations $\varphi \mapsto \bar{\varphi}$ and $\varphi \mapsto \hat{\varphi}$ define isometric involutions $C(G) \to C(G)$, referred to as *conjugation* and *inversion*, respectively. A subset $S \subseteq C(G)$ is called *conjugation* (resp. *inversion*) *invariant* if $\varphi \in S$ implies that $\bar{\varphi} \in S$ (resp. $\hat{\varphi} \in S$). We can similarly define isometric involutions of $L^p(G)$ in this way.

Theorem 3.24. $R(G)$ is a conjugation and inversion invariant (unital) subalgebra of $C(G)$.

Proof. Taking $V = k$ to be the trivial G-rep, $v = 1$, and $\alpha = \mathrm{id}_k$, we get $1 \in R(G)$. Let $\varphi, \psi \in R(G)$ and $c \in C(G)$. Let V and W be FD complex G-reps for which

$$\varphi(g) = \langle \alpha, g \cdot v \rangle \quad \text{and} \quad \psi(g) = \langle \beta, g \cdot w \rangle \quad \text{for all } g \in G, \tag{3.34}$$

where $v \in V$, $w \in W$, $\alpha \in V^*$, and $\beta \in W^*$ are fixed. Note $(c\varphi)(g) = \langle c\alpha, g \cdot v \rangle$, so that $c\varphi \in R(G)$. The functionals α and β induce $\gamma \in (V \oplus W)^*$ such that $\gamma(a+b) = \alpha(a) + \beta(b)$ for $a \in V$ and $b \in W$, in which case

$$(\varphi + \psi)(g) = \langle \gamma, g \cdot (v, w) \rangle. \tag{3.35}$$

Thus $\varphi + \psi \in R(G)$. Similarly, $\varphi\psi \in R(G)$ because we have an induced $\delta \in (V \otimes W)^*$ such that $\delta(a \otimes b) = \alpha(a)\beta(b)$ for all $a \in V$ and $b \in W$, so

$$(\varphi\psi)(g) = \langle \delta, g \cdot (v \otimes w) \rangle. \tag{3.36}$$

Having demonstrated that $R(G)$ is a subalgebra of $C(G)$, it remains to verify closure under conjugation and inversion. Let φ, V, v, and α be as before. For inversion invariance, we consider the evaluation map $\mathrm{ev}_v \in V^{**}$ defined by $\mathrm{ev}_v(\omega) = \omega(v)$, so that

$$\hat{\varphi}(g) = \langle \alpha, g^{-1} \cdot v \rangle = \langle \mathrm{ev}_v, g \cdot \alpha \rangle \tag{3.37}$$

implies $\hat{\varphi} \in R(G)$. As for conjugation invariance, we consider the *conjugate* representation $\overline{V}$ of V. As a vector space, $\overline{V}$ is the same underlying real TVS as V, but multiplication by i is $i \cdot v := -iv$, where omission of the $\cdot$ denotes scalar multiplication in V. Notice $\mathrm{End}(V) = \mathrm{End}(\overline{V})$, so that we naturally obtain a G-rep on $\overline{V}$ from any G-rep on V. Now $\bar{\alpha}(v) := \overline{\alpha(v)}$ is an element of $\overline{V}^*$, and then

$$\bar{\varphi}(g) = \langle \bar{\alpha}, g \cdot v \rangle. \tag{3.38}$$

Hence $\bar{\varphi} \in R(G)$, which completes the proof. $\square$

Corollary 3.25. $R(G)$ is equivalently the set of "right" matrix coefficients, i.e., functions $\varphi : G \to \mathbb{C}$ of the form $\varphi(g) = \langle \alpha, v \cdot g \rangle$ for some $v \in V$ and $\alpha \in V^*$, with V an FD *right* complex G-rep. In other words, $R(G) = R(G^{\text{opp}})$.

Proof. Since $R(G)$ is inversion invariant by Theorem 3.24, this is clear from Definition 3.22 and the correspondence between left and right representations (discussed at the beginning of the proof of Theorem 3.20). Note that $R(G^{\text{opp}})$ is precisely the set of "right" matrix coefficients, since left G-reps correspond with right G^{opp}-reps and vice versa. □

Another, very useful, characterization of $R(G)$ is as follows. Recall the definition of V_{fin} in the previous subchapter, and of the standard representations ρ and ℓ on $C(G)$.

Lemma 3.26. Let G be a compact group. Then $R(G) = C(G)_{\text{fin}}$, where $C(G)$ is equipped with either ρ or ℓ.

Proof. To differentiate $C(G)_{\text{fin}}$ with respect to ρ and ℓ, we refer to functions in the former space as *right-finite*, and those in the latter space as *left-finite*. We will show

$$\varphi \text{ left-finite} \implies \varphi \in R(G) \implies \varphi \text{ right-finite}.$$

The proof will then be complete upon replacing G by G^{opp} and applying Corollary 3.25.

First we prove that left-finite functions are matrix coefficients. To that end, let φ be left-finite on G. Suppose that $V \subseteq C(G)$ is an FD invariant subspace with respect to ℓ which contains φ. The evaluation map $\text{ev}_e : C(G) \to \mathbb{C}$ mapping $\psi \mapsto \psi(e)$ is linear and hence restricts to a linear functional α on V. Then simply note that $\varphi(g) = \langle \alpha, g^{-1} \cdot \varphi \rangle$. The latter function lies in $R(G)$ by inversion invariance.

It remains to show that matrix coefficients are right-finite. Let $\varphi \in R(G)$, written $\varphi(g) = \langle \alpha, g \cdot v \rangle$ for a G-rep V, $v \in V$, and $\alpha \in V^*$. We can associate to every $w \in V$ a function $T_w \in C(G)$ defined by $T_w(g) = \langle \alpha, g \cdot w \rangle$. These fit into a linear map $T : V \to C(G)$, $w \mapsto T_w$. Note that $\varphi = T(v) \in \operatorname{im} T$, and $\operatorname{im} T$ is FD because V is FD. But $T(w) \cdot g = T(g^{-1} \cdot w)$, so that $\operatorname{im} T$ is invariant with respect to ρ. Hence φ is right-finite. □

Assuming that $\dim \operatorname{Hom}_G(W_\pi, C(G)) < \infty$ for all $\pi \in \mathcal{I}_{\mathbb{C}}(G)$, we can apply Lemma 3.17 to obtain an isotypic decomposition of $R(G)$. In fact, we claim that $\operatorname{Hom}_G(W_\pi, C(G)) \cong W_\pi^*$ (as vector spaces, not G-reps), yielding linear isomorphisms

$$R(G) \cong \bigoplus_{\pi \in \mathcal{I}_{\mathbb{C}}(G)} (W_\pi \otimes W_\pi^*) \cong \bigoplus_{\pi \in \mathcal{I}_{\mathbb{C}}(G)} \operatorname{End}(W_\pi). \tag{3.39}$$

Furthermore, composing with the involution $\varphi \to \hat{\varphi}$ gives $R(G) = \oplus_{\pi \in \mathcal{I}_{\mathbb{C}}(G)} R_\pi(G)$, where $R_\pi(G) \cong \operatorname{End}(W_\pi)$ is the space of matrix coefficients with respect to W_π, i.e., functions $g \mapsto \langle \alpha, g \cdot w \rangle$ for $w \in W_\pi$ and $\alpha \in W_\pi^*$. This latter observation will be clear once we specify the isomorphism $\Psi_\pi : W_\pi^* \to \operatorname{Hom}_G(W_\pi, C(G))$.

Lemma 3.27. Let G be a compact group and $\pi \in \mathcal{I}_{\mathbb{C}}(G)$. For every $\alpha \in W_\pi^*$, define $\Psi_\pi(\alpha) : W_\pi \to C(G)$ by $\Psi_\pi(\alpha)(w)(g) = \langle \alpha, g^{-1} \cdot w \rangle$. Then

$$\Psi_\pi : W_\pi^* \to \mathrm{Hom}_G(W_\pi, C(G)), \quad \alpha \mapsto \Psi_\pi(\alpha) \tag{3.40}$$

is an isomorphism of vector spaces.

Proof. Certainly each $\Psi_\pi(\alpha)$ is linear. Given $w \in W$, $g, h \in G$, we have

$$\Psi_\pi(\alpha)(g \cdot w)(h) = \langle \alpha, (h^{-1}g) \cdot w \rangle = \Psi_\pi(\alpha)(w)(g^{-1}h) = (g \cdot \Psi_\pi(\alpha)(w))(h), \tag{3.41}$$

which verifies $\Psi_\pi(\alpha)(g \cdot w) = g \cdot \Psi_\pi(\alpha)(w)$. This ensures that $\Psi_\pi(\alpha) \in \mathrm{Hom}_G(W_\pi, C(G))$. The linearity of Ψ_π is clear. For injectivity, note that if $\alpha \neq 0$, then $\alpha(w) \neq 0$ for some $w \in W$, so $\Psi_\pi(\alpha)(w)(e) = \alpha(w) \neq 0$. Finally, we need to demonstrate surjectivity. Given $T \in \mathrm{Hom}_G(W_\pi, C(G))$, we can define $\alpha \in W_\pi^*$ by $\alpha(w) = T(w)(e)$. For $w \in W$, $g \in G$,

$$\Psi_\pi(\alpha)(w)(g) = T(g^{-1} \cdot w)(e) = (g^{-1} \cdot T(w))(e) = T(w)(g). \tag{3.42}$$

Thus $T = \Psi_\pi(\alpha)$, which completes the proof. □

In light of Lemma 3.27, we see that $R(G)$ is semisimple, and every W_π appears with multiplicity $d_\pi := \dim W_\pi$. The decomposition $R(G) = \oplus_{\pi \in \mathcal{I}_{\mathbb{C}}(G)} R_\pi(G)$ allows us to construct a basis for $R(G)$ as follows: for each $\pi \in \mathcal{I}_{\mathbb{C}}(G)$, fix a basis $e_1^\pi, \ldots, e_{d_\pi}^\pi$ of W_π. Then if we define

$$\pi_{ij}(g) = \langle (e_i^\pi)^*, g \cdot e_j^\pi \rangle \quad \text{for } g \in G, \tag{3.43}$$

where $1 \leq i, j \leq d_\pi$, the set of all π_{ij} as π ranges over $\mathcal{I}_{\mathbb{C}}(G)$ is a basis for $R(G)$. Choose $e_1^\pi, \ldots, e_{d_\pi}^\pi$ to be orthonormal with respect to a fixed G-invariant inner product $\langle -, - \rangle_\pi$ on W_π, so that

$$\pi_{ij}(g) = \langle e_i^\pi, g \cdot e_j^\pi \rangle_\pi. \tag{3.44}$$

Then for each $g \in G$, the $\pi_{ij}(g)$ are coefficients of a $d_\pi \times d_\pi$ unitary matrix. Ultimately, we will use the Peter-Weyl theorem to deduce that the π_{ij} form a Hilbert basis for $L^2(G)$. This is precisely the "convenient" Hilbert basis mentioned at the beginning of this chapter. Most of the difficulty lies in demonstrating that $R(G)$ is dense in $L^2(G)$, which will be one aspect of the Peter-Weyl theorem. However, at this point, we can deduce that the π_{ij} are orthogonal and compute the normalization constants needed to make this set ortho*normal*. The following theorem is often referred to as the *Schur orthogonality relations*.

Theorem 3.28 [Schur Orthogonality Relations]. If $\pi, \pi' \in \mathcal{I}_{\mathbb{C}}(G)$ are distinct, then $R_\pi(G) \perp R_{\pi'}(G)$. If $1 \leq i_1, j_1, i_2, j_2 \leq d_\pi$, then

$$\langle \pi_{i_1 j_1}, \pi_{i_2 j_2} \rangle = \frac{\delta_{i_1 i_2} \delta_{j_1 j_2}}{d_\pi}. \tag{3.45}$$

Proof. First we let $\pi, \pi' \in \mathcal{I}_{\mathbb{C}}(G)$, each equipped with a G-invariant inner product as in the preceding discussion. Fix $w \in W_\pi$ and $w' \in W_{\pi'}$. Then consider the map $\varphi : W_\pi \to W_{\pi'}$ given by

$$\varphi(v) = \int \pi'(g)w'\overline{\langle \pi(g)w, v\rangle_\pi}\, dg = \int \pi'(g)w'\langle v, \pi(g)w\rangle_\pi\, dg, \tag{3.46}$$

where integration on $W_{\pi'}$ is done component-wise (with respect to any basis). Now φ is certainly linear, and if $h \in G$, then the change of variables $g \mapsto hg$ yields

$$\varphi(\pi(h)v) = \int \pi'(g)w'\langle \pi(h)v, \pi(g)w\rangle_\pi\, dg = \pi'(h)\varphi(v). \tag{3.47}$$

Hence φ is G-linear. Now if $\alpha \in W_\pi^*$ and $\alpha' \in W_{\pi'}^*$ are represented by $v \in W_\pi$ and $v' \in W_{\pi'}$, respectively, with respect to the relevant inner products, then we get

$$\int \langle \alpha', \pi'(g)w'\rangle\overline{\langle \alpha, \pi(g)w\rangle}\, dg = \int \langle \pi'(g)w', v'\rangle_{\pi'}\overline{\langle \pi(g)w, v\rangle_\pi}\, dg = \langle \varphi(v), v'\rangle_{\pi'}. \tag{3.48}$$

If $\pi \neq \pi'$, then by Schur's lemma $\varphi = 0$, so (3.48) implies that $R_\pi(G) \perp R_{\pi'}(G)$.

If $\pi = \pi'$, then we can find $\lambda \in \mathbb{C}$ (possibly dependent on π, w, and w') for which φ is multiplication by λ. Then choosing v, w, v', w' to be basis vectors, from (3.48) we deduce

$$\langle \pi_{i_1 j_1}, \pi_{i_2 j_2}\rangle = \langle \varphi(e^\pi_{i_1}), e^\pi_{i_2}\rangle_\pi = \lambda\delta_{i_1 i_2}. \tag{3.49}$$

In particular, this inner product is 0 whenever $i_1 \neq i_2$. By either slightly tweaking φ or replacing π by its dual representation, we can also deduce that $\pi_{i_1 j_1} \perp \pi_{i_2 j_2}$ whenever $j_1 \neq j_2$. If $i_1 = i_2$ and $j_1 = j_2$, then we just need to show that $\lambda = 1/d_\pi$.

Fix $1 \le j \le d_\pi$, and let φ_j, λ_j be the instances of φ, λ corresponding to $w = w' = e^\pi_j \in W_\pi$. Then for each $1 \le i \le d_\pi$, taking $v = v' = e^\pi_i$ in (3.48) gives

$$\int |\pi_{ij}(g)|^2\, dg = \lambda_j. \tag{3.50}$$

If we sum over all $1 \le i \le d_\pi$, then because $\sum_i |\pi_{ij}(g)|^2 = 1$ for all $g \in G$, we get $1 = d_\pi\lambda_j$. Therefore $\lambda_j = 1/d_\pi$, which completes the proof. $\square$

Corollary 3.29. Let G be a compact group, and let V and W be two G-irreps with corresponding characters χ_V and χ_W. Then

$$\langle \chi_V, \chi_W\rangle = \dim \mathrm{Hom}_G(V, W) = \begin{cases} 1 & \text{if } V \cong W, \\ 0 & \text{otherwise.} \end{cases} \tag{3.51}$$

We are now ready to state the Peter-Weyl theorem. This is somewhat analogous to the Weierstrass theorem for $C([a, b])$, which also declares that an "algebraically defined" subset of continuous functions on a compact space (i.e., polynomials on $[a, b]$) is dense. Actually, we will make use of the Stone-Weierstrass theorem to deduce the $C(G)$ case from the $L^2(G)$ case, although one could also first prove the $C(G)$ case and then extend to $L^2(G)$ by density.

Theorem 3.30 [Peter-Weyl]. Let G be a compact group. Then $R(G)$ is dense in both $(C(G), \|\cdot\|)$ and $(L^2(G), \|\cdot\|_2)$.

Corollary 3.31. Let G be a compact group, and define the matrix coefficients π_{ij} as in (3.44). Then the set of π_{ij}, indexed by $\pi \in \mathcal{I}_{\mathbb{C}}(G)$ and $1 \leq i, j \leq d_\pi$, is a Hilbert basis for $L^2(G)$. An *orthonormal* Hilbert basis is given by $\sqrt{d_\pi}\pi_{ij}$.

Our main tool for proving Peter-Weyl is the notion of *convolution* on a compact group G. This is analogous to convolution on $\mathbb{R}^n$ as defined in the Appendix (B). In fact, both are instances of a more general construction that works on any locally compact Hausdorff group (so that the Haar measure exists).

Definition 3.32 [Convolution on G]. Let G be a compact group. Then if $\varphi \in C(G)$ and $f \in L^1(G)$, we define the *convolution* $\varphi * f$ on G as the function

$$(\varphi * f)(g) := \int \varphi(h)(h \cdot f)(g)\, dh = \int f(h)(\varphi \cdot h)(g)\, dh. \tag{3.52}$$

As with convolution on $\mathbb{R}^n$, we expect $\varphi * f$ to inherit the "best properties" from φ. For instance, $\varphi * f$ inherits continuity, which we now verify.

Lemma 3.33. Let $\varphi \in C(G)$ and $f \in L^1(G)$. Then $\varphi * f \in C(G)$.

Proof. Let $g_1, g_2 \in G$. Then we have

$$\begin{aligned} |(\varphi * f)(g_1) - (\varphi * f)(g_2)| &\leq \int |f(h)||\varphi(g_1 h^{-1}) - \varphi(g_2 h^{-1})|\, dh && (3.53)\\ &\leq \|f\|_1 \sup_{h \in G} |\varphi(g_1 h^{-1}) - \varphi(g_2 h^{-1})| && (3.54)\\ &= \|f\|_1 \omega_R(\varphi, \{g_1 g_2^{-1}\}). && (3.55) \end{aligned}$$

In particular, for every neighborhood U of the identity we have

$$\omega_R(\varphi * f, U) = \sup_{g_1 g_2^{-1} \in U} |(\varphi * f)(g_1) - (\varphi * f)(g_2)| \leq \|f\|_1 \omega_R(\varphi, U). \tag{3.56}$$

Because φ is right uniformly continuous, so is $\varphi * f$. $\square$

If $\varphi \in C(G)$, by Lemma 3.33 we obtain a linear operator $T_\varphi : L^1(G) \to C(G)$ via $T_\varphi(f) = \varphi * f$. Notice that $\|T_\varphi(f)\| \leq \|\varphi\|\|f\|_1$, so T_φ is continuous with operator norm at worst $\|\varphi\|$. In fact, T_φ is compact: if B is the closed unit ball in $L^1(G)$, then $T_\varphi(B)$ is uniformly bounded because T_φ is bounded, and equicontinuity is clear by (3.56), so Arzelà-Ascoli implies that $T_\varphi(B)$ is relatively compact.

Precomposing and postcomposing T_φ with the continuous linear embeddings $L^2(G) \hookrightarrow L^1(G)$ and $C(G) \hookrightarrow L^2(G)$ yields a compact operator $L^2(G) \to L^2(G)$, which we also refer to as T_φ (and this will be the implied meaning of T_φ from now on, unless stated otherwise). As an endomorphism of the Hilbert space $L^2(G)$, we would like to study the adjoint of T_φ. Define an isometric involution of $C(G)$ (or $L^p(G)$) with itself via $f \mapsto f^*$, where

$$f^* := \bar{\hat{f}} = \hat{\bar{f}} \implies f^*(g) = \overline{f(g^{-1})} \text{ for all } g \in G. \tag{3.57}$$

This leads us to an important lemma.

Lemma 3.34. Let $\varphi \in C(G)$. Then the adjoint T_φ^* of T_φ is T_{φ^*}.

Proof. Fix $f_1, f_2 \in L^2(G)$. Then by Fubini's theorem,

$$\langle T_\varphi f_1, f_2 \rangle = \int\int f_1(h)\varphi(gh^{-1})\overline{f_2(g)}\,dh\,dg = \int\int f_1(h)\varphi(gh^{-1})\overline{f_2(g)}\,dg\,dh. \tag{3.58}$$

If we pull out $f_1(h)$ from the first integral, we see that $\langle T_\varphi f_1, f_2\rangle = \langle f_1, \psi\rangle$, where

$$\psi(h) = \overline{\int \overline{f_2(g)}\varphi(gh^{-1})\,dg} = \int f_2(g)\overline{\varphi(gh^{-1})}\,dg = \int f_2(g)\varphi^*(hg^{-1})\,dg. \tag{3.59}$$

Therefore $\psi = T_{\varphi^*}(f_2)$, so that $T_\varphi^* = T_{\varphi^*}$, as desired. □

The last thing that we note is that T_φ commutes with ρ on $L^2(G)$. Indeed, for any $\varphi \in C(G)$, $f \in L^2(G)$, and $g, k \in G$ we have

$$T_\varphi(f \cdot k)(g) = \int \varphi(h) f(h^{-1}(gk^{-1}))\,dh = (T_\varphi f)(gk^{-1}) = (T_\varphi f \cdot k)(g). \tag{3.60}$$

Therefore each T_φ is ρ-invariant, so in particular it takes FD ρ-invariant subspaces of $L^2(G)$ (and hence of $C(G)$) to FD ρ-invariant subspaces. Since the image of T_φ is contained in $C(G)$, this means that $R(G)$ is invariant under T_φ by Lemma 3.26. We are now in a position to prove the $L^2(G)$ case of the Peter-Weyl theorem.

Proof. (Theorem 3.30, $L^2(G)$ case) It suffices to prove that $R(G)^\perp = \{0\}$ inside of $L^2(G)$. Fix $\varphi \in C(G)$, and consider the positive semidefinite self-adjoint operator

$$S_\varphi := T_\varphi T_\varphi^* = T_\varphi T_{\varphi^*}. \tag{3.61}$$

on $L^2(G)$. Then S_φ maps into $C(G)$, preserves $R(G)$, and by self-adjointness also preserves $R(G)^\perp$. In particular, S_φ restricts to a compact self-adjoint operator on $R(G)^\perp$, which is a closed subspace of $L^2(G)$ and hence a Hilbert space in its own right. If $S_\varphi|_{R(G)^\perp}$ is nonzero, it must have a nonzero eigenvalue λ with nonzero eigenspace E_λ. Now, $f \in E_\lambda$ implies that $f = \lambda^{-1}S_\varphi(f) \in C(G)$, and $\dim E_\lambda < \infty$ because S_φ is compact, so we are led to the contradiction $E_\lambda \subseteq R(G) \cap R(G)^\perp = \{0\}$.

Fix $f \in R(G)^{\perp}$. Then $S_\varphi f = 0$, so $T_{\varphi^*} f = 0$. As $T_{\varphi^*} f \in C(G)$, it is not just 0 a.e. but *everywhere*, in particular at the identity $e \in G$. Thus

$$\langle f, \varphi \rangle = \int f(h)\overline{\varphi(h)}\, dh = \int f(h)\varphi^*(h^{-1})\, dh = T_{\varphi^*} f(e) = 0. \tag{3.62}$$

As $\varphi \in C(G)$ was arbitrary, we conclude that $f \perp C(G)$. Since $C(G)$ is dense in $L^2(G)$, this means $f = 0$. Hence $R(G)^{\perp} = \{0\}$. $\square$

Since $R(G)$ is a unital *-subalgebra of $C(G)$ by Theorem 3.24, the density of $R(G)$ in $C(G)$ will follow from the Stone-Weierstrass theorem if $R(G)$ separates points. Indeed, we can prove this using the $L^2(G)$ case of the Peter-Weyl theorem, which have just proven.

Lemma 3.35. Let G be a compact group. Then $R(G)$ separates points.

Proof. Let $g \neq e$. Since G is compact and Hausdorff, it is normal. Choosing a neighborhood $U \ni e$ such that $\overline{U} \cap \overline{U}g^{-1} = \varnothing$, we can apply Urysohn's lemma to find $\varphi \in C(G)$ for which $\varphi|_{\overline{U}} \equiv 0$ and $\varphi|_{\overline{U}g^{-1}} \equiv 1$. Then

$$\|\varphi \cdot g - \varphi\|_2^2 \geq \int_U |\varphi(hg^{-1}) - \varphi(h)|^2\, dh = \int_U dh > 0, \tag{3.63}$$

by Lemma 3.9. In particular, $\rho(g)$ is not the identity on $L^2(G)$, so by the $L^2(G)$ case of Peter-Weyl, $\rho(g)$ cannot be the identity on $R(G)$. This holds for all $g \neq e$. Now let $g_1 \neq g_2$. Then $g_1^{-1}g_2 \neq e$, so we can find $\varphi \in R(G)$ for which

$$\varphi \cdot g_1^{-1}g_2 \neq \varphi \implies \varphi \cdot g_1^{-1} \neq \varphi \cdot g_2^{-1}. \tag{3.64}$$

Choose $h \in G$ such that $\varphi(hg_1) \neq \varphi(hg_2)$. Then $h^{-1} \cdot \varphi \in R(G)$ separates g_1 and g_2. $\square$

Using that $C(G)$ is dense in $L^p(G)$ for all $1 \leq p < \infty$, one can also deduce from the $C(G)$ case of the Peter-Weyl theorem that $R(G)$ is dense in $(L^p(G), \|\cdot\|_p)$. Again, we stress that as far as our exposition goes, this does *not* provide a proof of the $L^2(G)$ case, since we used that in order to prove the $C(G)$ case.

We end this subchapter with an application of Peter-Weyl to the study of *class* functions. If G is a group, then a *class* function on G is a function $f : G \to \mathbb{C}$ which is constant on conjugacy classes. That is, for all $g, h \in G$ we have $f(hgh^{-1}) = f(g)$. For instance, if V is an FD complex G-rep, then its character χ_V is a class function on G. The space of continuous class functions on G will be denoted $C_{\text{class}}(G) \subseteq C(G)$, and similarly we write $L^p_{\text{class}}(G) \subseteq L^p(G)$ (where a class function is understood in an a.e. sense, i.e., for almost every $g \in G$ we have $f(hgh^{-1}) = f(g)$ for all $h \in G$). Using the Haar averaging trick, we can "average" a function on G over conjugacy classes in order to obtain a class function.

More formally, we construct a linear operator

$$A : L^1(G) \to L^1(G), \quad A(f)(g) := \int f(hgh^{-1})\, dh. \tag{3.65}$$

If $1 \leq p < \infty$ and $f : G \to \mathbb{C}$ is measurable, note first that by Tonelli's theorem and the invariance of the Haar measure, we have

$$\int\int |f(hgh^{-1})|^p \, dh \, dg = \int\int |f(hgh^{-1})|^p \, dg \, dh = \int\int |f(g)|^p \, dg \, dh = \|f\|_p^p. \tag{3.66}$$

In particular, if $f \in L^p(G)$, then $h \mapsto f(hgh^{-1})$ is in $L^p(G)$ for a.e. $g \in G$. This ensures that $A(f)$ is well-defined in (3.65). If we further apply Jensen's inequality, then

$$\|A(f)\|_p^p \leq \int\int |f(hgh^{-1})|^p \, dh \, dg = \|f\|_p^p, \tag{3.67}$$

so that A restricts to a 1-Lipschitz map $L^p(G) \to L^p(G)$. In particular, $A(f) \in L^1(G)$ for all $f \in L^1(G)$. Observe as well that for every $g, h \in G$,

$$A(f)(hgh^{-1}) = \int f(khgh^{-1}k^{-1}) \, dk = \int f((kh)g(kh)^{-1}) \, dk = A(f)(g) \tag{3.68}$$

upon the change of variables $kh \mapsto k$. And certainly $A(f) = f$ for all class functions f. Therefore A projects $L^p(G)$ onto $L^p_{\text{class}}(G)$. In particular, $L^p_{\text{class}}(G)$ is closed in $L^p(G)$, for it is the kernel of the continuous linear map $\mathrm{id}_{L^p(G)} - A$. If we define $R_{\text{class}}(G) := A(R(G))$, then $R_{\text{class}}(G)$ is dense in $L^p_{\text{class}}(G)$ for every $1 \leq p < \infty$ by Peter-Weyl.

Lemma 3.36. Let G be a compact group. Then $R_{\text{class}}(G) = \operatorname{span}\{\chi_\pi \mid \pi \in \mathcal{I}_{\mathbb{C}}(G)\}$, where χ_π denotes the character of the irrep W_π. Each χ_π is called an *irreducible character* of G.

Proof. See the Appendix (C). □

Since the irreducible characters of G span a dense subspace of $L^2_{\text{class}}(G)$, which inherits a Hilbert space structure as a closed subspace of $L^2(G)$, from Corollary 3.29 we deduce

Corollary 3.37. If G is a compact group, then the irreducible characters of G form an orthonormal Hilbert basis for $L^2_{\text{class}}(G)$.

Corollary 3.38. If G is a first-countable compact group, then $\mathcal{I}_{\mathbb{C}}(G)$ is countable. This applies to all compact Lie groups.

Proof. By the Birkhoff-Kakutani theorem, every first-countable compact group is Polish and therefore second-countable [MZ55, Section 1.22], so $L^2(G)$ is separable by Theorems A.1 and A.2. Hence $L^2_{\text{class}}(G)$ is also separable. By Corollary 3.37, $\mathcal{I}_{\mathbb{C}}(G)$ must be countable. Of course, if G is a compact Lie group, we already know it is second-countable. □

It's also true that A continuously projects $C(G)$ onto $C_{\text{class}}(G)$, so that $R_{\text{class}}(G)$ is dense in $C_{\text{class}}(G)$. Assuming A maps continuous functions to continuous functions, certainly A is 1-Lipschitz as a map $C(G) \to C_{\text{class}}(G)$, for if $\varphi \in C(G)$ and $g \in G$ then

$$|A(\varphi)(g)| \leq \int |\varphi(hgh^{-1})|\, dh \leq \|\varphi\| \implies \|A(\varphi)\| \leq \|\varphi\|. \tag{3.69}$$

Justifying that $A(\varphi) \in C(G)$ for all $\varphi \in C(G)$ is a bit subtle, however. If $\varepsilon > 0$ and $A \subseteq G$, then for all $g_1 g_2^{-1} \in A$ we have

$$|A(\varphi)(g_1) - A(\varphi)(g_2)| \leq \int |\varphi(hg_1h^{-1}) - \varphi(hg_2h^{-1})|\, dh \leq \omega_R(\varphi, B), \tag{3.70}$$

where $B = \cup_{h \in G} hAh^{-1}$. By right uniform continuity we can find a neighborhood U of the identity e for which $\omega_R(\varphi, U) < \varepsilon$. Assuming that we can find a neighborhood $e \in V \subseteq U$ such that $hVh^{-1} \subseteq U$ for all $h \in G$, we are done. A topological group with the property that e contains a local basis of conjugation invariant neighborhoods is called SIN (*small invariant neighborhood*). Indeed, every compact group is SIN.

Theorem 3.39. Every compact group G is SIN.

Proof. See [GM67, Theorem 2.2]. □

When G is finite, every function $f : G \to \mathbb{C}$ is smooth, continuous, $L^p(G)$, and whatnot. We can identify $L^2(G)$ with the vector space $\mathbb{C}[G]$, equipped with the inner product

$$\langle f_1, f_2 \rangle = \frac{1}{|G|} \sum_{g \in G} f_1(g)\overline{f_2(g)}. \tag{3.71}$$

The subspace $L^2_{\text{class}}(G)$ is just the set of all class functions $f : G \to \mathbb{C}$, whose dimension is the class number of G. If C_G denotes the set of conjugacy classes of G, then $(L^2_{\text{class}}(G), \langle -, - \rangle)$ is naturally isomorphic to $\mathbb{C}[C_G]$, equipped with the inner product

$$\langle f_1, f_2 \rangle_C := \sum_{c \in C_G} \frac{|c|}{|G|} f_1(c)\overline{f_2(c)}. \tag{3.72}$$

Corollary 3.37 tells us that $|\mathcal{I}_{\mathbb{C}}(G)|$ is also the class number of G. In particular, there are only finitely many FD complex G-irreps up to isomorphism. Moreover, we can classify these irreps via the *character table*: the columns of this table are the conjugacy classes $c \in C_G$, the rows are the isomorphism classes $\pi \in \mathcal{I}_{\mathbb{C}}(G)$, and the (π, c)th entry is $\chi_\pi(c)$. With respect to the inner product (3.72), the rows of the character table are orthonormal.

Notice that $\chi_{U \oplus W} = \chi_U + \chi_W$ and $\chi_{U \otimes W} = \chi_U \chi_W$ for any two FD G-reps U and W. In particular, if V is an FD complex G-rep, and the multiplicity of π in V is a_π, we have $\chi_V = \sum a_\pi \chi_\pi$, so that up to isomorphism, V is completely determined by its character χ_V. One can, for example, apply this to the regular (permutation) representation of G on $\mathbb{C}[G]$ to deduce that

$$|G| = \sum_{\pi \in \mathcal{I}_{\mathbb{C}}(G)} (\dim W_\pi)^2 = \sum_\pi d_\pi^2. \tag{3.73}$$

3.4 Irreps of Connected Compact Lie Groups

With the Peter-Weyl theorem at our disposal, we have essentially reduced Fourier analysis on a compact group G to the classification of FD complex G-irreps. What can we say about $\mathcal{I}_{\mathbb{C}}(G)$? Or the dimensions of G-irreps? Their characters? These are all important questions if we hope to extract information from our orthonormal basis of matrix coefficients, necessary for Chapter 4. In order to obtain a satisfying answer to these questions, we specialize to connected compact Lie groups. There is a very long and detailed story behind this classification, which we lack the time or space to do justice, but beautiful answers to all of our questions nonetheless exist. The most important results are the *theorem of highest weight* and *Weyl's integral, character, and dimension formulae*. In this subchapter, we outline the classification by highest weight for both complex semisimple Lie algebras and connected compact Lie groups, along with Weyl's character and dimension formulae. Many proofs and details will be omitted, so we recommend [Hal15] and [Kna86] as references.

Fix a Lie group G. We begin by recalling some notions from the theory of Lie groups, all of which can be found in [Lee13]. A *Lie subgroup* H of G is a subgroup of G with a smooth structure making H into both an immersed submanifold of G and a Lie group. Importantly, every embedded Lie subgroup of G is closed, and conversely, by Cartan's closed subgroup theorem every closed subgroup H of G can be given a unique smooth structure such that H becomes an embedded Lie subgroup of G. All Lie subgroups of G are weakly embedded, so that if $\varphi : G \to G'$ is a homomorphism of Lie groups with image contained in a Lie subgroup of H', then the restriction $\varphi : G \to H'$ is smooth.

If N is closed normal subgroup of G, then by the quotient manifold theorem, the group G/N can be given a unique smooth structure such that the quotient map $q : G \to G/N$ becomes a homomorphism of Lie groups. More generally, if H is a closed subgroup of G then the set of (left or right) cosets G/H carries a unique smooth structure for which the projection $\pi : G \to G/H$ is a principal H-bundle, which is important in the study of *homogeneous spaces*. Because the kernel of any Lie group homomorphism is a closed normal subgroup of the domain, the first isomorphism theorem for groups has an analog in the category of Lie groups: if $\varphi : G \to G'$ is a surjective Lie group homomorphism, then one has an isomorphism of Lie groups $G/\ker(\varphi) \cong G'$.

Given a covering map $\pi : \tilde{G} \to G$ of connected spaces, where $e \in G$ is the identity of the Lie group G, for every $\tilde{e} \in \pi^{-1}(e)$ there is a unique Lie group structure on $\tilde{G}$ with respect to which $\tilde{e}$ is the identity and π is a Lie group homomorphism. In particular, every connected Lie group G has a unique (up to isomorphism) universal covering group, i.e., a simply connected Lie group $\tilde{G}$ covering G. Regardless of whether G is connected, we can consider the connected component G_e of G which contains the identity, called the *identity component* of G. Since G is locally path connected, its connected and path components agree. The set $\pi_0(G)$ of all components of G is readily equipped with a group structure for which G_e is the identity. When studying an arbitrary Lie group G, we can sometimes reduce to the connected case by restricting our attention to G_e, and even further to the simply connected case by passing to the universal covering group $\tilde{G}_e$ of G_e.

Example 3.40 [O_n, SO_n, and $\mathrm{Spin}(n)$]. The orthogonal group O_n is a compact Lie group with two connected components, based on the determinant ± 1. The identity component of O_n is the connected compact Lie group SO_n. If $n = 1$, then SO_1 is the trivial group, which is connected. If $n = 2$, then SO_2 is isomorphic to the circle group $\mathbb{T}$, which has fundamental group $\mathbb{Z}$ and universal cover $\mathbb{R}$. If $n \geq 3$, then $\pi_1(\mathrm{SO}_n) \cong \mathbb{Z}/2$. Therefore, the universal cover of SO_n is a compact double cover when $n \geq 3$, referred to as the *spin group* $\mathrm{Spin}(n)$.

Associated to every Lie group G is its Lie algebra $\mathfrak{g} = \mathrm{Lie}(G)$, the space of left-invariant vector fields on G. $\mathfrak{g}$ is a subalgebra of $(\mathfrak{X}(G), [-,-])$, the Lie algebra of all vector fields on G equipped with the Lie bracket. By evaluation at the identity $e \in G$, we obtain an isomorphism $\mathfrak{g} \cong T_eG$, where T_eG is the tangent space to G at e. In particular, $\dim \mathfrak{g} = \dim G$. Every homomorphism $\varphi : G \to G'$ of Lie groups induces a homomorphism of their corresponding Lie algebras $\dot{\varphi} : \mathfrak{g} \to \mathfrak{g}'$ in a functorial manner, where $\dot{\varphi}$ is identified with the differential $d\varphi_e : T_eG \to T_{e'}G'$ at the identity. Importantly, the diagram

$$\begin{tikzcd} \mathfrak{g} \arrow[r, "\dot{\varphi}"] \arrow[d, "\exp"'] & \mathfrak{g}' \arrow[d, "\exp"] \\ G \arrow[r, "\varphi"] & G' \end{tikzcd} \tag{3.74}$$

commutes, where $\exp : \mathfrak{g} \to G$ is the exponential map [Lee13, Chapter 20]. If H is a Lie subgroup of G, then the inclusion $\iota : H \hookrightarrow G$ gives rise to an embedding of Lie algebras $i : \mathfrak{h} \hookrightarrow \mathfrak{g}$. This allows us to view $\mathfrak{h}$ as a subalgebra of $\mathfrak{g}$, so by (3.74) $\exp : \mathfrak{h} \to H$ becomes a restriction of $\exp : \mathfrak{g} \to G$.

Example 3.41 [Matrix Groups]. Let V be an FD vector space over $k \in \{\mathbb{R}, \mathbb{C}\}$. Then the Lie group $\mathrm{GL}(V)$ is an open subset of the vector space $\mathrm{End}(V)$, allowing us to identify the corresponding Lie algebra $\mathfrak{gl}(V)$ with $\mathrm{End}(V)$. Under this identification, the Lie bracket is simply the commutator of linear operators, while the exponential map $\exp : \mathrm{GL}(V) \to \mathfrak{gl}(V)$ is the usual exponential map on Banach algebras, coming from the Taylor series

$$e^A = I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \cdots \tag{3.75}$$

If G is a Lie subgroup of $\mathrm{GL}(V)$ (i.e., a matrix group), we also view $\mathfrak{g}$ as a subalgebra of $\mathrm{End}(V)$. Of course, when $V = k^n$, we write $\mathrm{GL}_n k$ and $\mathfrak{gl}_n(k)$ rather than $\mathrm{GL}(k^n)$ or $\mathfrak{gl}(k^n)$ and view linear operators as $n \times n$ matrices in $M_n k \cong \mathrm{End}(k^n)$.

One can show, using the Peter-Weyl theorem, that every compact Lie group is isomorphic to a matrix group, for example because Lie groups have "no small subgroups" [Mil, Theorem 4.3.2]. In Chapter 4 we will be most interested in O_n, SO_n, and $\mathrm{Spin}(n)$, whose common Lie algebra is $\mathfrak{so}_n$. As a subalgebra of $M_n\mathbb{R}$, we can realize $\mathfrak{so}_n$ as the space of skew-symmetric $n \times n$ matrices, equipped with the commutator Lie bracket.

There are many important correspondences between a Lie group and its Lie algebra, which we summarize with the following theorem.

Theorem 3.42. Let G be a Lie group with $\mathfrak{g} = \mathrm{Lie}(G)$. Then every subalgebra $\mathfrak{h} \subseteq \mathfrak{g}$ is the Lie algebra of a unique connected Lie subgroup of G. If G is connected, then under this correspondence (connected) normal subgroups correspond to ideals.

Proof. See [Lee13, Theorem 19.26] and [Lee13, Theorem 20.28]. □

We already see from Theorem 3.42 that working with only *connected* Lie groups can simplify the picture. When it comes to studying representations of Lie groups, there are other reasons for restricting our attention to connected, or even simply connected, groups. For instance, by differentiation, we know how to pass from Lie group homomorphisms $\varphi : G \to G'$ to Lie algebra homomorphisms $\dot{\varphi} : \mathfrak{g} \to \mathfrak{g}'$, but when can this be reversed? In other words, when can we "integrate" a Lie algebra homomorphism $\dot{\varphi} : \mathfrak{g} \to \mathfrak{g}'$ to a Lie group homomorphism with $\dot{\varphi}$ as its differential?

Theorem 3.43. Let G and G' be Lie groups with corresponding Lie algebras $\mathfrak{g}$ and $\mathfrak{g}'$, and let $\dot{\varphi} : \mathfrak{g} \to \mathfrak{g}'$ be a Lie algebra homomorphism. Then if G is connected, there can be at most one $\varphi : G \to G'$ whose differential is $\dot{\varphi}$. If G is also simply connected, such a homomorphism φ exists.

Proof. See [Fre, Lecture 7] for a beautiful proof using the Frobenius theorem for foliations and the Maurer-Cartan equation. □

Corollary 3.44. If G and G' are simply connected Lie groups with isomorphic Lie algebras, then $G \cong G'$.

Given a real or complex FD representation of a Lie group G on a vector space V, which is equivalently a Lie group homomorphism $\rho : G \to \mathrm{GL}(V)$, we can differentiate to obtain a *real* Lie algebra homomorphism $\dot{\rho} : \mathfrak{g} \to \mathfrak{gl}(V)$. If V is a complex representation, then we can *complexify* $\dot{\rho}$ to obtain a complex homomorphism $\mathfrak{g}^{\mathbb{C}} \to \mathfrak{gl}(V)$, where $\mathfrak{g}^{\mathbb{C}} := \mathfrak{g} \otimes_{\mathbb{R}} \mathbb{C} = \mathfrak{g} \oplus i\mathfrak{g}$, by mapping $\xi + i\eta \mapsto \dot{\rho}(\xi) + i\dot{\rho}(\eta)$. This leads us to a general definition.

Definition 3.45. Let $\mathfrak{g}$ be a Lie algebra over a field k. Then a *representation* of $\mathfrak{g}$ over k, also called a $\mathfrak{g}$-rep, is a homomorphism $\dot{\rho} : \mathfrak{g} \to \mathfrak{gl}(V)$ of Lie algebras, where V is a k-vector space, the *representation space*. We can likewise form a category $\mathfrak{g}\text{-}\mathsf{Rep}_k$ of $\mathfrak{g}$-reps. As before, $\mathfrak{gl}(V)$ is just the associative algebra $\mathrm{End}(V)$ equipped with the commutator as its Lie bracket.

Remark 3.46. As with representations of groups, we may also abuse notation and refer to V, rather than $\dot{\rho}$, as the $\mathfrak{g}$-rep.

Definition 3.45 allows us to say that a real representation of a Lie group G on an FD vector space V induces a real representation of its Lie algebra $\mathfrak{g}$ on V. Similarly, a complex G-rep on V induces a complex $\mathfrak{g}^{\mathbb{C}}$-rep on V. If G is simply connected, then both of these correspondences are bijective by Theorem 3.43, with functorial inverse. Many definitions and results about representations of groups in subchapter 3.2 have analogs in the Lie algebra setting. For instance, we say that a subspace $W \subseteq V$ of a $\mathfrak{g}$-rep V is *invariant* whenever $\mathfrak{g} \cdot W \subseteq W$, which also leads to notions of subrepresentations and (ir)reducibility for $\mathfrak{g}$-reps. Differentiation preserves invariant subspaces, so if ρ is an FD G-irrep, $\dot{\rho}$ is a $\mathfrak{g}$-irrep. If G is connected, then the converse is also true.

Lemma 3.47. Let $\rho : G \to \mathrm{GL}(V)$ be a real/complex FD G-rep, where G is a connected Lie group. Then $W \subseteq V$ is ρ-invariant if and only if it is $\dot{\rho}$-invariant. In particular, ρ is irreducible if and only if $\dot{\rho}$ is irreducible.

Proof. We do this for the real case (and the complex case follows readily). Let $\mathfrak{g} = \mathrm{Lie}(G)$. It suffices to prove that $\dot{\rho}$-invariance implies ρ-invariance, so let $W \subseteq V$ be $\dot{\rho}$-invariant. By [Lee13, Propositions 7.14(c), 20.8(f)], we need only show that $\rho(e^{\xi})w \in W$ for all $\xi \in \mathfrak{g}$ and $w \in W$. Note

$$\rho(e^{\xi})w = e^{\dot{\rho}(\xi)}w = \sum_{n=0}^{\infty} \frac{\dot{\rho}(\xi)^n}{n!} w \tag{3.76}$$

by (3.74) and Example 3.41. Since $\dot{\rho}(\xi)^n w \in W$ for all $n \geq 0$, this completes the proof. $\square$

Let G be a simply connected Lie group with $\mathfrak{g} = \mathrm{Lie}(G)$. If we write $\mathcal{I}_{\mathbb{C}}(\mathfrak{g}^{\mathbb{C}})$ for the set of isomorphism classes of FD $\mathfrak{g}^{\mathbb{C}}$-reps, then there is a natural bijection between $\mathcal{I}_{\mathbb{C}}(G)$ and $\mathcal{I}_{\mathbb{C}}(\mathfrak{g}^{\mathbb{C}})$ (and likewise between $\mathcal{I}_{\mathbb{R}}(G)$ and $\mathcal{I}_{\mathbb{R}}(\mathfrak{g})$). Therefore, we can classify the complex irreps of any simply connected Lie group by classifying the irreps of its complexified Lie algebra. If G is not simply connected but connected, then we can pass to the universal covering group $\tilde{G}$, and $\mathrm{Lie}(\tilde{G}) \cong \mathfrak{g}$. Any G-rep can be restricted to a $\tilde{G}$-rep along the cover $\tilde{G} \to G$, and this restricted representation is irreducible if and only if the original G-rep is (since the covering map is surjective). However, there may be some $\tilde{G}$-reps which do not arise in this way, although those that do are the restriction of a *unique* G-rep. Thus, one approach to classifying the FD irreps of a connected Lie group G is to first classify those of $\mathfrak{g}^{\mathbb{C}}$, hence $\tilde{G}$, and then to specify which of these irreps actually descend to G-irreps. Importantly, if $\mathfrak{g}^{\mathbb{C}}$, or equivalently $\mathfrak{g}$, is semisimple (as a Lie algebra), we call G *semisimple*, and the $\mathfrak{g}^{\mathbb{C}}$-irreps can be classified via the *theorem of highest weight*. If furthermore G is compact, then the highest weights corresponding to G-irreps are the *analytically integral* ones. For a general connected compact Lie group G, both $\mathfrak{g} = \mathrm{Lie}(G)$ and $\mathfrak{g}^{\mathbb{C}}$ are reductive, meaning they split into abelian and semisimple parts, but an analogous *theorem of highest weight* still holds. All relevant terminology in this paragraph will be defined shortly.

The classification picture on both a Lie group and its (complexified) Lie algebra revolves around studying the natural *adjoint* representation. For a Lie group G, this is a representation on $\mathfrak{g} = \mathrm{Lie}(G)$ induced by conjugation: for each $g \in G$, we define the conjugation map $\mathrm{AD}_g : G \to G$ by $h \mapsto ghg^{-1}$, and then $\mathrm{Ad}_g := \dot{\mathrm{AD}}_g : \mathfrak{g} \to \mathfrak{g}$ is its differential. As a result, we obtain a G-rep on $\mathfrak{g}$

$$\mathrm{Ad} : G \to GL(\mathfrak{g}), \quad g \mapsto \mathrm{Ad}_g. \tag{3.77}$$

Differentiating leads to a $\mathfrak{g}$-rep $\mathrm{ad} : \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g})$, which can also be complexified into a $\mathfrak{g}^{\mathbb{C}}$-rep on $\mathfrak{g}^{\mathbb{C}}$. One can show by direct computation that $\mathrm{ad}_\xi(\eta) = [\xi, \eta]$. Hence, more generally, for any Lie algebra $\mathfrak{g}$ over a field k we define the adjoint representation

$$\mathrm{ad} : \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g}), \quad \xi \mapsto [\xi, -]. \tag{3.78}$$

That ad is a representation follows from the Jacobi identity.

Example 3.48 [Matrix Groups]. If $G \subseteq \mathrm{GL}(V)$ is a matrix group, then upon identifying $\mathfrak{g} = \mathrm{Lie}(G)$ as a subalgebra of matrices, the adjoint representation of G on $\mathfrak{g}$ is simply conjugation, i.e., $\mathrm{Ad}_A(B) = ABA^{-1}$.

We now briefly describe classification by highest weight for irreps of an FD complex semisimple Lie algebra $\mathfrak{g}$, before moving on to the case of a connected compact Lie group. Familiarity with the classical theory of *root systems*, as detailed in [Hal15, Chapter 8], will be assumed. To begin, fix a *Cartan subalgebra* (CSA) $\mathfrak{h} \subseteq \mathfrak{g}$—that is, a maximal abelian subalgebra of $\mathfrak{g}$ such that $\mathrm{ad}(\xi)$ is diagonalizable for every $\xi \in \mathfrak{h}$. It is a classical result that $\mathfrak{g}$ has a CSA, all of which are conjugate by an automorphism of $\mathfrak{g}$. In particular, $\dim \mathfrak{h}$, called the *rank* of $\mathfrak{g}$, is independent of $\mathfrak{h}$. For a $\mathfrak{g}$-rep $\dot{\rho} : \mathfrak{g} \to \mathfrak{gl}(V)$, a functional $\lambda \in \mathfrak{h}^*$ is said to be a *weight* of V when the corresponding *weight space*

$$V_\lambda = \{v \in V \mid \dot{\rho}(\xi)v = \lambda(\xi)v \;\text{ for all }\; \xi \in \mathfrak{h}\} \tag{3.79}$$

is nonzero. In other words, weights are simultaneous eigenvalues for the operators $\dot{\rho}(\xi)$, $\xi \in \mathfrak{h}$, while weight spaces are simultaneous eigenspaces. The nonzero weights of the adjoint representation are called *roots*, the set of which we denote Φ. It is immediate from the definition of a CSA that we obtain a simultaneous eigendecomposition

$$\mathfrak{g} = \mathfrak{g}_0 \oplus \Big(\bigoplus_{\alpha \in \Phi} \mathfrak{g}_\alpha\Big). \tag{3.80}$$

One can show that $\mathfrak{g}_0 = \mathfrak{h}$, that $\dim \mathfrak{g}_\alpha = 1$ for all $\alpha \in \Phi$, and that Φ forms a root system for $E = \mathrm{span}_{\mathbb{R}}(\Phi)$ with respect to the inner product induced by the *Killing form*. The Killing form is the ad-invariant symmetric bilinear form $\kappa(\xi, \eta) = \mathrm{tr}(\mathrm{ad}(\xi) \circ \mathrm{ad}(\eta))$ on $\mathfrak{g}$, which is nondegenerate by Cartan's criterion. Hence, κ induces an isomorphism $\mathfrak{h} \cong \mathfrak{h}^*$, carrying $-\kappa$ to a (real) inner product $(-, -)$ on E. One can show that weights of FD $\mathfrak{g}$-reps lie in E and are in fact integral with respect to Φ [Hal15, Proposition 9.2]. Fixing a base Δ for Φ, or equivalently a set of positive roots Φ^+, this allows us to say that a weight λ of a $\mathfrak{g}$-rep is *highest* whenever it is higher than all other weights (with respect to Δ).

Theorem 3.49 [Theorem of Highest Weight]. Let $\mathfrak{g}$, $\mathfrak{h}$, Φ, Δ, and $(E, (-,-))$ be as in the previous discussion. Then

1. Every FD $\mathfrak{g}$-irrep has a (unique) highest weight, and this highest weight is necessarily dominant integral.
2. Two FD $\mathfrak{g}$-irreps with the same highest weight are isomorphic.
3. Every dominant integral element of E is the highest weight of some FD $\mathfrak{g}$-irrep.

Proof. See [Hal15, Theorems 9.4, 9.5]. The most difficult statement to prove is (3), which usually involves the construction of *Verma modules*. $\square$

Parameterizing FD $\mathfrak{g}$-irreps by their highest weight allows us to study these irreps in light of the rich geometry of root systems. (All irreducible root systems may be classified in terms of their Dynkin diagrams [Hal15, Theorem 8.49].) The *Weyl character formula*, which we state shortly, is a prime example of root geometry translating into algebraic information. For an FD $\mathfrak{g}$-irrep $\dot{\pi}$, the character $\dot{\chi}_\pi : \mathfrak{h} \to \mathbb{C}$ of $\dot{\pi}$, by analogy with the character of a Lie group representation, is defined as $\dot{\chi}_\pi(\xi) = \operatorname{tr}(e^{\dot{\pi}(\xi)})$ for all $\xi \in \mathfrak{h}$. In particular, $\dot{\chi}_\pi(0)$ is the dimension of $\dot{\pi}$. Recall that the *Weyl vector* $\delta \in E$ is half of the sum of the positive roots. Also, the Weyl group W acts orthogonally on E, so the determinant $\varepsilon(w)$ of $w \in W$, viewed as an action on E, is always ± 1.

Theorem 3.50 [Weyl Character Formula]. Assume the same setup as Theorem 3.49. Let W be the Weyl group, δ the Weyl vector, and define $\varepsilon(w) \in \{\pm 1\}$ for each $w \in W$ as in the preceding paragraph. Then if $\dot{\pi}$ is an FD $\mathfrak{g}$-rep with highest weight λ,

$$\dot{\chi}_\pi(\xi) = \frac{\sum_{w \in W} \varepsilon(w) e^{\langle w \cdot (\lambda + \delta), \xi \rangle}}{\sum_{w \in W} \varepsilon(w) e^{\langle w \cdot \delta, \xi \rangle}} \quad \text{for all } \xi \in \mathfrak{h}. \tag{3.81}$$

Here $\langle -, - \rangle$ denotes the dual pairing (i.e., evaluation) $\mathfrak{h}^* \times \mathfrak{h} \to \mathbb{C}$.

Proof. See [Hal15, Theorem 10.14]. $\square$

Corollary 3.51 [Weyl Dimension Formula]. Assuming the setup of Theorem 3.50, we have

$$\dim \dot{\pi} = \frac{\prod_{\alpha \in \Phi^+} (\lambda + \delta, \alpha)}{\prod_{\alpha \in \Phi^+} (\delta, \alpha)}. \tag{3.82}$$

Proof. See [Hal15, Theorem 10.18]. Basically, one applies the Weyl Character formula to $\xi = 0$ and plays around with determinants. $\square$

The classification on a connected compact group G is similar. One chooses a *maximal torus* $T \subseteq G$, and then the complexified Lie algebra $\mathfrak{t}^{\mathbb{C}}$ of $\mathfrak{t} = \mathrm{Lie}(T)$ is a CSA of $\mathfrak{g}^{\mathbb{C}}$. If G is semisimple, then we can apply Theorem 3.49 to classify all $\mathfrak{g}^{\mathbb{C}}$-FD irreps, but to account for the irreps of the universal cover which may not descend, statement (3) in Theorem 3.49 has to be refined slightly: those dominant integral weights which correspond to FD G-irreps are precisely the *analytically integral* ones. These are the weights in $(\mathfrak{t}^{\mathbb{C}})^*$ whose restriction to $\mathfrak{t}$ lies in a special lattice called the *character lattice* $\Lambda \subseteq \mathfrak{t}^*$ of T. When G is not semisimple, we have to be a bit more careful to account for the center, but since we will only make use of this classification for the semisimple Lie groups SO_n $(n \geq 3)$ in Chapter 4, we defer to [Hal15, Chapters 11-12] for a full discussion. The relevant theorem is [Hal15, Theorem 12.6]. We will end this chapter with an overview of maximal tori and character lattices, which is enough to understand our definition of "analytically integral." That will complete the classification of G-irreps, at least in the semisimple case, modeled off of Theorem 3.49.

To motivate maximal tori, recall that our ultimate goal is to obtain a useful parameterization of the countable set $\mathcal{I}_{\mathbb{C}}(G)$ for G a (connected) compact Lie group. As a general fact, if k is an algebraically closed field and G is any group, then n-dimensional G-irreps are in bijection with homomorphisms $G \to \mathrm{GL}_n k$. Indeed, we already know that all n-dimensional G-reps are isomorphic to G-reps on k^n, but by Schur's lemma, any two isomorphic G-reps on a fixed FD vector space V must be identical (since an isomorphism between them is necessarily of the form $\lambda \cdot \mathrm{id}_V$ for some $\lambda \neq 0$). Therefore, if G is abelian, then G-irreps are in bijection with homomorphisms $G \to \mathrm{GL}_1(k) \cong k^\times$ by Corollary 3.14. In particular, irreps of abelian groups over algebraically closed fields are uniquely determined by their character. When $k = \mathbb{C}$ and G is a topological group, the same results hold, so long as we append "continuous" appropriately. If G is a compact group, then by the extreme value theorem one may deduce that all continuous homomorphisms $G \to \mathbb{C}^*$ have image lying in S^1. Thus, for a compact abelian group G (not necessarily a Lie group), there is a natural bijection between $\mathcal{I}_{\mathbb{C}}(G)$ and the *Pontryagin dual* $\hat{G} = \mathrm{Hom}(G, \mathbb{T})$, which is the group (under pointwise multiplication) of continuous homomorphisms $G \to \mathbb{T}$. Here $\mathbb{T}$ denotes the circle group, as in Example 3.8.

Theorem 3.52. Let G be a connected abelian Lie group. Then G is isomorphic to the product of an FD real vector space and a torus (a finite product $\mathbb{T} \times \cdots \times \mathbb{T}$ of circle groups). In particular, if G is compact, then it is isomorphic to a torus.

Proof. Let $\mathfrak{g} = \mathrm{Lie}(G)$. Since G is abelian, $\exp : \mathfrak{g} \to G$ is a homomorphism by, e.g., [Lee13, Corollary 20.11]. Because G is connected and $\exp(\mathfrak{g})$ is a subgroup of G containing a neighborhood of the identity, exp must be surjective [Lee13, Propositions 7.14(c), 20.8(f)]. Thus, by the first isomorphism theorem for Lie groups, we have $G \cong \mathfrak{g}/\ker(\exp)$. Now $\ker(\exp)$ is 0-dimensional since $\dim \mathfrak{g} = \dim G$, so as a discrete subgroup of the vector space $\mathfrak{g}$, it must be a lattice. If $\xi_1, \ldots, \xi_r$ is a basis for this lattice, we can complete it to a basis $\xi_1, \ldots, \xi_n$ of $\mathfrak{g}$, inducing an isomorphism $G \cong \mathbb{T}^r \times \mathbb{R}^{n-r}$. $\square$

In light of Theorem 3.52, we refer to any connected compact abelian Lie group T as a *torus group*. A *maximal torus* of a compact Lie group G is a torus subgroup $T \subseteq G$ such that, if T' is another torus group and $T \subseteq T' \subseteq G$, then $T = T'$. Every maximal torus turns out to be maximal abelian: if A is an abelian Lie subgroup of G and $T \subseteq A \subseteq G$, then $T = A$. It follows that $\mathfrak{t}^{\mathbb{C}}$ is a CSA of $\mathfrak{g}^{\mathbb{C}}$, where $\mathfrak{t} = \mathrm{Lie}(T)$, by Theorem 3.42 and the result that a connected Lie group is abelian if and only if its Lie algebra is abelian (the latter of which can be deduced from the Baker-Campbell-Hausdorff formula).

The idea behind studying maximal tori is that one can classify irreps of a torus quite naturally via its associated character lattice. If G is a connected compact Lie group, we can find a maximal torus, study the irreducible decomposition of the adjoint G-rep restricted to T, and bootstrap our way up from there. While a massive oversimplification, this is one approach to arrive at the theorem of highest weight for connected compact Lie groups *without* stepping too far into algebraic territory (as we did by first stating the classification for complex semisimple Lie algebras).

Example 3.53 [Unitary Group]. Let U_n be the connected compact Lie group of $n \times n$ unitary matrices. Then U_n has rank n, with a natural maximal torus given by the set of diagonal matrices $\mathrm{diag}(z_1, \ldots, z_n)$, where $z_i \in \mathbb{T}$ for all $1 \leq i \leq n$.

Theorem 3.54 [Torus Theorem]. Let G be a connected compact Lie group. Then every point of G lies in some maximal torus. Moreover, all maximal tori are conjugate and hence of the same dimension (which we call the *rank* of G).

Proof. See [Hal15, Theorem 11.9] and [Kna86, Corollary 4.35]. □

Corollary 3.55. Let G be a connected compact Lie group. Then the exponential map $\exp : \mathrm{Lie}(G) \to G$ is surjective.

Proof. We argued in the proof of Theorem 3.52 that the exponential map on every connected abelian group (hence every torus) is surjective. Thus, this follows immediately from Theorem 3.54. □

It remains to classify the irreps of an arbitrary torus group and define its associated character lattice. To that end, we fix a torus T of dimension r, so that there is an isomorphism $T \cong \mathbb{T}^r$. Since $\mathcal{I}_{\mathbb{C}}(T) \cong \hat{T} \cong (\hat{\mathbb{T}})^r$, we simply need to compute the Pontryagin dual of the circle. Now, for every integer n, the map $z \mapsto z^n$ from $\mathbb{C}^*$ to $\mathbb{C}^*$ restricts to a continuous homomorphism $\mathbb{T} \to \mathbb{T}$, yielding an injective homomorphism $\mathbb{Z} \hookrightarrow \hat{\mathbb{T}}$. If we can show that this homomorphism is actually surjective, and hence an isomorphism, we will obtain a parameterization $\mathbb{Z}^r \leftrightarrow \mathcal{I}_{\mathbb{C}}(T)$.

Theorem 3.56. Every $\varphi \in \hat{\mathbb{T}}$ is of the form $\varphi(z) = z^n$ for some $n \in \mathbb{Z}$.

Proof. Viewing $\mathbb{T}$ as a submanifold of $\mathbb{C}$ identifies $\mathrm{Lie}(\mathbb{T})$ with $i\mathbb{R}$ such that the exponential map $\exp : i\mathbb{R} \to \mathbb{T}$ is just $ix \mapsto e^{ix}$. Since $\dot{\varphi} : i\mathbb{R} \to i\mathbb{R}$ is linear, we can write $\dot{\varphi}(ix) = c \cdot ix$ for some $c \in \mathbb{R}$. But then we need $e^{cix} = \varphi(e^{ix})$ for all $x \in \mathbb{R}$, by (3.74). In particular,

$$e^{ci(x+2\pi)} = \varphi(e^{i(x+2\pi)}) = \varphi(e^{ix}) = e^{cix} \implies c \in \mathbb{Z}. \tag{3.83}$$

Hence, $\varphi(e^{ix}) = (e^{ix})^n$, where $n = c \in \mathbb{Z}$. Since exp is surjective onto $\mathbb{T}$, this means that $\varphi(z) = z^n$ for all $z \in \mathbb{T}$, as desired. $\square$

Using Theorem 3.56 to parameterize $\mathcal{I}_{\mathbb{C}}(T)$ by $\mathbb{Z}^r$ is not exactly natural, since an isomorphism $T \cong \mathbb{T}^r$ depends on a choice of basis for the *integral lattice* $\Pi := \ker(\exp) \subseteq \mathfrak{t}$. Observe, however, that Π induces a dual lattice $\Lambda \subseteq \mathfrak{t}^*$ with respect to the pairing $\mathfrak{t} \times \mathfrak{t}^* \to \mathbb{R}$, where Λ is the set of functionals which are integral on Π. The dual lattice Λ is called the *character lattice*. Identifying $\mathbb{R} \cong i\mathbb{R}$ via $x \mapsto 2\pi ix$ allows us to view the derivative $\dot{\varphi} : \mathfrak{t} \to i\mathbb{R}$ as a functional in the dual $\mathfrak{t}^*$, in which case (3.74) implies $e^{2\pi i\dot{\varphi}(\xi)} = \varphi(e^{\xi})$ for all $\xi \in \mathfrak{t}$. In particular, if $\xi \in \Pi$ then $\dot{\varphi}(\xi) \in \mathbb{Z}$, so that $\varphi \mapsto \dot{\varphi}$ maps $\hat{T} \to \Lambda$. By identifying this map with the isomorphism $\mathbb{Z}^r \to (\hat{\mathbb{T}})^r$, we see that $\hat{T} \to \Lambda$ is also an isomorphism. In other words, differentiation naturally identifies $\hat{T}$ with the character lattice Λ.

4 Superconcentration in the Orthogonally Invariant Case

In this final chapter, we outline progress towards proving our main Conjecture 1.14, using tools developed in the previous chapters. Ultimately, in Theorem 4.12 we reduce the conjecture to a much simpler, albeit computationally intensive, problem of bounding the expected smoothness of the free energy. To formulate this, we take a short expository digression in subchapter 4.2 to discuss the quadratic Casimir operator. Subchapters 4.1 and 4.3, on the other hand, reflect original research.

4.1 The Spectral Method for G-invariant SK Models

Inspired by Chatterjee's spectral approach to superconcentration, demonstrated for the Gaussian SK model in Chapter 2, we begin with a general framework for applying this method to G-invariant SK models. To that end, we fix a compact group G. Carrying our conventions over from Chapter 3, we let $\mathcal{I}_{\mathbb{C}}(G)$ denote the set of (isomorphism classes of) complex FD G-irreps, and we fix a representative W_π for each class $\pi \in \mathcal{I}_{\mathbb{C}}(G)$. We also let d_π be the dimension of W_π, equip W_π with a G-invariant inner product $\langle -, - \rangle_\pi$, fix an orthonormal basis $e_1^\pi, \ldots, e_{d_\pi}^\pi$ of W_π with respect to this inner product, and define the matrix coefficients π_{ij} as in equation (3.44). Then the Peter-Weyl theorem tells us that the set of all $\sqrt{d_\pi}\pi_{ij}$ forms an orthonormal Hilbert basis for $L^2(G)$.

Let N be a positive integer, $\Lambda \in \mathrm{Sym}_N(\mathbb{R})$, and $\beta > 0$. Let J be the disorder of the G-invariant SK model with structure Λ and inverse temperature β, and view the free energy $F_{N,\beta}$, abbreviated F, as a function $G \to \mathbb{R}$. In other words,

$$F(g) = \frac{1}{\beta} \log \sum_{\boldsymbol{\sigma} \in \{\pm 1\}^N} e^{\beta \mathcal{H}(\boldsymbol{\sigma})(g)}, \quad \text{where} \quad \mathcal{H}(\boldsymbol{\sigma})(g) = \boldsymbol{\sigma}^\intercal (g \cdot \Lambda) \boldsymbol{\sigma}. \tag{4.1}$$

Then $F \in C(G) \subseteq L^2(G)$, so by Parseval's identity we can write

$$\|F\|_2^2 = \sum_{\pi \in \mathcal{I}_{\mathbb{C}}(G)} d_\pi \sum_{i,j=1}^{d_\pi} \left| \int_G F(g) \overline{\pi_{ij}(g)} \, dg \right|^2. \tag{4.2}$$

Of course, we are interested in the variance of F with respect to the Haar probability measure on G, which is $\mathrm{Var}(F) = \|F\|_2^2 - (\int F(g)\, dg)^2$. Notice that the $(\int F(g)\, dg)^2$ term corresponds to the summand in (4.2) which arises from the trivial G-irrep, so that

$$\mathrm{Var}(F) = \sum_{\pi \in \mathcal{I}_{\mathbb{C}}^\times(G)} d_\pi \sum_{i,j=1}^{d_\pi} |a_{ij}^\pi|^2, \tag{4.3}$$

where $\mathcal{I}_{\mathbb{C}}^\times(G)$ is the set of *nontrivial* classes in $\mathcal{I}_{\mathbb{C}}(G)$ and $a_{ij}^\pi := \langle F, \pi_{ij} \rangle$.

Recalling the $(\mathbb{Z}/2)$-symmetry of the Gaussian SK model that allowed us to eliminate "odd" terms in the Hermite expansion of $\mathrm{Var}_{\gamma_N}(F_{N,\beta})$, we look for similar simplifications of (4.3) in cases of sufficient symmetry. More precisely, suppose that H is a discrete abelian subgroup of G for which F is left or right H-invariant, in the sense that $h \cdot F = F$ (resp. $F \cdot h = F$) for all $h \in H$. For instance, this is the case if each Hamiltonian $\mathcal{H}(\boldsymbol{\sigma})$ is left (resp. right) H-invariant. Because H is abelian, we can simultaneously diagonalize W_π with respect to the commuting unitary operators $\{\pi(h) \mid h \in H\}$. The set of simultaneous eigenspaces, which includes the subspace W_π^H of W_π fixed by H, is orthogonal, so we can assume each e_i^π is a simultaneous eigenvector, and more that $W_\pi^H = \operatorname{span}(e_1^\pi, \dots, e_{d_\pi^H}^\pi)$, where $d_\pi^H := \dim W_\pi^H$. We claim that $a_{ij}^\pi = 0$ if either $i > d_\pi^H$ or $j > d_\pi^H$.

Indeed, for such i, j, there must exist $h, h' \in H$ and $\lambda, \lambda' \neq 1$ for which $h \cdot e_i^\pi = \bar{\lambda} e_i^\pi$ and $(h')^{-1} \cdot e_j^\pi = \lambda' e_j^\pi$. Then for all $g \in G$,

$$(h \cdot \pi_{ij})(g) = \langle e_i^\pi, h^{-1} g \cdot e_j^\pi \rangle_\pi = \langle h \cdot e_i^\pi, g \cdot e_j^\pi \rangle_\pi = \bar{\lambda} \langle e_i^\pi, g \cdot e_j^\pi \rangle_\pi = \bar{\lambda} \cdot \pi_{ij}(g). \tag{4.4}$$

Likewise, because π is unitary, for all $g \in G$ we have

$$(\pi_{ij} \cdot h')(g) = \langle e_i^\pi, g(h')^{-1} \cdot e_j^\pi \rangle_\pi = \overline{\lambda'} \langle e_i^\pi, g \cdot e_j^\pi \rangle_\pi = \overline{\lambda'} \cdot \pi_{ij}(g). \tag{4.5}$$

Thus, if F is left H-invariant, then $a_{ij}^\pi = \langle F, \pi_{ij} \rangle = \langle h \cdot F, h \cdot \pi_{ij} \rangle = \lambda a_{ij}^\pi$ implies that $a_{ij}^\pi = 0$. Similarly, if F is right H-invariant, we have $a_{ij}^\pi = \lambda' a_{ij}^\pi$, so again $a_{ij}^\pi = 0$. Essentially, we can restrict our attention to W_π^H rather than all of W_π for each $\pi \in \mathcal{I}_\mathbb{C}(G)$, as far as (4.3) is concerned. This leads to the following analog of Lemma 2.33.

Lemma 4.1. Let H be a discrete abelian subgroup of G for which F is left or right H-invariant, and choose $e_1^\pi, \dots, e_{d_\pi}^\pi$ as in the previous discussion. Then

$$\operatorname{Var}(F) = \sum_{\pi \in \mathcal{I}_\mathbb{C}^\times(G)} d_\pi \sum_{i,j=1}^{d_\pi^H} |a_{ij}^\pi|^2. \tag{4.6}$$

In addition to the symmetry argument underpinning Lemma 2.33, there were two other tools that allowed us to control the variance of the Gaussian free energy in subchapter 2.4: the improved Gaussian Poincaré inequality (Theorem 2.30) and Gaussian integration by parts (Lemma 2.15). In order to avoid the formalization of Markov semigroups, we arrived at Theorem 2.30 by more direct means than Chatterjee, for whom Theorem 2.30 is a corollary of a general *improved Poincaré inequality* [Cha14, Theorem 6.2].

Broadly, given a probability measure of interest μ on a space $\mathcal{X}$, the key idea is that our "convenient" Hilbert basis for $L^2(\mu)$ is also an eigenbasis for a self-adjoint differential operator on $\mathcal{X}$, the *generator* of an associated Markov semigroup. For the N-dimensional standard Gaussian measure γ_N on $\mathbb{R}^N$, the generator is $Lf(x) = \Delta f(x) - x \cdot \nabla f(x)$, where Δ is the Laplacian. For the Haar measure on a connected compact Lie group G, the relevant generator is negative the *quadratic Casimir operator* $\Omega : C^\infty(G) \to C^\infty(G)$ with respect to some Ad-invariant inner product B on $\mathfrak{g} = \operatorname{Lie}(G)$. Algebraically, Ω is a central element of the *universal enveloping algebra* $U(\mathfrak{g})$. Geometrically, Ω is negative the Laplace-Beltrami operator with respect to the bi-invariant Riemannian metric on G obtained by translating B. We will formally define Ω in subchapter 4.2, but for now, it's enough to note its relevant properties: complex conjugation invariance ($\Omega\bar{f} = \overline{\Omega f}$), self-adjointness with respect to the $L^2(G)$ inner product, and diagonalizability over the matrix coefficients π_{ij}. The *Casimir eigenvalue* λ_π of each matrix coefficient π_{ij} is nonnegative, 0 only for the trivial G-irrep, and dependent solely on π (which justifies the notation λ_π).

The self-adjointness of the Laplace-Beltrami operator is an analog of integration by parts on a compact Riemannian manifold. Accordingly, we can attempt to bound the a^π_{ij} coefficients by iterating the self-adjointness of Ω several times and exploiting the smoothness of F, much like we did with Gaussian integration by parts in subchapter 2.4. Explicitly, for each $\pi \in \mathcal{I}^\times_\mathbb{C}(G)$ and $1 \le i, j \le d_\pi$, we note that

$$a^\pi_{ij} = \left\langle F, \frac{\Omega^r \pi_{ij}}{\lambda^r_\pi} \right\rangle = \lambda_\pi^{-r}\langle F, \Omega^r \pi_{ij}\rangle = \lambda_\pi^{-r}\langle \Omega^r F, \pi_{ij}\rangle. \tag{4.7}$$

In particular, by Jensen's inequality and the unitarity of π we can bound

$$\sum_{i=1}^{d_\pi}\sum_{j=1}^{d_\pi} |a^\pi_{ij}|^2 = \lambda_\pi^{-2r} \sum_{i=1}^{d_\pi}\sum_{j=1}^{d_\pi} \left| \int \Omega^r F(g) \overline{\pi_{ij}(g)}\, dg \right|^2 \tag{4.8}$$

$$\le \lambda_\pi^{-2r} \sum_{i=1}^{d_\pi}\sum_{j=1}^{d_\pi} \int |\Omega^r F(g)|^2 |\pi_{ij}(g)|^2\, dg \tag{4.9}$$

$$= d_\pi \lambda_\pi^{-2r} \int |\Omega^r F(g)|^2\, dg. \tag{4.10}$$

In the setting of Lemma 4.1, we can replace d_π with d^H_π in the first upper limit of each double sum, and d_π in (4.10) also becomes d^H_π. Because F is real-valued and Ω is complex conjugation invariant, $\Omega^r F$ is also real-valued, so $|\Omega^r F|^2 = (\Omega^r F)^2$.

Finally, we use Ω to formulate an improved Poincaré inequality on a connected compact Lie group G. For smooth functions $f_1, f_2 \in C^\infty(G)$, the *Dirichlet energy* $\mathcal{E}(f_1, f_2)$ of f_1 and f_2 is defined to be $\mathcal{E}(f_1, f_2) := \langle f_1, \Omega f_2\rangle$. The proof of the following theorem is adapted from Chatterjee's proof of [Cha14, Theorem 6.2].

Theorem 4.2 [Improved Poincaré on G]. Let G be a connected compact Lie group with quadratic Casimir operator Ω associated to an Ad-invariant inner product B on $\mathfrak{g} = \mathrm{Lie}(G)$. Fix $f \in C^\infty(G)$. Then for every $\ell > 0$,

$$\mathrm{Var}(f) \le \sum_{0<\lambda_\pi<\ell} d_\pi \sum_{i,j=1}^{d_\pi} |a^\pi_{ij}|^2 + \frac{\mathcal{E}(f,f)}{\ell}, \tag{4.11}$$

where $a^\pi_{ij} = \langle f, \pi_{ij}\rangle$ and the "$0 < \lambda_\pi < \ell$" subscript means that we sum over all $\pi \in \mathcal{I}_\mathbb{C}(G)$ with corresponding Casimir eigenvalue $0 < \lambda_\pi < \ell$. The upper limit of summation d_π can be replaced with d^H_π whenever H is a discrete abelian subgroup of G for which f is left or right H-invariant.

Proof. The Peter-Weyl expansion of f is $f = \sum d_\pi a_{ij}^\pi \pi_{ij}$, where the sum is over all $\pi \in \mathcal{I}_G(\mathbb{C})$ and $1 \le i, j, \le d_\pi$. Hence

$$\Omega f = \sum \lambda_\pi d_\pi a_{ij}^\pi \pi_{ij} \implies \mathcal{E}(f,f) = \sum \lambda_\pi d_\pi |a_{ij}^\pi|^2. \tag{4.12}$$

Denoting the trivial G-irrep by 1, by Parseval's identity we have

$$\mathrm{Var}(f) = \sum_{\pi \neq 1} d_\pi |a_{ij}^\pi|^2 = \sum_{0<\lambda_\pi<\ell} d_\pi |a_{ij}^\pi|^2 + \frac{1}{\ell} \sum_{\lambda_\pi \ge \ell} \ell d_\pi |a_{ij}^\pi|^2, \tag{4.13}$$

since $\lambda_\pi = 0$ if and only if $\pi = 1$. The desired inequality (4.11) follows from (4.13) and the simple bound

$$\sum_{\lambda_\pi \ge \ell} \ell d_\pi |a_{ij}^\pi|^2 \le \sum_{\lambda_\pi \ge \ell} \lambda_\pi d_\pi |a_{ij}^\pi|^2 \le \mathcal{E}(f,f). \tag{4.14}$$

This completes the proof. □

Remark 4.3. If ∇ is the gradient on G with respect to the Riemannian metric induced by B, then $\mathcal{E}(f,f) = \mathbb{E}[\|\nabla f\|^2]$ for all $f \in C^\infty(G)$, where for a vector field X we write $\|X\|^2 := B(X,X)$. We will see this concretely in subchapter 4.2.

Assume the setup of Lemma 4.1, where G is also a connected Lie group with quadratic Casimir operator Ω associated to some B. Then applying (4.10) and (4.11) to (4.6) yields

$$\mathrm{Var}(F) \le \mathbb{E}[(\Omega^r F)^2] \sum_{0<\lambda_\pi<\ell} \frac{d_\pi d_\pi^H}{\lambda_\pi^{2r}} + \frac{\mathbb{E}[\|\nabla F\|^2]}{\ell} \tag{4.15}$$

for all $r \ge 0$ and $\ell > 0$. Given $\ell > 0$, we let n_ℓ denote the number of $\pi \in \mathcal{I}_C(G)$ for which $0 < \lambda_\pi < \ell$. Also let λ_0 be the infimum of λ_π among $\pi \in \mathcal{I}_{\mathbb{C}}^\times(G)$ and D_ℓ (resp. D_ℓ^H) the supremum of d_π (resp. d_π^H) among $\pi \in \mathcal{I}_{\mathbb{C}}(G)$ with $\lambda_\pi < \ell$. Then by (4.15),

Lemma 4.4. For all $r \ge 0$ and $\ell > 0$, we have

$$\mathrm{Var}(F) \le n_\ell D_\ell D_\ell^H \lambda_0^{-2r} \mathbb{E}[(\Omega^r F)^2] + \frac{\mathbb{E}[\|\nabla F\|^2]}{\ell}. \tag{4.16}$$

We can always choose H to be the trivial subgroup of G, in which case $D_\ell^H = D_\ell$.

For fixed inverse temperature $\beta > 0$, suppose that in each dimension $N \ge 1$ we have a connected compact Lie group G_N acting smoothly on $\mathrm{Sym}_N\mathbb{R}$, a structure matrix $\Lambda_N \in \mathrm{Sym}_N(\mathbb{R})$, and associated free energy $F_N \in C^\infty(G)$. If we wish to control the asymptotic variance $\mathrm{Var}(F_N)$, by Lemma 4.4 we are reduced to considering the asymptotic behavior of the quantities n_ℓ, D_ℓ, λ_0, $\mathbb{E}[(\Omega^r F_N)^2]$, and $\mathbb{E}[\|\nabla F_N\|^2]$. When F_N has additional symmetry with respect to a discrete abelian subgroup H_N of G_N, in the sense of Lemma 4.1, we can further refine our variance bound in terms of $D_\ell^{H_N}$. While the presence of H_N-symmetry may not improve the asymptotics on $\mathrm{Var}(F_N)$, it does at least suggest that a superconcentration result for F_N may hold, by analogy with the Gaussian SK model.

4.2 The Quadratic Casimir Operator

This subchapter is dedicated to constructing the quadratic Casimir operator Ω and verifying the properties discussed in the previous subchapter. We begin with the *universal enveloping algebra* $U(\mathfrak{g})$ of a Lie algebra $\mathfrak{g}$ over k. Recall that every associative k-algebra A is a Lie algebra over k, with Lie bracket given by the commutator $[x, y] = xy - yx$, so it makes sense to discuss Lie algebra homomorphisms $\pi : \mathfrak{g} \to A$, i.e., k-linear maps for which $\pi([\xi, \eta]) = \pi(\xi)\pi(\eta) - \pi(\eta)\pi(\xi)$. The universal enveloping algebra $U(\mathfrak{g})$ is a unital associative k-algebra, equipped with a Lie algebra homomorphism $h : \mathfrak{g} \to U$, that satisfies a universal property: for every unital associative k-algebra A and Lie algebra homomorphism $\pi : \mathfrak{g} \to A$, there exists a unique unital k-linear ring homomorphism $\bar{\pi} : U(\mathfrak{g}) \to A$ such that

$$\begin{tikzcd} \mathfrak{g} \arrow[r, "\pi"] \arrow[d, "h"'] & A \\ U(\mathfrak{g}) \arrow[ur, "\exists! \bar{\pi}"'] & \end{tikzcd}$$

commutes. By the universal property, it's clear that $U(\mathfrak{g})$ is unique up to isomorphism. For existence, we can construct $U(\mathfrak{g})$ as a quotient of the tensor algebra $T(\mathfrak{g})$ by the two-sided ideal I generated by $\xi \otimes \eta - \eta \otimes \xi - [\xi, \eta]$ $(\xi, \eta \in \mathfrak{g})$, and define h to be the composition of the embedding $\mathfrak{g} \hookrightarrow T(\mathfrak{g})$ with the quotient map $T(\mathfrak{g}) \to T(\mathfrak{g})/I$. Importantly, h is injective, and hence an embedding, so that $\mathfrak{g}$ can be viewed as a subalgebra of $U(\mathfrak{g})$. In this sense, every Lie bracket is really a commutator in disguise. That h is injective follows from the powerful *Poincaré-Birkhoff-Witt* (PBW) theorem.

Theorem 4.5 [PBW]. Let $\mathfrak{g}$ be a Lie algebra with totally ordered basis $(\Xi, \leq)$. Let $h : \mathfrak{g} \to U(\mathfrak{g})$ be the canonical map into the universal enveloping algebra of $\mathfrak{g}$. Then the set of $h(\xi_1)h(\xi_2) \cdots h(\xi_n) \in U(\mathfrak{g})$, as $\xi_1 \leq \xi_2 \leq \cdots \leq \xi_n$ range over all (possibly empty) finite nondecreasing sequences in Ξ, is a basis for $U(\mathfrak{g})$. By convention, we associate 1 to the empty sequence.

Proof. See [Hal15, Section 9.4] for a proof in the FD case. The generalization to infinite dimensions goes through with essentially no modifications. □

Remark 4.6. By the well ordering theorem, every vector space has a basis, and this basis can be well ordered (hence totally ordered). Thus, even when $\mathfrak{g}$ is infinite dimensional, we can still conclude that $h : \mathfrak{g} \to U(\mathfrak{g})$ is injective. Of course, we don't need the axiom of choice for this when $\dim \mathfrak{g} < \infty$.

Suppose that V is a k-vector space and $\dot{\rho} : \mathfrak{g} \to \mathfrak{gl}(V)$ is a representation on V. Then by the universal property of $U(\mathfrak{g})$, the representation $\dot{\rho}$ extends to a unique unital k-linear ring map $U(\mathfrak{g}) \to \mathfrak{gl}(V)$, where $\mathfrak{gl}(V)$ is viewed as the associative algebra $\mathrm{End}(V)$.

The universal enveloping algebra has many applications, including the construction of Verma modules in the proof of Theorem 3.49, (3). In our case, we define Ω as an element of $U(\mathfrak{g})$. Suppose that $\mathfrak{g}$ is a real n-dimensional Lie algebra equipped with an ad-invariant inner product B. By ad-invariance, we mean that for all $\zeta, \xi, \eta \in \mathfrak{g}$, one has $B(\mathrm{ad}_\zeta \xi, \eta) + B(\xi, \mathrm{ad}_\zeta \eta) = 0$. For instance, if $\mathfrak{g}$ is semisimple then we can take $B = -\kappa$, where κ is the Killing form. We define the *quadratic Casimir operator* $\Omega \in U(\mathfrak{g})$ with respect to B as

$$\Omega := -\sum_{i=1}^{n} \xi_i^2, \tag{4.17}$$

where $\xi_1, \ldots, \xi_n$ is an orthonormal basis for $(\mathfrak{g}, B)$. By [Hal15, Proposition 10.5], Ω is well-defined (independent of the choice of orthonormal basis) and central in $U(\mathfrak{g})$.

If G is a Lie group with $\mathfrak{g} = \mathrm{Lie}(G)$ and B is an ad-invariant inner product on $\mathfrak{g}$ (which exists if, e.g., G is compact), then we still have to realize Ω as a differential operator on G. Since $\mathfrak{g}$ consists of vector fields on G, we can view each $\xi \in \mathfrak{g}$ as a first order differential operator $\xi : C^\infty(G) \to C^\infty(G)$, which induces a representation of $\mathfrak{g}$ on $C^\infty(G)$. This representation extends to a unital $\mathbb{R}$-linear ring map $U(\mathfrak{g}) \to \mathrm{End}_{\mathbb{R}}(C^\infty(G))$, which is how we can view elements of $U(\mathfrak{g})$ as operators $C^\infty(G) \to C^\infty(G)$. However, in the previous subchapter we viewed $C^\infty(G)$ as the space of complex-valued smooth functions on G (without comment, as it was demanded by the context), whereas in the present case $C^\infty(G)$ represents only real-valued functions. To avoid this ambiguity, we shall write $C^\infty(G, \mathbb{R})$ for the latter space. Then any $\mathbb{R}$-linear map $D : C^\infty(G, \mathbb{R}) \to C^\infty(G, \mathbb{R})$, such as the quadratic Casimir operator Ω, can be uniquely extended to a complex conjugation invariant $\mathbb{C}$-linear map $C^\infty(G) \to C^\infty(G)$ via $u + iv \mapsto Du + iDv$ for $u, v \in C^\infty(G, \mathbb{R})$. In this way, Ω becomes a complex conjugation invariant operator on $C^\infty(G)$.

Fix an n-dimensional compact Lie group G. Then by Theorem 3.21, we can chose an Ad-invariant inner product B on $\mathfrak{g} = \mathrm{Lie}(G)$, which is also ad-invariant by differentiation and the Leibniz rule. Let $\xi_1, \ldots, \xi_n$ be an orthonormal basis for $\mathfrak{g}$ with respect to B, and let Ω be the corresponding quadratic Casimir operator. Let $\rho : G \to V$ be an FD complex representation, and suppose that $\varphi \in C^\infty(G)$ is a matrix coefficient associated to ρ. In other words, for some $v \in V$, $\alpha \in V^*$ we can write $\varphi(g) = \langle \alpha, \rho(g)v \rangle$ for all $g \in G$ (where $\langle -, - \rangle$ is the duality pairing). Note that for all $\xi \in \mathfrak{g}$, we have

$$\xi\varphi(g) = \frac{d}{dt}\bigg|_{t=0} \langle \alpha, \rho(ge^{t\xi})v \rangle = \langle \alpha, \rho(g) \frac{d}{dt}\bigg|_{t=0} \rho(e^{t\xi})v \rangle = \langle \alpha, \rho(g)\dot{\rho}(\xi)v \rangle \tag{4.18}$$

by commuting the various linear operators with differentiation. In particular, $\xi\varphi$ is the matrix coefficient associated to ρ with $\dot{\rho}(\xi)v \in V$ and $\alpha \in V^*$. From (4.18), we deduce

$$\Omega\varphi(g) = \langle \alpha, \rho(g)\dot{\rho}(\Omega)v \rangle \tag{4.19}$$

under the extension of $\dot{\rho}$ to $U(\mathfrak{g})$. Thus, if $\dot{\rho}(\Omega) = \lambda_\rho \cdot \mathrm{id}_V$ for some constant λ_ρ, necessarily $\Omega\varphi = \lambda_\rho\varphi$. We claim that $\dot{\rho}(\Omega)$ is $\dot{\rho}$-linear, so when ρ is irreducible (and hence $\dot{\rho}$ is irreducible), such a λ_ρ exists by Schur's lemma—in particular, every matrix coefficient π_{ij} is an eigenvector for Ω, with corresponding Casimir eigenvalue λ_π dependent solely on π and not i, j. That $\dot{\rho}(\Omega)$ is $\dot{\rho}$-linear follows from the centrality of Ω.

Lemma 4.7. Let $\mathfrak{g}$ be a Lie algebra over an algebraically closed field k, and let $h : \mathfrak{g} \to U(\mathfrak{g})$ be the canonical embedding of $\mathfrak{g}$ into its universal enveloping algebra. Then if $\dot{\rho} : \mathfrak{g} \to \mathfrak{gl}(V)$ is a $\mathfrak{g}$-irrep and $\omega \in U(\mathfrak{g})$ is central, $\dot{\rho}(\omega) : V \to V$ is $\dot{\rho}$-linear.

Proof. Let $\xi \in \mathfrak{g}$. Because ω is central, we have $\xi\omega - \omega\xi = 0$ inside of $U(\mathfrak{g})$. Applying $\dot{\rho}$ yields $\dot{\rho}(\xi)\dot{\rho}(\omega) = \dot{\rho}(\omega)\dot{\rho}(\xi)$. Since ξ was arbitrary, $\dot{\rho}(\omega)$ is $\dot{\rho}$-linear. □

Remark 4.8. In the preceding discussion, we were ambiguous about when $\dot{\rho}$ should be viewed as a real $\mathfrak{g}$-rep or a complex $\mathfrak{g}^{\mathbb{C}}$-rep. Fortunately, the details work themselves out because $U(\mathfrak{g}^{\mathbb{C}}) \cong U(\mathfrak{g})^{\mathbb{C}}$. This isomorphism comes from complexifying the map $U(\mathfrak{g}) \to U(\mathfrak{g}^{\mathbb{C}})$ induced by $\mathfrak{g} \hookrightarrow \mathfrak{g}^{\mathbb{C}} \to U(\mathfrak{g}^{\mathbb{C}})$.

If $\pi \in \mathcal{I}_{\mathbb{C}}(G)$, we should be able to compute λ_π based on the associated highest weight and root system. Indeed, such a formula exists, assuming our choice of B is compatible with the inner product on the root system $(\Phi, E, (-,-))$.

Theorem 4.9. Let G be a compact connected semisimple Lie group, and choose Ω to be the quadratic Casimir operator with respect to the inner product $B = -\kappa$, where κ is the Killing form on $\mathfrak{g}$. With respect to the CSA $\mathfrak{t}^{\mathbb{C}}$ coming from a maximal torus, let $(\Phi, E, (-,-))$ be the associated root system, with Weyl vector δ. Then if π is a complex G-irrep with highest weight λ, we have

$$\lambda_\pi = (\lambda + \delta, \lambda + \delta) - (\delta, \delta) = (\lambda, \lambda + 2\delta) = (\lambda, \lambda) + \sum_{\alpha \in \Phi^+} (\lambda, \alpha). \tag{4.20}$$

Moreover, $\lambda_\pi \geq 0$, with equality only when $\lambda = 0$, i.e., when π is trivial.

Proof. See [Hal15, Proposition 10.6]. □

It remains to verify that Ω is self-adjoint and elaborate on Remark 4.3. If we extend B by translation to a bi-invariant Riemannian metric on G, then Ω corresponds to negative the Laplace-Beltrami operator Δ, and one has $\mathcal{E}(f, g) = \mathbb{E}[B(\nabla f, \nabla g)]$ by classical Riemannian geometry (since Δ is the divergence of ∇). In particular, the symmetry of B implies that $\mathcal{E}$ is symmetric, which in turn implies that Ω is self-adjoint. However, we can sidestep Riemannian geometry and deduce the self-adjointness of Ω from a more concrete statement of "integration by parts" on the Lie group G.

Lemma 4.10 [Integration by Parts on G]. Let G be a compact Lie group, $f_1, f_2 \in C^\infty(G)$, and $\xi \in \mathfrak{g}$. Then

$$\langle \xi f_1, f_2 \rangle = -\langle f_1, \xi f_2 \rangle. \tag{4.21}$$

Proof. By complex conjugation invariance, it's enough to assume that f_1, f_2 are real valued and prove $\int(\xi f_1)f_2 = -\int f_1(\xi f_2)$. The trick is to consider the two functions

$$F_1(t) = \int f_1(ge^{t\xi})f_2(g)\,dg \quad \text{and} \quad F_2(t) = \int f_1(g)f_2(ge^{t\xi})\,dg \tag{4.22}$$

on $\mathbb{R}$. By the invariance of the Haar measure, notice $F_1(t) = F_2(-t)$ for all $t \in \mathbb{R}$. Hence

$$\int(\xi f_1)f_2 = F_1'(0) = -F_2'(0) = \int f_1(\xi f_2), \tag{4.23}$$

where swapping a derivative and integral is justified by the mean value inequality and dominated convergence, since G is compact and f_1, f_2 are C^1. $\square$

Corollary 4.11. Ω is self-adjoint.

Proof. Clear because by two applications of Lemma 4.10, we have $\langle \xi^2 f_1, f_2\rangle = \langle f_1, \xi^2 f_2\rangle$ for all $f_1, f_2 \in C^\infty(G)$ and $\xi \in \mathfrak{g}$. $\square$

4.3 Reduction to Bounding Iterated Laplacians

We conclude by applying our work in subchapter 4.1 to an orthogonally invariant SK model. Throughout, we fix $\beta > 0$ and a sequence (Λ_N) of real $N \times N$ diagonal matrices with $\|\Lambda_N\| = O(1)$. Let $F_{N,\beta}$ be the free energy of the N-dimensional orthogonally invariant SK model with inverse temperature β and structure Λ_N, viewed as a Haar random variable on the connected compact Lie group SO_N. Observe that $F_{N,\beta}$ is both right and left invariant with respect to the discrete abelian subgroup $H_N \leq \mathrm{SO}_N$ of diagonal sign matrices. That is, H_N consists of diagonal $N \times N$ matrices with entries lying in $\{\pm 1\}$, an even number of which are -1. This parallels the "evenness" symmetry in the Gaussian case, which further suggests superconcentration of $F_{N,\beta}$, although it is not yet clear how to leverage this symmetry in resolving Conjecture 1.14 (we will ultimately bound away $D_\ell^{H_N} \leq D_\ell$).

Let $N \geq 1$, $M \geq 1$. Then applying Lemma 4.4 with $\ell = NM$ for each $r \geq 0$ yields

$$\mathrm{Var}(F_{N,\beta}) \leq P_{N,M} L_{N,\beta,r} + \frac{\mathbb{E}[\|\nabla F_{N,\beta}\|^2]}{NM}, \tag{4.24}$$

where $P_{N,M} := n_{NM} D_{NM} D_{NM}^{H_N}$ and $L_{N,\beta,r} := \lambda_0^{-2r}\mathbb{E}[(\Omega^r F_{N,\beta})^2]$. By analogy with our proof of Corollary 1.9, we conceive of a strategy to prove $\mathrm{Var}(F_{N,\beta}) = o_\beta(N)$. First, we show $\mathbb{E}[\|\nabla F_{N,\beta}\|^2] = O_\beta(N^2)$, so for every $\varepsilon > 0$ we may choose M large enough that

$$\frac{\mathrm{Var}(F_{N,\beta})}{N} \leq \frac{P_{N,M}}{N} L_{N,\beta,r} + \frac{\varepsilon}{2} \quad (N \geq 1, r \geq 0). \tag{4.25}$$

It then suffices to prove the existence of $r \geq 0$ and $N_0 \geq 1$ for which $N \geq N_0$ implies $P_{N,M} L_{N,\beta,r}/N < \varepsilon/2$. With this in mind, our next goal is to demonstrate that for fixed M, the quantity $P_{N,M}$ grows at worst polynomially in N, i.e., there exists a polynomial

$p_M \in \mathbb{R}[x]$ such that $P_{N,M} \leq p_M(N)$ for all N. Finally, superconcentration of $F_{N,\beta}$ is reduced to showing that $r \geq 0$ may always be chosen to ensure $L_{N,\beta,r}$ offsets polynomial growth of any order. In other words, for every $k \geq 1$, we want to find $r \geq 0$ such that $L_{N,\beta,r} = O_\beta(N^{-k})$. By demonstrating $\lambda_0 = N - 1$ for $N \geq 3$, we deduce:

Theorem 4.12. Suppose that for every positive integer k, there exists $r \geq 0$ such that $\mathbb{E}[(\Omega^r F_{N,\beta})^2] = O_\beta(N^{2r-k})$. Then $\mathrm{Var}(F_{N,\beta}) = o_\beta(N)$.

Our goal now is to justify the steps outlined in the previous paragraph, thereby proving Theorem 4.12. We will restrict our attention to $N \geq 3$ so that SO_N is semisimple. Recall that $\mathfrak{so}_N = \mathrm{Lie}(\mathrm{SO}_N)$ consists of $N \times N$ skew-symmetric matrices, where the Killing form is $\kappa(X, Y) = (N-2)\mathrm{tr}(XY)$. Rescale so that $B(X, Y) = -\frac{1}{2}\mathrm{tr}(XY) = \frac{1}{2}\mathrm{tr}(X^\intercal Y)$ is half the Frobenius metric on $M_N\mathbb{R}$, in which case Theorem 4.9 still holds if we appropriately rescale the root system inner product. Letting E_{ij} denote the $N \times N$ matrix with (i, j)th entry equal to 1 and all other entries equal to 0, the set $\{X_{ij} = E_{ij} - E_{ji} \mid 1 \leq i < j \leq N\}$ is an orthonormal basis for $(\mathfrak{so}_N, B)$. Letting $n = \lfloor N/2 \rfloor$, we take the standard maximal torus $T \subseteq \mathrm{SO}_N$ consisting of block-diagonal matrices $\mathrm{diag}(A_1, \ldots, A_n, A')$, where $A_i \in \mathrm{SO}_2$ for $1 \leq i \leq n$ and A' is either empty (the unique 0×0 matrix) or the 1×1 matrix $A' = (1)$, depending on the parity of N. In particular, SO_N has rank n and the subalgebra $\mathfrak{t} = \mathrm{Lie}(T)$ has basis $X_{12}, X_{34}, X_{56}, \ldots, X_{2n-1,2n}$. If N is even, then the corresponding root system is D_n, while if N is odd, the root system is B_n. With respect to our rescaling of the inner product, the longest roots have length $\sqrt{2}$ (see [Kna86, Examples 7.7.2, 7.7.3]), so we can identify $\mathfrak{t}^* \cong \mathbb{R}^n$ and consult [Kna86, Appendix C.1] to label the SO_N-irreps.

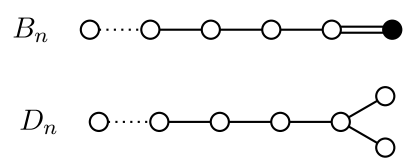

Figure 12: Dynkin diagrams for B_n and D_n

By studying the root system associated to SO_N, we can use Theorem 4.9 and the Weyl dimension formula to upper bound $P_{N,M}$ polynomially in N and show $\lambda_0 = N - 1$. Then, we compute derivatives of $F_{N,\beta}$ in exponential coordinates, which we can bound in Haar expectation using the tools of *Weingarten calculus* [CŚ06]. The latter calculations are quite messy, but since we hope to present the details in a future paper, we omit them in this thesis for brevity. The idea, however, is simple: although there is no $1/N$ factor in the orthogonally invariant Hamiltonian, which prevents us from directly proving analogs to Lemma 2.35, we *can* formulate these analogs in expectation. At present, this appears to be a promising approach to dealing with $\mathbb{E}[(\Omega^r F_{N,\beta})^2]$ and resolving Conjecture 1.14. Fortunately, in the case of the Dirichlet energy, we can deduce the desired bound $O_\beta(N^2)$ without invoking these more complicated tools. In fact, Tobias and I have independently proven that we can uniformly bound $\|\nabla F_{N,\beta}\|^2 \leq 4N^2 \|\Lambda_N\|^2$. I present my proof of this fact.

Lemma 4.13. For any $\beta > 0$, we have $\|\nabla F_{N,\beta}\|^2 \leq 4N^2\|\Lambda_N\|^2$.

Proof. Write $F = F_{N,\beta}$. Then with respect to B we have $\Omega = -\sum_{i<j} X_{ij}^2$ and

$$\|\nabla F\|^2 = \sum_{1\leq i<j\leq N} (X_{ij}F)^2. \tag{4.26}$$

Now a straightforward calculation shows that, for each $1 \leq i < j \leq N$,

$$X_{ij}F = 2\langle \boldsymbol{\sigma}^\intercal J X_{ij}\boldsymbol{\sigma}\rangle, \tag{4.27}$$

where $\langle - \rangle$ denotes the Gibbs average. It follows then from Jensen's inequality that

$$\|\nabla F\|^2 \leq 4\sum_{i<j}\langle(\boldsymbol{\sigma}^\intercal J X_{ij}\boldsymbol{\sigma})^2\rangle = 4\Big\langle\sum_{i<j}(\boldsymbol{\sigma}^\intercal J X_{ij}\boldsymbol{\sigma})^2\Big\rangle. \tag{4.28}$$

Fix $\boldsymbol{\sigma} \in \{\pm 1\}^N$ and let $\boldsymbol{\tau}_{ij} = X_{ij}\boldsymbol{\sigma} = \sigma_j e_i - \sigma_i e_j$ for each $i < j$. Then because $J = J^\intercal$,

$$\sum_{i<j}(\boldsymbol{\sigma}^\intercal J X_{ij}\boldsymbol{\sigma})^2 = \sum_{i<j}(J\boldsymbol{\sigma})^\intercal \boldsymbol{\tau}_{ij}\boldsymbol{\tau}_{ij}^\intercal J\boldsymbol{\sigma} = (J\boldsymbol{\sigma})^\intercal\Big[\sum_{i<j}\boldsymbol{\tau}_{ij}\boldsymbol{\tau}_{ij}^\intercal\Big]J\boldsymbol{\sigma}. \tag{4.29}$$

By direct computation, $\boldsymbol{\tau}_{ij}\boldsymbol{\tau}_{ij}^\intercal = E_{ii} + E_{jj} - \sigma_i\sigma_j(E_{ij} + E_{ji})$. Summing over all $i < j$ leads to two terms, namely $\sum_{i<j}(E_{ii} + E_{jj}) = (N-1)I_N$ and $\sum_{i<j}\sigma_i\sigma_j(E_{ij} + E_{ji}) = \boldsymbol{\sigma}\boldsymbol{\sigma}^\intercal - I_N$, whose difference is $\sum_{i<j}\boldsymbol{\tau}_{ij}\boldsymbol{\tau}_{ij}^\intercal = NI_N - \boldsymbol{\sigma}\boldsymbol{\sigma}^\intercal$. Therefore

$$(J\boldsymbol{\sigma})^\intercal\Big[\sum_{i<j}\boldsymbol{\tau}_{ij}\boldsymbol{\tau}_{ij}^\intercal\Big]J\boldsymbol{\sigma} = N\|J\boldsymbol{\sigma}\|^2 - (\boldsymbol{\sigma}^\intercal J\boldsymbol{\sigma})^2 \leq N\|J\boldsymbol{\sigma}\|^2. \tag{4.30}$$

But as $\|J\| = \|\Lambda_N\|$ and $\|\boldsymbol{\sigma}\| = \sqrt{N}$, we have $\|J\boldsymbol{\sigma}\|^2 \leq N\|\Lambda_N\|^2$. Putting everything together gives the desired inequality $\|\nabla F_{N,\beta}\|^2 \leq 4N^2\|\Lambda_N\|^2$. $\square$

We return to the problem of computing λ_0 and bounding $P_{N,M}$. Suppose that an irrep $\pi \in \mathcal{I}_{\mathbb{C}}(\mathrm{SO}_N)$ corresponds to a dominant analytically integral element $\lambda \in \mathbb{R}^n$ with coordinates $\lambda = \lambda_1 e_1 + \cdots + \lambda_n e_n$. Then each $\lambda_i \in \mathbb{Z}$ and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. The λ_i are also nonnegative, except possibly λ_n when N is even. However, by H_N-invariance, a_{ij}^π vanish when $\lambda_n < 0$. This detail is irrelevant asymptotically, but for simplicity it allows us to view λ as a Young diagram of length at most n. By [Kna86, Appendix C.1], we compute λ_π and d_π using Theorem 4.9 and the Weyl dimension formula, respectively. For one,

$$\lambda_\pi = \sum_{i=1}^n \lambda_i(\lambda_i + N - 2i). \tag{4.31}$$

From this formula we see immediately that $\lambda_\pi \geq N - 1$ if $\pi \neq 1$, which is achieved when $\lambda = e_1$. This weight is in fact analytically integral, corresponding to the complexification of the standard representation $\mathrm{SO}_N \circlearrowright \mathbb{R}^N$, so $\lambda_0 = N-1$. Combined with Lemma 4.13, this is already sufficient for deducing the temperature independent bound

$$\operatorname{Var}(F_{N,\beta}) \leq \frac{4N^2\|\Lambda_N\|^2}{N-1}, \tag{4.32}$$

which is $O(N)$ under our assumption $\|\Lambda_N\| = O(1)$. This is at least a reassuring sanity check—certainly a necessary condition for *super*concentration is concentration. Next we claim that, for fixed $M \geq 1$, the maximum number of rows and columns among Young diagrams representing π with $\lambda_\pi < NM$ is $O_M(1)$. In particular, $n_{NM} = O_M(1)$.

Lemma 4.14. Given $N \geq 3$, $M \geq 1$, suppose that $\lambda_\pi < NM$ for some irrep π with corresponding $\lambda \in \mathbb{R}^n$. Then $\lambda_1 < \max(2, M)$, and the length ℓ of λ (the number of nonzero λ_i) is at most $(N - \sqrt{N^2 - 4MN})/2 = O_M(1)$ when $N \geq 4M$.

Proof. Certainly $\lambda_1(\lambda_1 + N - 2) < NM$. If $\lambda_1 \geq \max(2, M)$, then $\lambda_1(\lambda_1 + N - 2) \geq NM$, so we must have $\lambda_1 < \max(2, M)$. As for ℓ, notice that $\ell(N-\ell) = \sum_{i=1}^{\ell}(1 + N - 2i) < NM$, so by the quadratic formula, either

$$\ell > \frac{N + \sqrt{N^2 - 4MN}}{2} \quad \text{or} \quad \ell < \frac{N - \sqrt{N^2 - 4MN}}{2}. \tag{4.33}$$

Necessarily $\ell \leq n \leq N/2$, removing the first option if N is sufficiently large ($N \geq 4M$). As $N \to \infty$, note that $(N - \sqrt{N^2 - 4MN})/2 \to M$. $\square$

Finally, we bound the dimension d_π. If N is odd, then $d_\pi = \mu_\pi \nu_\pi$, where

$$\mu_\pi = \prod_{i=1}^{n} \frac{\lambda_i + \frac{N}{2} - i}{\frac{N}{2} - i} \quad \text{and} \quad \nu_\pi = \prod_{1 \leq i < j \leq n} \frac{(\lambda_i + \lambda_j + N - i - j)(\lambda_i - \lambda_j + j - i)}{(N - i - j)(j - i)}. \tag{4.34}$$

If N is even, the formula is the same except with the μ_π term absent, i.e., $d_\pi = \nu_\pi$.

Lemma 4.15. Let M be a positive integer. Then there exists a polynomial $p \in \mathbb{R}[x]$ such that $d_\pi \leq p(N)$ for all $N \geq 3$ and π with $\lambda_\pi < NM$.

Proof. Restricting our attention to λ representing π with $\lambda_\pi < NM$, it suffices to prove that both μ_π and ν_π grow at worst polynomially in N. We begin with μ_π (and N odd). Since $\lambda_i \leq \lambda_1$ for all i and all multiplicands are 1 when $i > \ell$,

$$\mu_\pi = \prod_{i=1}^{\ell} \left(\frac{\lambda_i}{\frac{N}{2} - i} + 1 \right) \leq \left(\frac{\lambda_1}{\frac{N}{2} - \ell} + 1 \right)^{\ell} \leq (2\lambda_1 + 1)^{\ell}. \tag{4.35}$$

But λ_1 and ℓ are both $O_M(1)$, so we conclude that μ_π is $O_M(1)$.

It remains to bound ν_π. We can groups the factors in this product based on whether $j \le \ell$ or $j > \ell$. In the first case, we get the term

$$\prod_{1 \le i < j \le \ell} \frac{(\lambda_i + \lambda_j + N - i - j)(\lambda_i - \lambda_j + j - i)}{(N - i - j)(j - i)}, \tag{4.36}$$

where each multiplicand may be bounded by $(2\lambda_1 + 1)(\lambda_1 + 1) = O_M(1)$. Since there are $\ell(\ell - 1)/2 = O_M(1)$ multiplicands in total, this term is $O_M(1)$. In the second case, we note that $\lambda_k = 0$ if $k > \ell$, so the term of interest is

$$\prod_{1 \le i \le \ell < j \le n} \frac{(\lambda_i + N - i - j)(\lambda_i + j - i)}{(N - i - j)(j - i)} \le \left(\prod_{j=\ell+1}^{n} \frac{(\lambda_1 + N - 1 - j)(\lambda_1 + j - 1)}{(N - \ell - j)(j - \ell)} \right)^{\ell}. \tag{4.37}$$

Since $\ell = O_M(1)$, it's enough to polynomially bound the parenthesized product in equation (4.37). But this product is equal to

$$\frac{(N - n - \ell - 1)!(N + \lambda_1 - \ell)!(n + \lambda_1 - 1)!}{(N - 1)!(n - \ell)!(N - n + \lambda_1 - 2)!(\lambda_1 + \ell - 1)!}. \tag{4.38}$$

If N is even, then $N = 2n$ and the above quantity simplifies to

$$\frac{(n - \ell - 1)!(2n + \lambda_1 - \ell)!(n + \lambda_1 - 1)!}{(2n - 1)!(n - \ell)!(n + \lambda_1 - 2)!(\lambda_1 + \ell - 1)!} = \frac{(n + \lambda_1 - 1)(2n + \lambda_1 - \ell)!}{(n - \ell)(2n - 1)!(\lambda_1 + \ell - 1)!}. \tag{4.39}$$

If N is odd, then $N = 2n + 1$ and we obtain instead

$$\frac{(n - \ell)!(2n + 1 + \lambda_1 - \ell)!(n + \lambda_1 - 1)!}{(2n)!(n - \ell)!(n + \lambda_1 - 1)!(\lambda_1 + \ell - 1)!} = \frac{(2n + 1 + \lambda_1 - \ell)!}{(2n)!(\lambda_1 + \ell - 1)!}. \tag{4.40}$$

Asymptotically we can ignore $(n + \lambda_1 - 1)/(n - \ell) \to 1$ in (4.39), since λ_1 and ℓ are $O_M(1)$, and for both (4.39) and (4.40) we can bound away $(\lambda_1 + \ell - 1)! \ge 1$. Thus, we just need to examine the behavior of

$$\frac{(2n + \lambda_1 - \ell)!}{(2n - 1)!} \quad \text{and} \quad \frac{(2n + 1 + \lambda_1 - \ell)!}{(2n)!}. \tag{4.41}$$

Since $\lambda_1 - \ell = O_M(1)$, both of these quotients behave like $N^{O_M(1)}$. $\square$

By Lemma 4.15, D_{NM} grows at worst polynomially in N. As $D_{NM}^{H_N} \le D_{NM}$ and $n_{NM} = O_M(1)$, we conclude that $P_{N,M}$ also grows at worst polynomially in N. Combined with Lemma 4.13 and the fact that $\lambda_0 = N - 1$, this completes the proof of Theorem 4.12. Of course, bounding away $D_{NM}^{H_N} \le D_{NM}$ seems potentially nonoptimal because we do not take advantage of the H_N-symmetry, as remarked at the beginning of this subchapter. Even bounding away the d_π and $d_\pi^{H_N}$ terms by D_{NM} and $D_{NM}^{H_N}$ may be nonoptimal. Regardless, without incorporating these tighter bounds, we have reduced Conjecture 1.14 to average smoothness bounds on the free energy via Theorem 4.12.

5 Appendix

Here we include several auxiliary proofs that were felt relatively disconnected from the main ideas of the thesis, but which are nonetheless important. There were more results that I had originally wished to include in section A but unfortunately did not have time to detail, so I recommend [Kal21] as a general reference for further reading.

A Measure and Probability Theory

Theorem A.1. Let $(\mathcal{X}, \Sigma, \mu)$ be a countably generated σ-finite measure space. Then the normed space $L^p(\mu)$ is separable for all $1 \leq p < \infty$.

Proof. Assume that μ is finite, for the σ-finite case can be deduced easily from the finite case. Recall that a countably generated measurable space is one such that $\Sigma = \sigma(\mathcal{A})$, where $\mathcal{A} \subseteq \mathcal{X}$ is countable. Since the algebra generated by a countable family of sets is countable (just take all finite unions of finite intersections), we may as well assume that $\mathcal{A}$ is an algebra. Working over $\mathbb{R}$, we consider the set

$$S = \operatorname{span}_{\mathbb{Q}}(\{\mathbb{I}_A \mid A \in \mathcal{A}\}),$$

where $\mathbb{I}_A$ denotes the indicator function of A. In other words, S is the set of all finite rational-coefficient linear combinations of indicators on sets of $\mathcal{A}$ (if we were working over $\mathbb{C}$, we could replace $\mathbb{Q}$ with the number field $\mathbb{Q}(i)$ of Gaussian rationals). Let S_μ denote those functions in S which have μ-finite support. Then S_μ is countable, and we claim that S_μ is dense in $L^p(\mu)$ for all $1 \leq p < \infty$. Since $\mathbb{Q}$ is dense in $\mathbb{R}$ and simple functions with μ-finite support are dense in $L^p(\mu)$ [Rud87, Theorem 3.13], it suffices to prove that

$$\mathcal{B} := \{B \in \Sigma \mid \forall \varepsilon > 0, \text{ there exists } A \in \mathcal{A} \text{ with } \mu(A \triangle B) < \varepsilon\} \tag{A.1}$$

is the entirety of Σ. Here $\triangle$ denotes the symmetric difference, so that $|\mathbb{I}_A - \mathbb{I}_B| = \mathbb{I}_{A\triangle B}$.

Now, certainly $\mathcal{A} \subseteq \mathcal{B} \subseteq \Sigma$, so because $\mathcal{A}$ is a π-system such that $\sigma(\mathcal{A}) = \Sigma$, it suffices by Dynkin's π-λ theorem to prove that $\mathcal{B}$ is a λ-system. Notice $\varnothing, \mathcal{X} \in \mathcal{A} \subseteq \mathcal{B}$. The identity $A \triangle B = A^c \triangle B^c$, along with the fact that $\mathcal{A}$ is closed under complements, proves that $\mathcal{B}$ is closed under complements. Finally, let $B_1, B_2, \ldots$ be a countable family of disjoint sets in $\mathcal{B}$, and set $B = \cup B_n$. If $\varepsilon > 0$, for each $n \geq 1$ we can choose $A_n \in \mathcal{A}$ for which

$$\mu(A_n \triangle B_n) \leq \frac{\varepsilon}{2^{n+1}}. \tag{A.2}$$

Since μ is finite, we can choose $N \geq 1$ large enough so that $\mu(\cup_{n>N} B_n) < \varepsilon/2$. Then if we let $A = \cup_{n \leq N} A_n \in \mathcal{A}$, $B_{\leq N} = \cup_{n\leq N} B_n$, and $B_{>N} = \cup_{n>N} B_n$, we have

$$A \triangle B \subseteq (A \triangle B_{\leq N}) \cup B_{>N} \subseteq B_{>N} \cup \Big(\bigcup_{n=1}^{N} A_n \triangle B_n\Big). \tag{A.3}$$

It follows readily that $\mu(A \triangle B) \leq \mu(\cup_{n>N} B_n) + \sum_{n=1}^{N} \mu(A_n \triangle B_n) < \varepsilon$. $\square$

Theorem A.2. Let (X, τ) be a second-countable topological space. Then the Borel σ-algebra $\mathcal{B}(X)$ is countably generated.

Proof. Let $\mathcal{A}$ be a countable basis for τ. Let $\mathcal{A} \subseteq \Sigma \subseteq \mathcal{B}(X)$, where Σ is a σ-algebra. Then Σ contains all countable unions of sets in $\mathcal{A}$. But since $\mathcal{A}$ is a basis for τ, every open set is a union of sets in $\mathcal{A}$, and as $\mathcal{A}$ is countable, any such union may be considered countable. It follows that $\tau \subseteq \Sigma \subseteq \mathcal{B}(X)$, so by definition, $\Sigma = \mathcal{B}(X)$. Hence $\mathcal{B}(X) = \sigma(\mathcal{A})$, which proves that $\mathcal{B}(X)$ is countably generated. $\square$

B Smooth Approximation by Convolution

Our goal is to make precise and justify the approximation arguments in the proofs of Theorem 2.10 and Theorem 2.30. In both cases, we have already proven the theorem if $f \in C_c^\infty(\mathbb{R}^N)$ is a compactly supported smooth function on $\mathbb{R}^N$. For general f, we desire to find a sequence of $f_n \in C_c^\infty(\mathbb{R}^N)$ which converges to f in a suitable manner.

First we consider Theorem 2.10. By assumption, $f \in C^1(\mathbb{R}^N) \cap L^2(\gamma_N)$, and thus it suffices to prove

$$\|f\|_{\gamma_N}^2 \leq \left(\int f \, d\gamma_N\right)^2 + \sum_{i=1}^N \|\partial^i f\|_{\gamma_N}^2, \tag{B.1}$$

where $\|\cdot\|_{\gamma_N}$ is the norm on $L^2(\gamma_N)$. We can assume the first order partials of f lie in $L^2(\gamma_N)$, for otherwise the inequality is trivial. Since (B.1) is known to hold with f compactly supported and smooth, we just need to construct $f_n \in C_c^\infty(\mathbb{R}^N)$ so that f_n and $\partial^i f_n$ converge in $L^2(\gamma_N)$ to f and $\partial^i f$, respectively ($1 \leq i \leq N$).

Similarly, we can reduce Theorem 2.30 to the claim that, for $f \in C^M(\mathbb{R}^N)$ with $\partial^{\boldsymbol{k}} f \in L^2(\gamma_N)$ for all $|\boldsymbol{k}| \leq M$, one can find $f_n \in C_c^\infty(\mathbb{R}^N)$ such that $\partial^{\boldsymbol{k}} f_n \to \partial^{\boldsymbol{k}} f$ in $L^2(\gamma_N)$ for all $|\boldsymbol{k}| \leq M$. The desired claim for Theorem 2.10 is then the special case where $M = 1$. Therefore, Theorem 2.10 and Theorem 2.30 reduce to the following.

Theorem B.1. Let $f \in C^M(\mathbb{R}^N)$ such that $\partial^{\boldsymbol{k}} f \in L^2(\gamma_N)$ for all $|\boldsymbol{k}| \leq M$. Then there exist $f_n \in C_c^\infty(\mathbb{R}^N)$ such that for all $|\boldsymbol{k}| \leq M$, $\partial^{\boldsymbol{k}} f_n \to \partial^{\boldsymbol{k}} f$ in $L^2(\gamma_N)$.

Slightly more generally, we can consider the Gaussian *Sobolev space* $W^{M,p}(\gamma_N)$ for each $p \geq 1$, which is the Banach space of measurable $f : \mathbb{R}^N \to \mathbb{C}$ such that $\partial^{\boldsymbol{k}} f$ exists weakly in $L^p(\gamma_N)$ for all $|\boldsymbol{k}| \leq M$, equipped with the norm

$$\|f\|_{M,p} := \sum_{|\boldsymbol{k}| \leq M} \|\partial^{\boldsymbol{k}} f\|_p. \tag{B.2}$$

Then the space $C_c^\infty(\mathbb{R}^N)$ is dense in $W^{M,p}(\mathbb{R}^N)$. We will not concern ourselves with the abstraction of weak derivatives, but the argument we provide to prove Theorem B.1 is readily adapted to prove the more general claim that $C_c^\infty(\mathbb{R}^N)$ is dense in $W^{M,p}(\gamma_N)$.

Let $L^1_{\mathrm{loc}}(\mathbb{R}^N)$ denote the space of functions $f : \mathbb{R}^N \to \mathbb{C}$ such that $f|_K \in L^1(K)$ for all compact subsets $K \subseteq \mathbb{R}^N$. We recall the following definition:

Definition B.2 [Convolution]. Let $f \in C_c(\mathbb{R}^N)$ and $g \in L^1_{\mathrm{loc}}(\mathbb{R}^N)$. The *convolution* $f * g$ of f and g is

$$(f * g)(x) := \int_{\mathbb{R}^N} f(x-y)g(y)\,dy = \int f(y)g(x-y)\,dy. \tag{B.3}$$

Note that $f \in C_c(\mathbb{R}^N)$ and $g \in L^1_{\mathrm{loc}}(\mathbb{R}^N)$ ensure that $f * g$ is a well-defined, since

$$\int_{\mathbb{R}^N} |f(x-y)g(y)|\,dy \le \|f\|_\infty \int_{x-\mathrm{supp}(f)} |g(y)|\,dy < \infty. \tag{B.4}$$

In fact, $f * g$ is continuous. To see this, let $\varepsilon > 0$ and $x \in \mathbb{R}^N$. Since $f \in C_c(\mathbb{R}^N)$, in particular f is uniformly continuous, so we can find $\delta > 0$ such that $|h| < \delta$ implies $|f(z+h) - f(z)| < \varepsilon/(\|g\|_1 + 1)$ for all $z \in \mathbb{R}^N$. Thus, if $|h| < \delta$, then

$$|(f*g)(x+h) - (f*g)(x)| \le \int_{\mathbb{R}^N} |f(x-y+h) - f(x-y)||g(y)|\,dy \tag{B.5}$$

$$\le \delta\|g\|_1 \le \varepsilon \frac{\|g\|_1}{\|g\|_1 + 1} < \varepsilon. \tag{B.6}$$

The heuristic is that $f * g$ inherits the "best" properties from f. We have just seen that continuity is one such property, but it turns out that smoothness is inherited as well.

Theorem B.3. Let $f \in C^M_c(\mathbb{R}^N)$ and $g \in L^1_{\mathrm{loc}}(\mathbb{R}^N)$. Then $f * g \in C^M(\mathbb{R}^N)$, and for each $|\boldsymbol{k}| \le M$ we have

$$\partial^{\boldsymbol{k}}(f * g) = \partial^{\boldsymbol{k}} f * g. \tag{B.7}$$

Proof. It suffices to do this when $M = 1$. The general claim follows by induction on M. Thus, suppose that $f \in C^1_c(\mathbb{R}^N)$, $g \in L^1_{\mathrm{loc}}(\mathbb{R}^N)$, and $1 \le i \le N$. Let e_i be the ith standard basis vector of $\mathbb{R}^N$. Let t_n be a sequence of nonzero real numbers with $t_n \to 0$. Then

$$\frac{(f*g)(x+t_ne_i) - (f*g)(x)}{t_n} = \int \frac{f(x-y+t_ne_i) - f(x-y)}{t_n} g(y)\,dy. \tag{B.8}$$

Assuming that we can justify the application of dominated convergence to the integral on the right-hand side of (B.8), we deduce that $\partial^i(f * g) = \partial^i f * g$.

Note that by the mean value inequality applied to f, we have

$$\left|\frac{f(x-y+t_ne_i) - f(x-y)}{t_n} g(y)\right| \le \|\partial^i f\|_\infty |g(y)| \mathbb{I}(y \in A_{x,n}), \tag{B.9}$$

where $A_{x,n} = \{y \in \mathbb{R}^N \mid f(x-y+t_ne_i) \neq 0,\ f(x-y) \neq 0\}$. If $T > 0$ is such that $|t_n| \le T$ for all n, then $A_{x,n} \subseteq (x - \mathrm{supp} f) + \bar{B}_T(0)$, which is compact, so that the right-hand side of (B.9) is integrable. Hence, dominated convergence is justified. $\square$

Given an arbitrary function $g \in L^1_{\mathrm{loc}}(\mathbb{R}^N)$, we can construct a family of smooth functions $(g_\varepsilon)_{\varepsilon>0}$ that approximate g by defining $g_\varepsilon = \varphi_\varepsilon * g$, where $(\varphi_\varepsilon)_{\varepsilon>0}$ is a family of smooth density functions that "approximate the identity," i.e., the δ distribution at 0, as $\varepsilon \to 0$. Namely, we shall want each φ_ε to be smooth, supported on the closed ball $\bar{B}_\varepsilon(0)$ of radius ε centered at 0, bounded between 0 and 1, and integrating to 1. Such functions are often called *mollifiers*. A function φ which satisfies the previous definition in the case $\varepsilon = 1$ is a *standard mollifier*, and from it we can construct a sequence of mollifiers (φ_ε), called the standard mollifiers *associated* to φ, by defining

$$\varphi_\varepsilon(x) := \varepsilon^{-N}\varphi(x/\varepsilon). \tag{B.10}$$

As an example, from smooth bump functions one can construct radially symmetric standard mollifiers, such as

$$\varphi(x) \propto \mathbb{I}(|x| < 1)e^{-1/(1-|x|^2)}. \tag{B.11}$$

The proportionality constant is chosen to ensure $\int \varphi(x)\,dx = 1$. We now make precise the claim that g_ε is a very good approximation to g when (φ_ε) is a sequence of mollifiers associated to a standard mollifier φ.

Theorem B.4. Let (φ_ε) be a sequence of mollifiers in N dimensions associated to a standard mollifier φ. Then if $1 \le p < \infty$ and $g \in L^p(\mathbb{R}^N)$, the sequence of smooth functions $g_\varepsilon := \varphi_\varepsilon * g$ converges to g in $L^p(\mathbb{R}^N)$ as $\varepsilon \to 0$.

Proof. For each $\varepsilon > 0$, let μ_ε denote the probability measure on $\mathbb{R}^N$ defined such that $d\mu_\varepsilon(x) = \varphi_\varepsilon(x)\,dx$. Note that

$$g_\varepsilon(x) - g(x) = \int g(x-y) - g(x)\,d\mu_\varepsilon(y). \tag{B.12}$$

Thus, by Jensen's inequality and Tonelli's theorem we can write

$$\|g_\varepsilon - g\|_p^p = \int |g_\varepsilon(x) - g(x)|^p\,dx \le \int h(y)\,d\mu_\varepsilon(y), \tag{B.13}$$

where $h(y) := \int |g(x-y) - g(x)|^p\,dx$. We claim that $h(y) \to 0$ as $y \to 0$. Because μ_ε is supported on $[-\varepsilon, \varepsilon]$, this implies $\int h(y)\,d\mu_\varepsilon(y) \to 0$ as $\varepsilon \to 0$.

It remains to justify our claim that $\lim_{y\to 0} h(y) = 0$. If $g \in C_c(\mathbb{R}^N)$, this follows readily from dominated convergence. More generally, one uses the density of $C_c(\mathbb{R}^N)$ in $L^p(\mathbb{R}^N)$. Since we will only apply Theorem B.4 in the case that g is compactly supported and continuous, the details are left to the reader. $\square$

We are finally in a position to prove Theorem B.1. Let $f \in C^M(\mathbb{R}^N)$ be as in the theorem statement. While we do not need to work with the entire Sobolev space $W^{M,2}(\gamma_N)$, we can still define the (possibly unbounded) norm $\|\cdot\|_{M,2}$ as in (B.2) on $C^M(\mathbb{R}^N)$. Then if X is the normed space of functions $f \in C^M(\mathbb{R}^N)$ where $\|f\|_{M,2} < \infty$, equipped with $\|\cdot\|_{M,2}$, Theorem B.1 is reduced to proving that $C_c^\infty(\mathbb{R}^N)$ is dense in X. The proof is divided into two steps. Let $Y := C_c^M(\mathbb{R}^N)$ be equipped with the norm $\|\cdot\|_{M,2}$, which is a subspace of X. We first use Theorem B.3 and Theorem B.4 to prove that $C_c^\infty(\mathbb{R}^N)$ is dense in Y, and then we show, by a standard *cutoff* argument, that Y is dense in X.

Our first order of business is to prove that $C_c^\infty(\mathbb{R}^N)$ is dense in Y, so let $f \in Y$. Fixing a standard mollifier φ with associated (φ_ε), this ensures that each $f_\varepsilon := \varphi_\varepsilon * f = f * \varphi_\varepsilon$ is a *compactly supported* (clear from Definition B.2) smooth function, that Theorem B.3 applies to f_ε with φ_ε in place of g, and that Theorem B.4 applies to the partials of f_ε with the partials of f replacing g. More precisely, for each $|\boldsymbol{k}| \le M$ we get $\partial^{\boldsymbol{k}} f_\varepsilon = \varphi_\varepsilon * \partial^{\boldsymbol{k}} f$, and thus $\partial^{\boldsymbol{k}} f_\varepsilon \to \partial^{\boldsymbol{k}} f$ in $L^2(\mathbb{R}^N)$ as $\varepsilon \to 0$. By dominated convergence, as $d\gamma_N(x) = \psi(x)\,dx$ for a bounded function $\psi(x)$, this implies that $\partial^{\boldsymbol{k}} f_\varepsilon \to \partial^{\boldsymbol{k}} f$ also in $L^2(\gamma_N)$, so $f_\varepsilon \to f$ in Y.

It remains to verify that Y is dense in X. We accomplish this using a sequence of smooth *cutoff* functions. Namely, we can find a constant $A > 0$ and a sequence of smooth functions $0 \le \chi_n \le 1$ on $\mathbb{R}^N$ for which χ_n is 1 on $\bar{B}_n(0)$, 0 outside $B_{n+1}(0)$, and $\|\partial^{\boldsymbol{k}}\chi_n\|_\infty \le A$ for all $|\boldsymbol{k}| \le M$. Much like with mollifiers, we can define radially symmetric χ_n starting from just a particular choice of $\chi = \chi_1$ in 1 dimension by setting

$$\chi_n(x) := \chi(\max\{|x| - n + 1, 0\}). \tag{B.14}$$

Let $f \in X$ and define $f_n = f\chi_n$ for all positive integers n. Then $f_n \in Y$, and for all $|\boldsymbol{k}| \le M$ we note that $\partial^{\boldsymbol{k}} f_n = \partial^{\boldsymbol{k}} f$ for $|x| < n$ and $\partial^{\boldsymbol{k}} f_n = 0$ for $|x| > n+1$. Hence

$$\|\partial^{\boldsymbol{k}} f_n - \partial^{\boldsymbol{k}} f\|_{\gamma_N}^2 = \int_{n\le|x|\le n+1} |\partial^{\boldsymbol{k}} f_n - \partial^{\boldsymbol{k}} f|^2\, d\gamma_N + \int_{|x|>n+1} |\partial^{\boldsymbol{k}} f|^2\, d\gamma_N. \tag{B.15}$$

The right summand approaches 0 as $n \to \infty$ by the dominated convergence theorem, since $\partial^{\boldsymbol{k}} f \in L^2(\gamma_N)$. For the left summand, we can use the generalized Leibniz rule to write

$$\partial^{\boldsymbol{k}} f_n - \partial^{\boldsymbol{k}} f = \sum_{0<\boldsymbol{\ell}\le\boldsymbol{k}} \binom{\boldsymbol{k}}{\boldsymbol{\ell}} \partial^{\boldsymbol{\ell}}\chi_n \partial^{\boldsymbol{k}-\boldsymbol{\ell}} f, \tag{B.16}$$

where the binomial coefficients are defined as usual, but with the factorial convention for multi-indices. By the triangle inequality, (B.16) allows us to bound

$$|\partial^{\boldsymbol{k}} f_n - \partial^{\boldsymbol{k}} f| \le A' \sum_{0<\boldsymbol{\ell}\le\boldsymbol{k}} |\partial^{\boldsymbol{k}-\boldsymbol{\ell}} f|, \tag{B.17}$$

where A' is equal to A times the maximum of the binomial coefficients. Then we can use the Minkowski inequality along with dominated convergence to show that also

$$\lim_{n\to\infty} \int_{n\le|x|\le n+1} |\partial^{\boldsymbol{k}} f_n - \partial^{\boldsymbol{k}} f|^2\, d\gamma_N = 0, \tag{B.18}$$

so $\|\partial^{\boldsymbol{k}} f_n - \partial^{\boldsymbol{k}} f\|_{\gamma_N}^2 \to 0$. Thus $f_n \to f$ in X, as desired.

C Omitted Proofs

The proofs in this section are of results which have already been stated in the body of the thesis, so we do not bother restating the theorems/lemmas.

Proof. (Theorem 1.13) First we claim that G is almost surely invertible. It suffices to prove that the set of singular matrices has Lebesgue measure 0 inside $M_N\mathbb{R}$, for G is a continuous random variable. The set of singular matrices is the vanishing locus of the determinant, a nonzero polynomial in N^2 variables. We claim more generally that for any positive integer n and every nonzero $p \in \mathbb{R}[x_1, \ldots, x_n]$, the vanishing locus $\mathcal{V}(p)$ of p has measure 0.

We prove our claim by induction on n. If $n = 1$, then since every nonzero $p \in \mathbb{R}[x]$ has finitely many roots, the set $\mathcal{V}(p)$ is finite and in particular of measure 0. Now suppose our claim holds for some positive integer n, and let $p \in \mathbb{R}[x_1, \ldots, x_n, x_{n+1}]$ be nonzero. For each $x \in \mathbb{R}^n$, let $p_x \in \mathbb{R}[x_1]$ denote the polynomial $p_x(x_1) := p(x_1, x)$. There are finitely many polynomials $q_0, q_1, \ldots, q_d \in \mathbb{R}[x_1, \ldots, x_n]$ for which $p_x(x_1) = \sum_{i=0}^{d} q_i(x) x_1^i$, so that $p_x \equiv 0$ if and only if $q_i(x) = 0$ for all $i \leq d$. In particular,

$$E := \{x \in \mathbb{R}^n \mid p_x \equiv 0\} = \cap_{i=0}^{d} \mathcal{V}(q_i) \tag{C.1}$$

has measure 0 in $\mathbb{R}^n$ by the induction hypothesis. Hence by Tonelli's theorem and the already proven base case, we have

$$\lambda_{n+1}(\mathcal{V}(p)) = \int_{\mathbb{R}^n} \lambda_1(\mathcal{V}(p_x))\, d\lambda_n(x) = \int_{\mathbb{R}^n - E} \lambda_1(\mathcal{V}(p_x))\, d\lambda_n(x) = 0, \tag{C.2}$$

where λ_k is the Lebesgue measure on $\mathbb{R}^k$. This proves our claim.

Now, define a Borel measurable function $f : M_N\mathbb{R} \to \mathrm{O}_N$ by letting $f(A)$ be the matrix obtained by applying Gram-Schmidt to the columns of A if $A \in \mathrm{GL}_N\mathbb{R}$, and otherwise $f(A) = I$. We wish to show that $f(G) \sim \mathrm{Haar}(\mathrm{O}_N)$. The distribution of $f(G)$ is a Borel probability measure on the Polish space O_N, and hence it is automatically a Riesz measure. It is thus sufficient to prove that this distribution is left-invariant, or equivalently $Of(G) \sim f(G)$ for every $O \in \mathrm{O}_N$. Write $g_1, \ldots, g_N$ for the columns of G, which are independently distributed $\mathcal{N}(0, I_N)$. Then the columns of OG are $Og_1, \ldots, Og_N$, and $O \sim OG$. Hence $f(G) \sim f(OG)$, so it suffices to prove that $f(OA) = Of(A)$ for all $A \in \mathrm{GL}_N\mathbb{R}$. Intuitively, a basis for $\mathbb{R}^N$ determines a complete flag, and Gram-Schmidt constructs an orthonormal basis preserving this flag in an "oriented" fashion, which commutes with rotations.

Let $a_1, \ldots, a_N$ be a basis for $\mathbb{R}^N$, and suppose that $b_1, \ldots, b_N$ is the basis resulting from applying Gram-Schmidt to $a_1, \ldots, a_N$. Let $c_1, \ldots, c_N$ be the basis resulting from applying Gram-Schmidt to $Oa_1, \ldots, Oa_N$. By induction on i we will demonstrate that $Ob_i = c_i$ for all $1 \leq i \leq N$. For the base case $i = 1$, just note that

$$Ob_1 = O\left(\frac{a_1}{\|a_1\|}\right) = \frac{Oa_1}{\|Oa_1\|} = c_1. \tag{C.3}$$

Now suppose that the claim holds for all $j < i$, where $2 \leq i \leq N$. Then

$$Ob_i = O\left(\frac{a_i - \sum_{j<i}\langle a_i, b_j\rangle b_j}{\|a_i - \sum_{j<i}\langle a_i, b_j\rangle b_j\|}\right) = \frac{Oa_i - \sum_{j<i}\langle Oa_i, Ob_j\rangle Ob_j}{\|Oa_i - \sum_{j<i}\langle Oa_i, Ob_j\rangle Ob_j\|} \tag{C.4}$$

$$= \frac{Oa_i - \sum_{j<i}\langle Oa_i, c_j\rangle c_j}{\|Oa_i - \sum_{j<i}\langle Oa_i, c_j\rangle c_j\|} = c_i. \tag{C.5}$$

□

Proof. (Theorem 2.2) First suppose that μ satisfies the isoperimetric inequality (2.2). Let f be a 1-Lipschitz function on $\mathcal{X}$, $m \in \mathrm{med}(f)$, and $A = \{f \le m\}$. Then $\mu(A) \ge 1/2$, and so for every $\varepsilon > 0$,

$$\mu(\mathcal{X} - A^{\varepsilon}) \le C \exp(-\varepsilon^2/2\sigma^2). \tag{C.6}$$

Now fix $\varepsilon > 0$. Suppose that $x \in \mathcal{X}$ is such that $f(x) - m \ge \varepsilon$. Then if $y \in A$, we have

$$f(y) \le m \le f(x) - \varepsilon \implies f(x) - f(y) \ge \varepsilon. \tag{C.7}$$

Because f is 1-Lipschitz, $d(x,y) \ge |f(x) - f(y)| \ge f(x) - f(y) \ge \varepsilon$, so $\{f - m \ge \varepsilon\} \subseteq \mathcal{X} - A^{\delta}$ for every $\delta < \varepsilon$. In particular,

$$\mu(f - m \ge \varepsilon) \le \mu(\mathcal{X} - A^{\delta}) \le C \exp(-\delta^2/2\sigma^2) \tag{C.8}$$

for all $\delta < \varepsilon$. Letting $\delta \to \varepsilon^-$, we obtain (2.6).

Conversely, suppose that (2.6) holds for all 1-Lipschitz f, medians m, and $\varepsilon > 0$. Let $A \subseteq \mathcal{X}$ be measurable such that $\mu(A) \ge 1/2$. Consider the function $f : \mathcal{X} \to \mathbb{R}$ defined by $f(x) = d(x, A)$. Thus $\{f \le \varepsilon\} = A^{\varepsilon}$ for every $\varepsilon > 0$. First we claim that f is 1-Lipschitz. Let $\eta > 0$. Then we can find $z \in A$ such that $d(y, z) \le d(y, A) + \eta$. Also $d(x, A) \le d(x, z)$ because $z \in A$, so

$$d(x, A) - d(y, A) \le d(x, z) - d(y, z) + \eta \le d(x, y) + \eta \tag{C.9}$$

by the triangle inequality. Letting $\eta \to 0^+$, we get $f(x) - f(y) \le d(x, y)$. Swapping x and y gives $f(y) - f(x) \le d(x, y)$, so $|f(x) - f(y)| \le d(x, y)$. This verifies that f is 1-Lipschitz.

Now let $\varepsilon > 0$. By (2.6), for all $m \in \mathrm{med}(f)$ we have

$$\mu(f - m \le \varepsilon) \ge \mu(f - m < \varepsilon) \ge 1 - C \exp(-\varepsilon^2/2\sigma^2). \tag{C.10}$$

Since $\{f \le 0\}$ and $\{f \ge 0\}$ are both supersets of A, and $\mu(A) \ge 1/2$, it follows that $0 \in \mathrm{med}(f)$, so we can take $m = 0$ in (C.10). Moreover, notice that $\{f \le \varepsilon\} = A^{\varepsilon}$ by definition. Therefore we obtain the isoperimetric inequality (2.2). $\square$

Proof. (Lemma 3.16) First suppose that V is semisimple, and let $V = \oplus_{i \in I} U_i$ be a decomposition into G-irreps. If $W \subseteq V$ is invariant, then for each $i \in I$ the intersection $W \cap U_i$ is either all of U_i (so that $U_i \subseteq W$) or 0. This implies $W = \oplus_{j \in J} U_j$, where $J = \{i \in I \mid U_i \subseteq W\}$, and we can take $W' = \oplus_{i \in I - J} U_i$. Hence V is completely reducible.

Conversely, suppose that V is completely reducible. If V is FD, we can proceed by induction. When V is irreducible, certainly V is semisimple. Otherwise, we can split $V = W \oplus W'$, where W and W' are proper invariant subspaces of V. If both W and W' are irreducible, then we are done. Otherwise, we further split W and W' as needed. This process eventually terminates for dimensional reasons. If V is infinite-dimensional, the same argument works, but we need some form of *transfinite* induction, such as Zorn's lemma. $\square$

Proof. (Lemma 3.36) Let $\rho : G \to \mathrm{GL}(V)$ be a G-irrep with character χ_V, and for $v \in V$ and $\alpha \in V^*$ define the matrix coefficient $\varphi(g) = \langle \alpha, g \cdot v\rangle$. It suffices to show that $A(\varphi)$ is a scalar multiple of χ_V. The pair (v, α) corresponds to an endomorphism $T : V \to V$ by $T(w) = \alpha(w)v$, and notice $\varphi(g) = \mathrm{tr}(T \circ \rho(g))$. We can average T to obtain another endomorphism

$$A(T) : V \to V, \quad w \mapsto \int (g \cdot T)(w)\, dg. \tag{C.11}$$

Here G acts as usual on $\mathrm{End}(V) \cong V \otimes V^*$ by $g \cdot T = \rho(g) \circ T \circ \rho(g)^{-1}$. Invariance implies that $A(T)$ is G-linear, so $A(T) = \lambda \cdot \mathrm{id}_V$ for some $\lambda \in \mathbb{C}$ by Schur's lemma. Hence

$$A(\varphi)(g) = \int \varphi(hgh^{-1})\, dh = \int \mathrm{tr}(T \circ \rho(hgh^{-1}))\, dh = \int \mathrm{tr}((h \cdot T) \circ \rho(g))\, dh \tag{C.12}$$

$$= \mathrm{tr}\left(\int h \cdot T\, dh \circ \rho(g)\right) = \lambda \mathrm{tr}(\rho(g)) = \lambda \chi_V(g). \tag{C.13}$$

□

D Supplementary Code

All simulations done for this thesis are original, as are the *Mathematica* calculations detailed in Table 8. I thank my friend and colleague Lev Kruglyak for his assistance. Below is a link to a GitHub repository containing the relevant code, for completeness and reproducibility:

https://github.com/ajlamotta/Senior-Thesis.

6 Bibliography

[BKS03] I. Benjamini, G. Kalai, and O. Schramm. "First Passage Percolation Has Sublinear Distance Variance". In: *Ann. Probab.* 31.4 (2003), pp. 1970–1978.

[BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford Series in Probability and Statistics. Oxford University Press, 2013. ISBN: 9780199535255. URL: https://books.google.com/books/about/Concentration_Inequalities.html?id=koNqWRluhP0C.

[BS15] Bhaswar B. Bhattacharya and Subhabrata Sen. "High Temperature Asymptotics of Orthogonal Mean-Field Spin Glasses". In: *Journal of Statistical Physics* 162.1 (Oct. 2015), pp. 63–80. ISSN: 1572-9613. DOI: 10.1007/s10955-015-1406-7. URL: http://dx.doi.org/10.1007/s10955-015-1406-7.

[Cha09] Sourav Chatterjee. *Disorder Chaos and Multiple Valleys in Spin Glasses*. 2009. arXiv: 0907.3381 [math.PR]. URL: https://arxiv.org/abs/0907.3381.

[Cha14] Sourav Chatterjee. *Superconcentration and Related Topics*. Springer Monographs in Mathematics. Springer, Cham, 2014, pp. x+156. ISBN: 978-3-319-03885-8. DOI: 10.1007/978-3-319-03886-5. URL: https://doi.org/10.1007/978-3-319-03886-5.

[Cha23] Sourav Chatterjee. "Spin Glass Phase at Zero Temperature in the Edwards-Anderson Model". In: *arXiv* (2023). DOI: 10.48550/arXiv.2301.04112. eprint: 2301.04112. URL: https://arxiv.org/abs/2301.04112.

[CL24] Wei-Kuo Chen and Wai-Kit Lam. "Universality of Superconcentration in the Sherrington-Kirkpatrick Model". In: *Random Structures Algorithms* 64.2 (2024), pp. 267–286. ISSN: 1042-9832,1098-2418. DOI: 10.1002/rsa.21183. URL: https://doi.org/10.1002/rsa.21183.

[CŚ06] Benoît Collins and Piotr Śniady. "Integration with respect to the Haar measure on unitary, orthogonal and symplectic group". In: *Comm. Math. Phys.* 264.3 (2006), pp. 773–795. ISSN: 0010-3616,1432-0916. DOI: 10.1007/s00220-006-1554-3. URL: https://doi.org/10.1007/s00220-006-1554-3.

[CX21] Hong-Bin Chen and Jiaming Xia. *Limiting Free Energy of Multi-Layer Generalized Linear Models*. 2021. arXiv: 2108.12615 [math.PR].

[DMS17] Amir Dembo, Andrea Montanari, and Subhabrata Sen. "Extremal Cuts of Sparse Random Graphs". In: *Ann. Probab.* 45.2 (2017), pp. 1190–1217. ISSN: 0091-1798,2168-894X. DOI: 10.1214/15-AOP1084. URL: https://doi.org/10.1214/15-AOP1084.

[DS94] Persi Diaconis and Mehrdad Shahshahani. "On the Eigenvalues of Random Matrices". In: *Journal of Applied Probability* 31A (1994), pp. 49–62.

[EMS23] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. "Local Algorithms for Maximum Cut and Minimum Bisection on Locally Treelike Regular Graphs of Large Degree". In: *Random Structures Algorithms* 63.3 (2023), pp. 689–715. ISSN: 1042-9832,1098-2418. DOI: 10.1002/rsa.21149. URL: https://doi.org/10.1002/rsa.21149.

[FH91] William Fulton and Joe Harris. *Representation Theory*. Vol. 129. Graduate Texts in Mathematics. A first course, Readings in Mathematics. Springer-Verlag, New York, 1991, pp. xvi+551. ISBN: 0-387-97527-6. DOI: 10.1007/978-1-4612-0979-9. URL: https://doi.org/10.1007/978-1-4612-0979-9.

[FLS22] Zhou Fan, Yufan Li, and Subhabrata Sen. *TAP Equations for Orthogonally Invariant Spin Glasses at High Temperature*. 2022. arXiv: 2202.09325 [math.PR]. URL: https://arxiv.org/abs/2202.09325.

[Fre] Daniel Freed. *Differential Geometry*. URL: https://people.math.harvard.edu/~dafr/diffgeom.pdf.

[FW24] Zhou Fan and Yihong Wu. *The Replica-Symmetric Free Energy for Ising Spin Glasses With Orthogonally Invariant Couplings*. 2024. arXiv: 2105.02797 [math.PR]. URL: https://arxiv.org/abs/2105.02797.

[Gaa] Onno van Gaans. *Probability Measures on Metric Spaces*. URL: https://pub.math.leidenuniv.nl/~gaansowvan/jancol1.pdf.

[GM67] Siegfried Grosser and Martin Moskowitz. "On central topological groups". In: *Trans. Amer. Math. Soc.* 127 (1967), pp. 317–340. ISSN: 0002-9947,1088-6850. DOI: 10.2307/1994651. URL: https://doi.org/10.2307/1994651.

[Gue03] Francesco Guerra. "Broken Replica Symmetry Bounds in the Mean Field Spin Glass Model". In: *Communications in Mathematical Physics* 233.1 (2003), pp. 1–12.

[Hal15] Brian Hall. *Lie groups, Lie algebras, and representations*. Second. Vol. 222. Graduate Texts in Mathematics. An elementary introduction. Springer, Cham, 2015, pp. xiv+449. ISBN: 978-3-319-13466-6. DOI: 10.1007/978-3-319-13467-3. URL: https://doi.org/10.1007/978-3-319-13467-3.

[HW65] J. M. Hammersley and D. J. A. Welsh. "First-Passage Percolation, Subadditive Processes, Stochastic Networks, and Generalized Renewal Theory". In: *Bernoulli 1713 Bayes 1763 Laplace 1813: Anniversary Volume*. Springer, Berlin, Heidelberg, 1965, pp. 61–110. DOI: 10.1007/978-3-642-99884-3_7. URL: https://doi.org/10.1007/978-3-642-99884-3_7.

[Inc] Wolfram Research Inc. *Mathematica, Version 14.1*. Champaign, IL, 2024. URL: https://www.wolfram.com/mathematica.

[Jon+23] Chris Jones et al. "Random Max-CSPs Inherit Algorithmic Hardness From Spin Glasses". In: *14th Innovations in Theoretical Computer Science Conference*. Vol. 251. LIPIcs. Leibniz Int. Proc. Inform. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2023, Art. No. 77, 26. ISBN: 978-3-95977-263-1. DOI: 10.4230/lipics.itcs.2023.77. URL: https://doi.org/10.4230/lipics.itcs.2023.77.

[Kal21] Olav Kallenberg. *Foundations of Modern Probability*. Third. Vol. 99. Probability Theory and Stochastic Modelling. Springer, Cham, 2021, pp. xii+946. ISBN: 978-3-030-61871-1. DOI: 10.1007/978-3-030-61871-1. URL: https://doi.org/10.1007/978-3-030-61871-1.

[Kes86] H. Kesten. "Aspects of First-Passage Percolation". In: *École d'été de probabilités de Saint-Flour XIV*. Vol. 1180. Lecture Notes in Mathematics. Berlin: Springer, 1986, pp. 125–264.

[Kes93] H. Kesten. "On the Speed of Convergence in First-Passage Percolation". In: *Ann. Appl. Probab.* 3.2 (1993), pp. 296–338.

[Kna86] Anthony W. Knapp. *Representation Theory of Semisimple Groups*. Vol. 36. Princeton Mathematical Series. An overview based on examples. Princeton University Press, Princeton, NJ, 1986, pp. xviii+774. ISBN: 0-691-08401-7. DOI: 10.1515/9781400883974. URL: https://doi.org/10.1515/9781400883974.

[Led01] Michel Ledoux. *The Concentration of Measure Phenomenon*. Vol. 89. Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, 2001, pp. x+181. ISBN: 0-8218-2864-9. DOI: 10.1090/surv/089. URL: https://doi.org/10.1090/surv/089.

[Lee13] John M. Lee. *Introduction to Smooth Manifolds*. Second. Vol. 218. Graduate Texts in Mathematics. Springer, New York, 2013, pp. xvi+708. ISBN: 978-1-4419-9981-8.

[Mil] Dragan Miličić. *Notes on Representations of Compact Groups*. URL: https://www.math.utah.edu/~milicic/Math_6260/compact.pdf.

[MZ55] Deane Montgomery and Leo Zippin. *Topological Transformation Groups*. Interscience Publishers, New York-London, 1955, pp. xi+282.

[Pan18] Dmitry Panchenko. "On the K-SAT Model with Large Number of Clauses". In: *Random Structures Algorithms* 52.3 (2018), pp. 536–542. ISSN: 1042-9832,1098-2418. DOI: 10.1002/rsa.20748. URL: https://doi.org/10.1002/rsa.20748.

[Par79] Georgio Parisi. "Infinite Number of Order Parameters for Spin-Glasses". In: *Physical Review Letters* 43.23 (Dec. 1979), pp. 1754–1756. DOI: 10.1103/PhysRevLett.43.1754.

[Rud87] Walter Rudin. *Real and Complex Analysis*. Third. McGraw-Hill Book Co., New York, 1987, pp. xiv+416. ISBN: 0-07-054234-1.

[Rud91] Walter Rudin. *Functional Analysis*. Second. International Series in Pure and Applied Mathematics. McGraw-Hill, Inc., New York, 1991, pp. xviii+424. ISBN: 0-07-054236-8.

[SK75] David Sherrington and Scott Kirkpatrick. "Solvable Model of a Spin-Glass". In: *Physical Review Letters* 35.26 (Dec. 1975), pp. 1792–1796. DOI: 10.1103/PhysRevLett.35.1792. URL: https://link.aps.org/doi/10.1103/PhysRevLett.35.1792.

[Sub21] Eliran Subag. *TAP Approach for Multi-Species Spherical Spin Glasses I: General Theory*. 2021. arXiv: 2111.07132 [math.PR].

[Tal06] Michel Talagrand. "The Parisi Formula". In: *Annals of Mathematics* 163.1 (2006), pp. 221–263. DOI: 10.4007/annals.2006.163.221. URL: https://annals.math.princeton.edu/wp-content/uploads/annals-v163-n1-p04.pdf.

[Tal11] Michel Talagrand. *Mean Field Models for Spin Glasses. Volume I*. Vol. 54. Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics. Basic examples. Springer-Verlag, Berlin, 2011, pp. xviii+485. ISBN: 978-3-642-15201-6. DOI: 10.1007/978-3-642-15202-3. URL: https://doi.org/10.1007/978-3-642-15202-3.

[Ton17] David Tong. *Statistical Field Theory. University of Cambridge Part III Mathematical Tripos*. 2017. URL: https://www.damtp.cam.ac.uk/user/tong/sft/one.pdf.

[TW09] Craig A. Tracy and Harold Widom. "The Distributions of Random Matrix Theory and Their Applications". In: *New Trends in Mathematical Physics*. Ed. by Vladas Sidoravičius. Dordrecht: Springer Netherlands, 2009, pp. 753–765. ISBN: 978-90-481-2810-5. DOI: 10.1007/978-90-481-2810-5_48.