Analytical Methods for Multivariate Stable Probability Distributions

by

Marcus Pivato

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of of Mathematics
University of Toronto

Analytical Methods for Multivariate Stable Probability Distributions,
Thesis submitted for the degree of Doctor of Philosophy,
2001,
Marcus J. Pivato,
Department of Mathematics,
University of Toronto

Abstract

Stable probability distributions (SPDs) are generalizations of the familiar Gaussian distribution, but have infinite variance. Thus, statistical techniques relying on finite variance are inapplicable to SPDs. In particular, the correlation structure of multivariate SPDs is much more complex than a Gaussian; it is described by a measure on the sphere, called a **spectral measure**, and is poorly understood.

In this work, the relationship between multivariate SPDs and their spectral measures is illuminated, and tools are developed for the statistical analysis of multivariate SPDs. Methods from nonabelian harmonic analysis are applied to express the spectral measure using spherical Fourier series; this leads to an efficient and practical method for estimating spectral measures from empirical data, even in very high dimensions. Formulae are computed which relate the estimation error of a spectral measure back to the estimation error of the original SPD. These results are then applied to the identification and analysis of stable stochastic processes.

A mio nonno,
Antonio Pivato

Acknowledgements

I would like to thank Luis Seco for suggesting this direction of research, and for inspiring others, through his example, to confront difficult problems, and solve them.

I also thank committee members Jeremy Quastel, Keith Knight, and Don McLeish for their careful reading of the thesis and many valuable suggestions and corrections.

I would also like to thank Thomas Hurd, Jeffrey Boland, and Nicolas Hernandez for many fruitful discussions, Fotios Harmantzis for suggesting some excellent literature, and Gustavo Comezana for many entertaining conversations and much programming wisdom.

I would like to thank Ida Bulat, the Department of Mathematics graduate secretary and Angel of Mercy, for her limitless patience and generosity.

Finally, I thank Reem Yassawi, and Joseph, Emma, and Juliana Pivato, for their unwavering support and encouragement.

Contents

List of Figures

Chapter 1

Introduction

1.1 Introducing Stable Random Variables

Stable probability distributions are the natural generalizations of the normal distribution, and share with it two key properties:

- **Stability:** The normal distribution is **stable** in the sense that, if $\mathbf{X}$ and $\mathbf{Y}$ are independent random variables, with identical normal distributions, then $\mathbf{X} + \mathbf{Y}$ is also normal, and its distribution is the same as $\mathbf{X}$ and $\mathbf{Y}$ once it is "renormalized" by multiplying by $2^{-1/2}$. Formally, we can write:

$$\frac{1}{2^{1/2}}(\mathbf{X} + \mathbf{Y}) \quad \stackrel{\cong}{_{distr}} \quad \mathbf{X} \quad \stackrel{\cong}{_{distr}} \quad \mathbf{Y}$$

  In a similar fashion, if $\mathbf{X}$ and $\mathbf{Y}$ are independent, identically distributed (i.i.d) stable random variables, then $\mathbf{X}+\mathbf{Y}$ is also stable, and its distribution is the same as $\mathbf{X}$ and $\mathbf{Y}$ when renormalized by $2^{-1/\alpha}$ (see Theorem 74 on page 130). The **stability exponent** α ranges from 0 to 2 . When $\alpha = 2$, we have the familiar normal distribution.

- **Renormalization Limit:** The Central Limit Theorem says that the normal distribution is the natural limiting distribution of a "suitably renormalized" infinite sum of independent random variables with finite variance. If $\mathbf{X}_1, \mathbf{X}_2, \ldots$ is a sequence of such variables, then the random variables

$$\frac{1}{N^{1/2}} \sum_{n=1}^{N} \mathbf{X}_n$$

converge, in density, to a normal distribution. Similarly, if $\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3, \ldots$ are independent random variables whose distributions decay according to a power law with exponent $-1-\alpha$, then the random variables

$$\frac{1}{N^{1/\alpha}} \sum_{n=1}^{N} \mathbf{Y}_n$$

converge, in distribution, to an α-stable distribution (see Theorem 74 on page 130).

Thus, stable distributions model random aggregations of many small, independent perturbations. For example, stable distributions may model the motions of Markovian stochastic processes whose increments exhibit power laws. Power law distributions arise with surprising frequency in certain systems, especially those involving many independent interacting units with sensitive dependencies between them (see Appendix A.3 on page 137, Appendix B.2 on page 146, or Appendix C on page 153). A better understanding of stable distributions may be important to the study of such phenomena.

1.2 Statement of the Problem

Although one-dimensional stable distributions are well-understood, there are many open questions in the multivariate regime. The simplicity of the multivariate Gaussian universe does not extend to non-Gaussian multivariate stable distributions. An N-dimensional Gaussian distribution is completely determined by its $N \times N$ covariance matrix, which transforms nicely under linear changes of coordinates. In particular, by orthogonally diagonalizing the matrix, we can find an orthonormal basis for $\mathbb{R}^N$, relative to which the multivariate normal variable is revealed as a sum of independent univariate normal variables —this is the so-called *Principal Component Analysis*.

For a general multivariate stable distribution, however, the situation is much more complex. Since the marginals do not have finite variance, it does not make sense to define a "covariance matrix" in the usual way; none of the integrals would converge. Various modified notions of "covariance" have been proposed (see, for example, [154]), but these do not transform in any simple way under changes of coordinates. In particular, there is nothing analogous to a "principal components analysis".

The correlation structure of a stable distribution on $\mathbb{R}^D$ is determined by an arbitrary measure on the sphere $\mathbb{S}^{D-1}$, called the **spectral measure**

(see Theorem 77 on page 143). This measure is essentially an "infinite-dimensional" data-structure, so it is clear that, in general, no $N \times N$ matrix can possibly be adequate for representing it. A "principal components" type decomposition is only valid when the spectral measure consists of $2D$ antipodaly positioned atoms.

The goal of the present work is to illuminate the relationship between multivariate stable distributions and their spectral measures, and to develop tools for statistical analysis in the multivariate stable universe.

1.3 Organization of this Work

All relevant mathematical background material is developed in the appendices, and frequent reference is made to these throughout the text. I have attempted to maintain consistent notational conventions; an index of notation appears on page 271.

In Part I, I develop efficient methods for estimating the spectral measure of a multivariate stable distribution from empirical sample data.

In Chapter 3, I develop bounds relating the distance between two stable probability distributions to the distance between their respective spectral measures. In Chapter 4, I deal with the characteristic function, and its image, the **spherical log-characteristic function**. In Chapters 5 and 6, I develop "convolution" style methods to extract the spectral measure from information about the spherical log characteristic function.

In Part II, I briefly explore some other statistical methods in the stable universe. In Chapter 7, I investigate applications of of **normal rank correlation analysis** to multivariate stable distributions, and demonstrate that NRC analysis clearly distinguishes between distributions with independent coordinates and those with correlated coordinates.

In Chapter 8, I explore the problem of analyzing data from "mixed" stable distributions: that is, the distributions of sums of independent multivariate stable random variables $\mathbf{X}_1, \ldots, \mathbf{X}_N$, where $\mathbf{X}_n$ has stability exponent α_n, and $\alpha_1 < \ldots < \alpha_N$.

In Chapter 9, I develop techniques for identifying a stable stochastic process from empirical data. I consider the case where the process is defined via a stochastic integral over random α-stable noise. I show how to recover the stability exponent α of the noise and (if the noise is multidimensional) its spectral measure (§9.2). In the case of a *moving average* process, I show how

to reconstruct the shape of the averaging kernel from the spectral measure of marginals of the process (§9.1). Finally, in §9.3 I show how to decompose a broad class of moving average process into linear combinations of Ornstein-Uhlenbeck type processes, thereby providing a simple physical mechanism for them.

Statement of Originality

Material in the appendices is not original, but is provided as background material. All results proved in the numbered chapters are, to the best of my knowledge, original work, unless explicitly stated otherwise. When non original results appear in the numbered chapters, they will be indicated explicitly by a $(*)$.

Part I

Estimation of Spectral Measures

Chapter 2

Motivation and Context

A multivariate stable probability distribution is completely described by its **shift parameter** $\vec{\mu}$ (which, for $\alpha > 1$, is the mean of the distribution) and a **spectral measure** Γ, a measure on the sphere $\mathbb{S}^{D-1}$ which describes the multivariate correlation structure of the distribution. Estimation of $\vec{\mu}$ can be performed componentwise, by simply applying the assorted estimation procedures developed in the univariate case (see Appendix A.4 on page 140). Estimation of Γ is more difficult. Γ is not a parameter, or even a finite collection of parameters, but is, in a sense, an "infinite-dimensional" data structure. Furthermore, Γ is only indirectly visible; the image of Γ under a sort of "spherical convolution" appears in the *logarithm* of the *characteristic function* of the distribution, and there is no more direct way to observe it.

2.1 Summary of Previous results

Early on, Press [141] developed an estimation scheme for multivariate stable distributions, through a straightforward generalization of his one-dimensional method. Press's method, however, only works for "pseudo-Gaussian" distributions[1], with log-characteristic functions of the form:

$$\Phi_{\mathbf{X}}(\vec{\xi}) \quad = \quad \left\langle \vec{\xi}, \vec{\mu} \right\rangle \mathbf{i} \; + \; \left\langle \vec{\xi}, \Omega\vec{\xi} \right\rangle^{\alpha/2}$$

where Ω is some symmetric, positive semidefinite "covariance matrix". If Ω has unit eigenvectors $\vec{\omega}_1, \ldots, \vec{\omega}_D$, with eigenvalues $\lambda_1, \ldots \lambda_D$ (ie. as a covariance matrix, we have "principal components" $\lambda_1\vec{\omega}_1, \ \ldots, \ \lambda_1\vec{\omega}_1$), then

[1] **Note:** "Pseudo-Gaussian" distributions are *not* the same as a *sub-Gaussian* distributions; indeed, in a qualitative sense, they are almost opposite.

the spectral measure of this distribution is symmetric and atomic, with atoms at each of $\pm\vec{\omega}_1, \ldots, \pm\vec{\omega}_D$, with masses $\lambda_1, \ldots \lambda_D$ —in other words:

$$\Gamma = \sum_{d=1}^{D} \lambda_d \left(\delta_{\vec{\omega}_d} + \delta_{-\vec{\omega}_d} \right)$$

Press proposes to solve for the components of the matrix Ω by empirically estimating the log characteristic function at some collection of frequencies $\{\vec{\xi}_1, \ldots, \vec{\xi}_N\}$, where $N = D(D+1)/2$, and then solving a system of N linear equations. He claims that his method will generalize to a *sum* of pseudo-Gaussians:

$$\Phi_{\mathbf{X}}(\vec{\xi}) \quad = \quad \left\langle \vec{\xi}, \vec{\mu} \right\rangle \mathbf{i} \; + \; \sum_{m=1}^{M} \left\langle \vec{\xi}, \Omega_m \vec{\xi} \right\rangle^{\alpha/2}$$

(where $\Omega_1, \ldots, \Omega_M$ are linearly independent, symmetric, positive semidefinite matrices). However, in this case, one no longer ends up with a system of linear equations, so it is not clear that the method is tractable. In any event, Press's method only applies to multivariate distributions with particularly simple atomic spectral measures, which furthermore must be symmetrically distributed. Empirical evidence (see, for example, [17]) suggests that the stable distributions found in financial data are significantly skewed; symmetry is not a reasonable assumption.

Cheng, Rachev and Xin [166],[19] develop a more sophisticated method, by expressing a stable random vector in polar coordinates, and then examining the order statistics of the radial component, as a function of the angular component. They utilize the theorem of Araujo and Giné (Theorem 78 on page 145) stating that the radial distribution decays most slowly in those angular directions with the heaviest concentration of spectral mass; these differences in decay rate are then used to estimate the density distribution of the spectral measure.

More generally, Hurd *et al.* [113] consider any multivariate, infinitely-divisible distribution ρ whose Lévy-Khintchine measure λ takes the form

$$d\lambda[r.\theta] \quad = \quad f(r) \; dr \; d\Gamma[\theta]$$

where $\theta \in \mathbb{S}^{D-1}$ and Γ is some "spectral measure" on $\mathbb{S}^{D-1}$, while $r \in [0, \infty)$, and $f : [0, \infty) \longrightarrow [0, \infty)$ is some function asymptotically of order $f(r) \sim \mathcal{O}\left(r^{-\alpha-1}\right)$. A result similar to that of Araujo and Giné (Theorem 78 on page 145) is shown for this class of distributions, providing a mechanism for

estimating Γ from empirical data by looking at the angular distribution of extremal events.

Nolan, Panorska, and McCulloch [81] develop a method based upon a discrete approximation of the spectral measure. If the spectral measure is treated as a sum of a finite number of atoms,

$$\Gamma \;=\; \sum_{\mathbf{a}\in\mathcal{A}} \gamma_{\mathbf{a}}\delta_{\mathbf{a}},$$

then, for any fixed $\mathbf{s} \in \mathbb{S}^{D-1}$, the "convolution kernel" $\eta_{\mathbf{s}}$ (see Section 4.1 on page 43) can be restricted to a function $\eta_{\mathbf{s}} : \mathcal{A} \longrightarrow \mathbb{C}$. The set of all discrete measures supported on $\mathcal{A}$ is a finite-dimensional vector space, which we can identify with $\mathbb{C}^{\mathcal{A}}$, and $\eta_{\mathbf{s}}$ is just a linear functional on this vector space. If $\Xi \subset \mathbb{S}^{D-1}$ is some finite set, then we can define a linear map:

$$F : \mathbb{C}^{\mathcal{A}} \longrightarrow \mathbb{C}^{\Xi}$$

where, for each $\mathbf{s} \in \Xi$,

$$F(\Gamma)_{\mathbf{s}} \;=\; \Phi(\mathbf{s}) \;=\; \int_{\mathbb{S}^{D-1}} \eta_{\mathbf{s}}\, d\Gamma$$

The image $F(\Gamma)$ is really just the restriction of the *spherical log-characteristic function* $\mathbf{g}$ introduced in Section 4.1.

The method of Nolan *et al.* then comes down to *inverting* this linear transformation to recover Γ from an empirical estimate of $\mathbf{g} = \Phi_{|\Xi}$. They explicitly implemented their method in the two-dimensional case (ie. when the spectral measure lives on a circle), and tested it against a variety of distributions. They found that it worked fairly well for a variety of measures on the circle, and consistently outperformed the method of Chen *et al.*

Nonetheless, the method of Nolan *et al.* is limited for a number of reasons. First, it depends upon a discrete approximation of the spectral measure. Although error bounds on such discrete approximations are available (see, for example, Section 3.3 on page 34, and also [174]), it is unsatisfactory to always have to resort to such an approximation. In particular, because of the aforementioned theorem of Araujo and Giné [1], the asymptotic decay of the probability distribution is distinctly slower along rays passing through atoms of a spectral measure than it is along the rays in between them. This angular fluctuation in asymptotic decay rate will not be seen in a stable distribution with a continuous spectral measure. Hence, in approximating a continuous measure by a discrete one, we may introduce anomalous asymptotic behaviour to the distribution.

Furthermore, since the method of Nolan *et al.* involves matrix inversion, it is extremely numerically expensive in a high-dimensional space; a good approximation of the spectral measure may require a large number of atoms, so high dimensionality may be unavoidable. Suppose that, to approximate the spectral measure with a certain accuracy, we need a set of atoms with density $\epsilon \sim 1/N$ over the surface of $\mathbb{S}^{D-1}$. Such a set has cardinality of order $\mathcal{O}\left(N^{D-1}\right)$, thus, we are looking at inverting a matrix with $\mathcal{O}\left(N^{D-1}\right)$ rows and columns —a computation of order $\mathcal{O}\left(N^{3(D-1)}\right)$. Applying the inverse matrix to the vector $\mathbf{g}$ is a further computation of order $\mathcal{O}\left(N^{2(D-1)}\right)$.

Finally, the "convolution kernel" η has a strongly "smoothing" effect when transforming Γ into $\mathbf{g}$ (see § 4.3 on page 45, for example). This means that the linear transformation F has many small eigenvalues, which means that the inverse operation is extremely sensitive; small estimation errors in $\mathbf{g}$ can blow up into big errors in Γ. (Indeed, we must be careful when positioning our estimator atoms, or the transformation will not be invertible at all.) These issues are addressed in [78]. The methods of Chen *et al.* and Nolan *et al.* are also discussed in section 5 of [128].

Finally, J. Nolan [130], also develops techniques for identifying when a multivariate distribution is stable, by applying maximum likelihood estimation techniques (see also [127]) to the parameters of the one-dimensional marginals. He also provides a complete parameter estimation scheme in the special case of sub-Gaussian distributions, but does not address the issue of spectral measures in general.

2.2 Overview of Part I:

We will now explore the problem of estimating a multivariate stable distribution from empirical sample data.

In Chapter 3, I develop bounds relating the distance (relative to a variety of different norms) between two stable probability distributions to the distance between their respective spectral measures. These bounds can be used to compute the rate-of-convergence of some approximating series of measures, and to get explicit bounds on estimation errors.

In Chapter 4, we examine the **spherical log-characteristic function**, normally indicated as "$\mathbf{g}$" (see §4.1 for definition). In § 4.3 on page 45 we

characterize the smoothness properties of $\mathbf{g}$, and in § 4.4 on page 47, we use this smoothness to develop estimation procedures for $\mathbf{g}$.

Knowledge of the $\mathbf{g}$ is sufficient to describe the characteristic function of $\mathbf{X}$. However, $\mathbf{g}$ is not very convenient if we want to simulate $\mathbf{X}$ —for this, we need explicit knowledge of the spectral measure Γ. However, Γ is not directly visible in empirical data; instead, all we see is $\mathbf{g}$, which is the *image* of Γ via the linear transformation described in §4.1. The problem, then is to recover Γ from $\mathbf{g}$, by reversing this linear transformation.

The theorems in §4.3 reveal the essential problem: $\mathbf{g}$ is much "smoother" than Γ. The "high frequency" or highly "singular" parts of Γ are very much suppressed when it is converted into $\mathbf{g}$. Hence, the transform $\Gamma \longrightarrow \mathbf{g}$ has many very small eigenvalues; this is exhibited explicitly in Proposition 23 on page 45. Hence, the inverse transform $\mathbf{g} \longrightarrow \Gamma$ will have many very large eigenvalues, associated to the "high frequency" parts of Γ. This means that small errors in our empirical estimate of $\mathbf{g}$ may get blown up into very large errors in our estimation of Γ —or at least, of its high frequency part. The "low frequency" part of Γ should be fairly easy to estimate.

In §4.1, we describe the equation

$$\mathbf{g}(\mathbf{x}) \;=\; \int_{\mathbb{S}^{D-1}} \eta^{\alpha}(\mathbf{x},\theta)\; d\Gamma[\theta] \qquad\qquad (A)$$

as a kind of "convolution" of the measure Γ with the kernel η^{α}. We could reconstruct Γ from the empirical estimate of $\mathbf{g}$ if we could somehow "deconvolve" the η out.

Of course, the expression (A) is not "really" a convolution, because, if $D \neq 1, 2, 4$, then $\mathbb{S}^{D-1}$ is not a topological group [6]. However, we can make sense of this "pseudoconvolution" in at least two ways, and each provides a method of "deconvolution" for recovering Γ.

In Chapter 5, by treating $\mathbb{S}^{D-1}$ as a homogeneous Riemannian manifold, and recognizing that the value $\eta^{\alpha}(\mathbf{x},\theta)$ depends only on the *distance* between $\mathbf{x}$ and θ, we can compute the (de)convolution of η^{α} with an arbitrary function on $\mathbb{S}^{D-1}$ in terms of the **eigenfunctions of the Laplacian** on $\mathbb{S}^{D-1}$. It turns out that the eigenfunctions of the Laplacian on $\mathbb{S}^{D-1}$ are functions known as **spherical harmonics**; they span $\mathbf{L}^2(\mathbb{S}^{D-1})$, and behave nicely under convolution.

In Chapter 6, by considering the action of $\mathbb{SO}^D[\mathbb{R}]$ on $\mathbb{S}^{D-1}$, and the fact that η^{α} is "nicely behaved" under this action, we can "pull back" equation

(A) to a *true* convolution on $\mathbb{SO}^D[\mathbb{R}]$. At this point, we can employ the machinery of nonabelian group representation theory. The **Peter-Weyl Theorem** says that $\mathbf{L}^2\left(\mathbb{SO}^D[\mathbb{R}]\right)$ has an orthonormal basis given by the matrix coefficients of its irreducible unitary representations. A function in $\mathbf{L}^2$ can thus be represented as a linear combination of these **coefficient functions**, in a direct generalization of the Fourier series of classical abelian harmonic analysis. Convolution of two functions corresponds to matrix-wise multiplication of these coefficients, in direct generalization to the familiar multiplication formula for classical Fourier convolution. Hence, we can "deconvolve" by multiplying by suitable matrix inverses.

Chapter 3

Error Propagation: From Spectral Measure to Probability Distribution

Suppose ρ_1 and ρ_2 are stable probability measures with spectral measures Γ_1 and Γ_2, respectively. If Γ_1 is a "good approximation" of Γ_2, then is ρ_1 a "good approximation" of ρ_2?

In this chapter we develop bounds relating the distance between Γ_1 and Γ_2 to that between ρ_1 and ρ_2, relative to a variety of different norms. These bounds can be used to compute the rate of convergence of an approximating series of measures, and to get explicit bounds on estimation errors.

The basic strategy is as follows: Suppose that, for $k = 1, 2$, ρ_k is a stable probability measure on $\mathbb{R}^D$, with characteristic function $\chi_k : \mathbb{R}^D \longrightarrow \mathbb{C}$, log-characteristic function $\Phi_k : \mathbb{R}^D \longrightarrow \mathbb{C}$, and spectral measure Γ_k on $\mathbb{S}^{D-1}$. Thus:

- Φ_k is a function of Γ_k via the transformation given by equation (B.1) of Theorem 77 on page 143.
- χ_k is a function of Φ_k: $\chi_k(\vec{\xi}) = \exp[\Phi_k(\vec{\xi})]$.
- ρ_k is a function of χ_k via inverse Fourier-transform.

If Γ_1 is "close" to Γ_2, we want to establish that ρ_1 is "close" to ρ_2. Our general strategy will be:

- First show that, if χ_1 is "close" to χ_2, then ρ_1 is "close" to ρ_2.
- Next, if Φ_1 is "close" to Φ_2, then χ_1 is "close" to χ_2.

- Finally, if Γ_1 is "close" to Γ_2, then Φ_1 is "close" to Φ_2.

Here "close" will be interpreted in several ways. First, in §3.1, we consider convergence in the weak* topology. Then, in §3.2, we will simultaneously consider $\mathbf{L}^1$ and $\mathbf{L}^\infty$ convergence.

3.1 Weak* Convergence

Proposition 1: *Let ρ and $\left\{\rho_n|_{n\in\mathbb{N}}\right\}$ be stable probability measures on $\mathbb{R}^D$, with spectral measures Γ and $\left\{\Gamma_n|_{n\in\mathbb{N}}\right\}$ on $\mathbb{S}^{D-1}$, respectively.*

$$\textit{If } \Gamma_n \xrightarrow[n\to\infty]{} \Gamma \quad \textit{in the weak* topology on } \mathcal{M}_{\mathcal{EAS}}\left[\mathbb{S}^{D-1}\right]$$

$$\textit{then } \rho_n \xrightarrow[n\to\infty]{} \rho \quad \textit{in the weak* topology on } \mathcal{M}_{\mathcal{EAS}}\left[\mathbb{R}^D\right].$$

Proof: Let ρ_n (respectively ρ) have Fourier transform χ_n (respectively χ), and log-Fourier transform Φ_n (respectively Φ).

Claim 1: $\Phi_n \to \Phi$, *pointwise.*

Proof: Define $\eta : \mathbb{R}^D \times \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$ by $\eta(\vec{\xi},\mathbf{s}) = \left|\left\langle\vec{\xi},\mathbf{s}\right\rangle\right|^\alpha + \mathcal{B}_\alpha\left\langle\vec{\xi},\mathbf{s}\right\rangle^{\langle\alpha\rangle}\mathbf{i}$. Let $\vec{\xi}\in\mathbb{R}^D$. Then $\Phi_n\left(\vec{\xi}\right) = \int_{\mathbb{S}^{D-1}}\eta\left(\vec{\xi},\mathbf{s}\right)\,d\Gamma_n[\vec{s}] \xrightarrow[n\to\infty]{} \int_{\mathbb{S}^{D-1}}\eta\left(\vec{\xi},\mathbf{s}\right)\,d\Gamma[\vec{s}] = \Phi\left(\vec{\xi}\right)$, by weak*-convergence ..... □ [Claim 1]

Thus, since $\chi_n = \exp(\Phi_n)$, we conclude that $\chi_n \xrightarrow[n\to\infty]{} \chi$, pointwise. Now apply the Lévy Continuity Theorem ([46] or [177], Thm. 2.5.1) and conclude that $\rho_n \xrightarrow[n\to\infty]{wk*} \rho$. □

3.2 $\mathbf{L}^p$ Convergence

Suppose ρ_1 and ρ_2 are two stable probability measures on $\mathbb{R}^D$ with density functions $F_1, F_2 : \mathbb{R}^D \longrightarrow [0,\infty)$ and spectral measures Γ_1 and Γ_2. The goal of this section is to establish bounds on the distance $\|F_1 - F_2\|_p$ (for $p\in[1,\infty]$) in terms of the distance between their spectral measures. The formal statements of this result are Corollary 14 and Corollary 15 on page 33. First there is a lot of machinery, however.

3.2.1 Consequences of Fourier Convergence

Suppose that the Fourier transform of ρ_1 is "close" to that of ρ_2 with respect to the $\mathbf{L}^p$ norm, for some $p \in [1, \infty]$. First, how close is ρ_1 to ρ_2? Second, what conditions on Γ_1 and Γ_2 can make this happen?

Notation: For $\vec{\xi} \in \mathbb{R}^D$, define $\mathcal{E}_{\vec{\xi}} : \mathbb{R}^D \longrightarrow \mathbb{C}$ by: $\mathcal{E}_{\vec{\xi}}(\vec{x}) = \exp\left[2\pi\mathbf{i} \cdot \left\langle \vec{\xi}, \mathbf{x} \right\rangle\right]$.

If $g : \mathbb{R}^D \longrightarrow \mathbb{C}$, then $\widehat{g} : \mathbb{R}^D \longrightarrow \mathbb{C}$ is the **Fourier Transform** of g: for all $\boldsymbol{\xi} \in \mathbb{R}^D$, $\widehat{g}(\boldsymbol{\xi}) = \int_{\mathbb{R}^D} g(\mathbf{x}) \cdot \mathcal{E}_{\vec{\xi}}(\mathbf{x})\, d\mathbf{x}$.

Lemma 2: *(Consequences of $\mathbf{L}^\infty$ Fourier Convergence)*
Let ρ_1, ρ_2 be two probability measures on $\mathbb{R}^D$, having Fourier transforms χ_1, χ_2 respectively, and log-Fourier transforms Φ_1, Φ_2 respectively.
Let $\epsilon > 0$. If $\|\chi_1 - \chi_2\|_\infty < \epsilon$, then:

1. *Let $\vec{\xi}_1, \ldots, \vec{\xi}_N \in \mathbb{R}^D$ and $\widehat{f}_1, \ldots, \widehat{f}_N \in \mathbb{C}$, and consider the trigonometric polynomial $f(\vec{x}) = \sum_{n=1}^N \widehat{f}_n \mathcal{E}_{\vec{\xi}_n}$. For any $p, q \in [1, \infty]$ so that $\frac{1}{p} + \frac{1}{q} = 1$,*

$$\left|\mathbf{E}_{\rho_1}^{xpct}[f] - \mathbf{E}_{\rho_2}^{xpct}[f]\right| < \left\|\widehat{f}_1, \widehat{f}_2, \ldots, \widehat{f}_N\right\|_p \cdot N^{1/q} \cdot \epsilon$$

   (where we interpret $N^{1/\infty} = 1$).

2. *For any $f \in \mathbf{L}^1(\mathbb{R}^D)$ with $\widehat{f} \in \mathbf{L}^1(\mathbb{R}^D)$, $\left|\mathbf{E}_{\rho_1}^{xpct}[f] - \mathbf{E}_{\rho_2}^{xpct}[f]\right| < \left\|\widehat{f}\right\|_1 \cdot \epsilon$.*

Proof:

Proof of Part 1: Let $\rho = \rho_1 - \rho_2$, so that ρ has Fourier transform $\chi = \chi_1 - \chi_2$. Then $\mathbf{E}_\rho^{xpct}[f] = \langle f, \rho \rangle = \sum_{n=1}^N \widehat{f}_n \cdot \left\langle \mathcal{E}_{\vec{\xi}_n}, \rho \right\rangle = \sum_{n=1}^N \widehat{f}_n \cdot \chi(\vec{\xi}_n)$.

$$\begin{aligned}
\text{Thus, } \left|\mathbf{E}_{\rho_1}^{xpct}[f] - \mathbf{E}_{\rho_2}^{xpct}[f]\right| &= \left|\mathbf{E}_\rho^{xpct}[f]\right| = \left|\sum_{n=1}^N \widehat{f}_n \cdot \chi(\vec{\xi}_n)\right| \\
&\leq_{(1)} \left\|\widehat{f}_1, \ldots, \widehat{f}_N\right\|_p \cdot \left\|\chi(\vec{\xi}_1), \ldots, \chi(\vec{\xi}_N)\right\|_q \\
&\leq \left\|\widehat{f}_1, \ldots, \widehat{f}_N\right\|_p \cdot N^{1/q} \cdot \max_{1 \leq n \leq N} \left\{\left|\chi\left(\vec{\xi}_n\right)\right|\right\} \\
&\leq \left\|\widehat{f}_1, \ldots, \widehat{f}_N\right\|_p \cdot N^{1/q} \cdot \epsilon
\end{aligned}$$

where step (1) is just Hölder's inequality.

Proof of Part 2: $\mathbf{E}_{\rho}^{xpct}[f] = \langle f, \rho \rangle = \left\langle \widehat{f}, \widehat{\rho} \right\rangle = \left\langle \widehat{f}, \chi \right\rangle$, by Parseval's Formula (see [84], p.132). Thus, $|\mathbf{E}_{\rho_1}^{xpct}[f] - \mathbf{E}_{\rho_2}^{xpct}[f]| = |\mathbf{E}_{\rho}^{xpct}[f]| = \left|\left\langle \widehat{f},\ \chi \right\rangle\right| \leq \left\|\widehat{f}\right\|_1 \|\chi\|_\infty < \left\|\widehat{f}\right\|_1 \epsilon$, by Hölder's inequality. □

Lemma 3: *(Consequences of $\mathbf{L}^1$ Fourier Convergence)*
Let $\left\{\rho_n|_{n\in\mathbb{N}}\right\}$ be a sequence of probability measures on $\mathbb{R}^D$, with density functions $\left\{F_n|_{n\in\mathbb{N}}\right\}$ and Fourier transforms $\left\{\chi_n|_{n\in\mathbb{N}}\right\}$, and let ρ be another probability measure on $\mathbb{R}^D$ with density function F and Fourier transform χ.

Suppose that $\left\{\chi_n|_{n\in\mathbb{N}}\right\}$ and χ are all in $\mathbf{L}^1[\mathbb{R}^D]$.

1. *For all $n \in \mathbb{N}$, $\|F_n - F\|_\infty < \|\chi_n - \chi\|_1$.*
2. *If $\lim_{n\to\infty} \|F_n - F\|_\infty = 0$, then $\lim_{n\to\infty} \|F_n - F\|_1 = 0$, and thus, $\lim_{n\to\infty} \|F_n - F\|_p = 0$, for all $p \in [1, \infty]$.*

Proof:

Proof of Part 1: Since $\left\{\chi_n|_{n\in\mathbb{N}}\right\}$ and χ are all in $\mathbf{L}^1(\mathbb{R}^D)$, use the Fourier Inversion formula to write:

$$F_n = \widehat{\chi_n} \text{ and } F = \widehat{\chi}$$

Thus, $\|F_n - F\|_\infty = \|\widehat{\chi_n} - \widehat{\chi}\|_\infty = \left\|\widehat{\chi_n - \chi}\right\|_\infty \leq \|\chi_n - \chi\|_1$.

Proof of Part 2 (Convergence in $\mathbf{L}^1$): Fix $\epsilon > 0$. Since ρ is a probability distribution, let $\mathbb{K} \subset \mathbb{R}^D$ be some large compact set so that $\rho[\mathbb{K}] > 1 - \frac{\epsilon}{8}$, and thus, $\rho[\mathbb{R}^D \setminus \mathbb{K}] < \frac{\epsilon}{8}$.

Let $K = \mathcal{L}^{bsg}[\mathbb{K}]$, and let N be large enough so that, for all $n > N$, $\|F - F_n\|_\infty < \frac{\epsilon}{8K}$. Thus, $\rho_n[\mathbb{K}] = \int_{\mathbb{K}} F_n \, d\mathcal{L}^{bsg} > \int_{\mathbb{K}} F \, d\mathcal{L}^{bsg} - \|F - F_n\|_\infty \cdot \mathcal{L}^{bsg\,D}[\mathbb{K}] > 1 - \frac{\epsilon}{8} - \frac{\epsilon}{8K}K = 1 - \frac{\epsilon}{4}$, and therefore, $\rho_n[\mathbb{R}^D \setminus \mathbb{K}] < \frac{\epsilon}{4}$.

Thus, $\|F_n - F\|_1 = \int_{\mathbb{R}^D} |F_n - F| \, d\mathcal{L}^{bsg}$

$$\begin{aligned} &= \int_{\mathbb{K}} |F_n - F| \; d\mathcal{L}^{bsg} + \int_{\mathbb{R}^D \setminus \mathbb{K}} |F_n - F| \; d\mathcal{L}^{bsg} \\ &\leq \|F_n - F\|_\infty \, \mathcal{L}^{bsg}[\mathbb{K}] + \int_{\mathbb{R}^D \setminus \mathbb{K}} |F_n| \; d\mathcal{L}^{bsg} + \int_{\mathbb{R}^D \setminus \mathbb{K}} |F| \; d\mathcal{L}^{bsg} \\ &\leq \frac{\epsilon}{8} + \frac{\epsilon}{4} + \frac{\epsilon}{4} = \frac{\epsilon}{2} < \epsilon. \end{aligned}$$

Proof of Convergence in $\mathbf{L}^p$: Let $1 < p < \infty$. Then for any $n \in \mathbb{N}$, $(F_n - F) \in \mathbf{L}^1 \cap \mathbf{L}^\infty(\mathbb{R}^D) \subset \mathbf{L}^p(\mathbb{R}^D)$, and furthermore,

$$\|F_n - F\|_p \leq \|F_n - F\|_1^{(1/p)} \cdot \|F_n - F\|_\infty^{1-(1/p)}$$

(see, for example, [48], Theorem 6.10, p. 177).

Thus, convergence in $\mathbf{L}^1$ and $\mathbf{L}^\infty$ entails convergence in every $\mathbf{L}^p$. ___□

3.2.2 Sufficient conditions for Fourier Convergence, via log-Fourier convergence

Proposition 4: *(A Sufficient condition for $\mathbf{L}^\infty$ Fourier Convergence)*

Let ρ_1, ρ_2 be two probability measures on $\mathbb{R}^D$, having Fourier transforms χ_1, χ_2 respectively, and log-Fourier transforms Φ_1, Φ_2 respectively.
Let $\alpha \in [0, 2)$, and suppose $C_0 > 0$ so that, for all $\vec{\xi} \in \mathbb{R}^D$,

$$\mathbf{re}\left[\Phi_1(\vec{\xi})\right], \; \mathbf{re}\left[\Phi_2(\vec{\xi})\right] < -C_0|\vec{\xi}|^\alpha.$$

1. *If $\alpha \neq 1$, and there is $C_1 > 0$ such that $\left|\Phi_1(\vec{\xi}) - \Phi_2(\vec{\xi})\right| < C_1 \cdot |\vec{\xi}|^\alpha$,*
   then: $\|\chi_1 - \chi_2\|_\infty < \left(\frac{C_1}{e \cdot C_0}\right)$.

2. *If $\alpha = 1$, and there are $C_1, C_2 > 0$ such that*

$$\left|\Phi_1(\vec{\xi}) - \Phi_2(\vec{\xi})\right| < C_1 \cdot |\vec{\xi}| + C_2 \cdot |\vec{\xi}| \log |\vec{\xi}|,$$

$$\textit{then:} \quad \|\chi_1 - \chi_2\|_\infty < \left(\frac{C_1 - C_2 \log(C_0)}{e \cdot C_0} + \frac{C_2}{C_0}\frac{4}{e^2}\right)$$

Proof:

Claim 1: *For any $x, z \in \mathbb{C}$, with $\mathbf{re}[x] < \mathbf{re}[z]$,*
$|\exp(x) - \exp(z)| < \exp(\mathbf{re}[z]) \cdot |x - z|$.

Proof: By the Intermediate Value Theorem, we can find a point y on the line segment between x and z, so that $(\exp(x) - \exp(z)) = \exp'(y) \cdot (x - z)$. But since $\mathbf{re}\,[x] \leq \mathbf{re}\,[y] \leq \mathbf{re}\,[z]$, therefore $|\exp'(y)| = |\exp(y)| = \exp(\mathbf{re}\,[y]) < \exp(\mathbf{re}\,[z])$. Thus, $|\exp(x) - \exp(z)| = |\exp'(y)| \cdot |x - z| < \exp(\mathbf{re}\,[z]) \cdot |x - z|$. .....□ [Claim 1]

Proof of Part 1: For any $\vec{\xi} \in \mathbb{R}^D$,

$$\begin{aligned}
\left|\chi_1(\vec{\xi}) - \chi_2(\vec{\xi})\right| &= \left|\exp(\Phi_1(\vec{\xi})) - \exp(\Phi_2(\vec{\xi}))\right| \\
&\leq_{(1)} \exp\left(\max\left\{\mathbf{re}\left[\Phi_1(\vec{\xi})\right],\ \mathbf{re}\left[\Phi_2(\vec{\xi})\right]\right\}\right) \cdot \left|\Phi_1(\vec{\xi}) - \Phi_2(\vec{\xi})\right| \\
&\leq_{(2)} \exp\left(-C_0 \left|\vec{\xi}\right|^\alpha\right) \cdot C_1 \cdot \left|\vec{\xi}\right|^\alpha \\
&=_{(3)} \frac{C_1}{C_0} \exp(-z) \cdot z \\
&\leq \frac{C_1}{C_0} \max_{z \geq 0} \{\exp(-z) \cdot z\} \\
&\leq_{(4)} \frac{C_1}{C_0} \frac{1}{e}.
\end{aligned}$$

(1) By **Claim 1**.

(2) By the hypotheses.

(3) Here, we define $z = C_0 \left|\vec{\xi}\right|^\alpha$.

(4) Consider the first term. The function $\exp(-z) \cdot z$ is maximized when

$$0 = \frac{d}{dz} z \exp(-z) = (1 - z) \exp(-z), \qquad (*)$$

which occurs when $z = 1$. (Expression $(*)$ is positive when $z < 1$ and negative when $z > 1$, so we know this is a maximum.) At $z = 1$, the expression $\exp(-z) \cdot z$ has the value $1/e$.

Proof of Part 2: For any $\vec{\xi} \in \mathbb{R}^D$,

$$\begin{aligned}
\left|\chi_1(\vec{\xi}) - \chi_2(\vec{\xi})\right| &= \left|\exp(\Phi_1(\vec{\xi})) - \exp(\Phi_2(\vec{\xi}))\right| \\
&\leq_{(1)} \exp\left(\max\left\{\mathbf{re}\left[\Phi_1(\vec{\xi})\right],\ \mathbf{re}\left[\Phi_2(\vec{\xi})\right]\right\}\right) \cdot \left|\Phi_1(\vec{\xi}) - \Phi_2(\vec{\xi})\right| \\
&\leq_{(2)} \exp\left(-C_0 \left|\vec{\xi}\right|\right) \cdot \left(C_1 \cdot \left|\vec{\xi}\right| + C_2 \cdot \left|\vec{\xi}\right| \log \left|\vec{\xi}\right|\right) \\
&=_{(3)} \frac{C_1}{C_0} \exp(-z) \cdot z \quad + \quad \frac{C_2}{C_0} \exp(-z) \cdot z[\log(z) - \log(C_0)]
\end{aligned}$$

$$\begin{aligned} &= \left[\frac{C_1}{C_0} - \frac{C_2}{C_0}\log(C_0)\right] \cdot \exp(-z) \cdot z \\ &\qquad + \frac{C_2}{C_0}\exp(-z) \cdot z\log(z) \\ &\leq_{(4)} \left(\frac{C_1 - C_2\log(C_0)}{C_0 \cdot e}\right) + \frac{C_2}{C_0}\left|\exp(-z) \cdot z\log(z)\right| \\ &\leq_{(5)} \left(\frac{C_1 - C_2\log(C_0)}{C_0 \cdot e}\right) + \frac{C_2}{C_0}\frac{4}{e^2}. \end{aligned}$$

(1) By **Claim 1**.

(2) By the hypotheses.

(3) Here, we define $z = C_0\left|\vec{\xi}\right|$.

(4) As in the proof of **Part 1**.

(5) Let $f(z) = |\exp(-z)z\log(z)|$. Now, if $z \in [0,1]$, then $|z\log(z)| < \frac{1}{e} < \frac{4}{e^2}$. On the other hand, if $z > 1$, then $|z\log(z)| < z^2$, so $f(z) < g(z) = \exp(-z) \cdot z^2$. The function $g(z)$ takes its extremal values when $0 = g'(z) = \exp(-z)(2z - z^2)$, which happens when $z = 0$ or 2. $z = 0$ is a minimum, while $z = 2$ is a maximum; the value of $g(2)$ is $\frac{4}{e^2}$. ___□

Now we want to find sufficient conditions for $\mathbf{L}^1$ convergence of the Fourier transforms. This will require a little more machinery.

Definition 5: $\eta_{\vec{\xi}}$

For any $\vec{\xi} \in \mathbb{R}^D$, let $\eta_{\vec{\xi}} : \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$ be defined:

$$\eta_{\vec{\xi}}(\mathbf{s}) = \left|\left\langle \vec{\xi}, \mathbf{s} \right\rangle\right|^{\alpha} + \mathcal{B}_\alpha \left\langle \vec{\xi}, \mathbf{s} \right\rangle^{\langle\alpha\rangle} \mathbf{i}.$$

Definition 6: $\mathcal{K}(\rho)$

Let ρ be a probability measure on $\mathbb{R}^D$ with Fourier transform χ.

$$\text{Define } \mathcal{K}(\rho) = \min_{\theta \in \mathbb{S}^{D-1}} \left|\log|\chi(\theta)|\right| \tag{3.1}$$

Remark 7: Since ρ is a probability measure, $|\chi(\vec{\xi})| \leq 1$ for all $\vec{\xi} \in \mathbb{R}^D$, with equality at $\vec{\xi} = 0$. If $|\chi(\vec{\xi})| = 1$ for some $\vec{\xi} \neq 0$, this means that the probability measure ρ is totally "flat" in the dimension spanned by $\vec{\xi}$; in other words, the support of ρ lies entirely in some affine subspace of $\mathbb{R}^D$ perpendicular to $\vec{\xi}$.

Hence, the following conditions are equivalent:

1. $\mathcal{K}(\rho) > 0$.
2. $|\chi(\theta)| < 1$, for every $\theta \in \mathbb{S}^{D-1}$.
3. **supp** $[\rho]$ is a large enough subset of $\mathbb{R}^D$ that its affine closure is all of $\mathbb{R}^D$.

If ρ is an α-stable probability measure, with log-characteristic function Φ and spectral measure Γ, then notice that, for any $\vec{\xi} \in \mathbb{R}^D$, $\log \left|\chi(\vec{\xi})\right| = \mathbf{re}\left[\Phi(\vec{\xi})\right]$. Thus,

$$\mathcal{K}(\rho) \;=\; \min_{\theta \in \mathbb{S}^{D-1}} |\mathbf{re}\left[\Phi(\theta)\right]| \tag{3.2}$$

By thinking of a random vector with distribution ρ as a stochastic integral of the α-stable noise on $\mathbb{S}^{D-1}$ with intensity Γ (see Example 98 on page 188) , it is easy to see that the linear span of **supp** $[\rho]$ is the same as the linear span of **supp** $[\Gamma]$. Thus, the following conditions are equivalent:

1. $\mathcal{K}(\rho) > 0$.
2. $\mathbf{re}\left[\Phi(\theta)\right] < 0$, for every $\theta \in \mathbb{S}^{D-1}$.
3. **supp** $[\Gamma]$ is a large enough subset of $\mathbb{S}^{D-1}$ that its linear span is all of $\mathbb{R}^D$.

Proposition 8: *(A Sufficient condition for* $\mathbf{L}^1$ *Fourier Convergence)*

Let $\alpha \in [0, 2)$*, and let* ρ_1, ρ_2 *be two* α*-stable probability distributions on* $\mathbb{R}^D$*, having Fourier transforms* χ_1, χ_2*, log-Fourier transforms* Φ_1, Φ_2*, and spectral measures* Γ_1, Γ_2*, respectively. Let* $\mathcal{K} = \min\{\mathcal{K}(\rho_1), \mathcal{K}(\rho_2)\}$.

1. *If* $\alpha \neq 1$*, then* $\|\chi_1 - \chi_2\|_1 \;<\; \dfrac{\mathcal{C}_\alpha}{\mathcal{K}^{1+D/\alpha}} \cdot \left\|(\Phi_1 - \Phi_2)_{|\mathbb{S}^{D-1}}\right\|_1$.
2. *If* $\alpha = 1$*, then*

$$\|\chi_1 - \chi_2\|_1 \;<\; \frac{\mathcal{C}_1}{\mathcal{K}^{D+1}} \cdot \left\|(\Phi_1 - \Phi_2)_{|\mathbb{S}^{D-1}}\right\|_1 + |\mathcal{B}_1| \cdot \mathcal{A} \cdot \mathcal{F}(\mathcal{K}) \cdot \|\Gamma_1 - \Gamma_2\|_{var} .$$

$$\begin{aligned}
\textit{where we define} \quad \mathcal{A} &= \mathbf{Area}\left[\mathbb{S}^{D-1}\right] \\
\mathcal{C}_\alpha &= \int_0^\infty \exp\left(-s^\alpha\right) \cdot s^{D+\alpha-1}\, ds, \\
(\textit{thus}, \ \mathcal{C}_1 &= \int_0^\infty \exp\left(-s\right) \cdot s^{D}\, ds), \\
\textit{and} \quad \mathcal{F}(\mathcal{K}) &= \frac{\mathcal{E}}{\mathcal{K}^{D+1}} + \mathcal{C}_1 \cdot \max\left\{\frac{1}{e\cdot(D+1)}, \ \frac{|\log(\mathcal{K})|}{\mathcal{K}^{D+1}}\right\}. \\
\textit{where} \quad \mathcal{E} &= \int_0^\infty \exp\left(-s\right) \cdot s^{D} \cdot |\log(s)|\ ds.
\end{aligned}$$

Proof:

Claim 1: *Let* $\vec{\xi} \in \mathbb{R}^D$, *and let* $\theta = \frac{\vec{\xi}}{|\vec{\xi}|}$.

If $\alpha \neq 1$, *then* $\left|\chi_1(\vec{\xi}) - \chi_2(\vec{\xi})\right| \ \leq \ \exp\left(\mathbf{re}\left[\Phi_{max}(\theta)\right]\left|\vec{\xi}\right|^\alpha\right)\cdot\left|\vec{\xi}\right|^\alpha \cdot |\Phi_1(\theta) - \Phi_2(\theta)|.$

If $\alpha = 1$, *then* $\left|\chi_1(\vec{\xi}) - \chi_2(\vec{\xi})\right| \ \leq$

$$\exp\left(\mathbf{re}\left[\Phi_{max}(\theta)\right]\left|\vec{\xi}\right|\right)\cdot\left|\vec{\xi}\right|\cdot\Big(\ |\Phi_1(\theta) - \Phi_2(\theta)| \ + \ \left|\mathcal{B}_1 \cdot \log|\vec{\xi}|\right| \cdot |c_{\Gamma_1} - c_{\Gamma_2}|\Big),$$

where $\Phi_{max}(\theta)$ *is whichever of* $\{\Phi_1(\theta), \Phi_1(\theta)\}$ *has the larger real part, and where, for* $k = 1, 2$,

$$c_{\Gamma_k} = \int_{\mathbb{S}^{D-1}} \mathbf{u}\ d\Gamma_k[\mathbf{u}]$$

is the **centroid** *of* Γ_k.

Proof:

Claim 1.1: $\mathbf{re}\left[\Phi_k(\vec{\xi})\right] \ = \ \left|\vec{\xi}\right|^\alpha \mathbf{re}\left[\Phi_k(\theta)\right].$

Proof: First, note that $\mathbf{re}\left[\eta_{\vec{\xi}}(\mathbf{s})\right] \ = \ \left|\left\langle\vec{\xi}, \mathbf{s}\right\rangle\right|^\alpha \ = \ \left|\vec{\xi}\right|^\alpha |\langle\theta, \mathbf{s}\rangle|^\alpha \ =$ $\left|\vec{\xi}\right|^\alpha \mathbf{re}\left[\eta_\theta(\mathbf{s})\right]$. Thus, for $k = 1, 2$, $\mathbf{re}\left[\Phi_k(\vec{\xi})\right] \ = \ \int_{\mathbb{S}^{D-1}} \mathbf{re}\left[\eta_{\vec{\xi}}(\mathbf{s})\right] d\Gamma[\mathbf{s}] \ =$ $\left|\vec{\xi}\right|^\alpha \int_{\mathbb{S}^{D-1}} \mathbf{re}\left[\eta_\theta(\mathbf{s})\right]\ d\Gamma[\mathbf{s}] \ = \ \left|\vec{\xi}\right|^\alpha \mathbf{re}\left[\Phi_k(\theta)\right]$. ..... □ [Claim 1.1]

Claim 1.2: *If* $\alpha \neq 1$, *then* $\mathbf{im}\left[\Phi_k(\vec{\xi})\right] \ = \ \left|\vec{\xi}\right|^\alpha \mathbf{im}\left[\Phi_k(\theta)\right]$.

If $\alpha = 1$, *then* $\mathbf{im}\left[\Phi_k(\vec{\xi})\right] \ = \ \left|\vec{\xi}\right|\Big(\mathbf{im}\left[\Phi_k(\theta)\right] \ + \ \mathcal{B}_1 \log|\vec{\xi}|\,\langle\theta, c_{\Gamma_k}\rangle\Big)$.

Proof: If $\alpha \neq 1$, then the proof is just like **Claim 1.1.** $\mathbf{im}\left[\eta_{\vec{\xi}}(\mathbf{s})\right] = \mathcal{B}_\alpha \left\langle \vec{\xi}, \mathbf{s} \right\rangle^{\langle\alpha\rangle} = \mathcal{B}_\alpha \left|\vec{\xi}\right|^\alpha \langle\theta, \mathbf{s}\rangle^{\langle\alpha\rangle} = \left|\vec{\xi}\right|^\alpha \mathbf{im}\left[\eta_\theta(\mathbf{s})\right]$. Thus, for $k = 1, 2$,

$$\begin{aligned}\mathbf{im}\left[\Phi_k(\vec{\xi})\right] &= \int_{\mathbb{S}^{D-1}} \mathbf{im}\left[\eta_{\vec{\xi}}(\mathbf{s})\right] d\Gamma[\mathbf{s}] = \left|\vec{\xi}\right|^\alpha \int_{\mathbb{S}^{D-1}} \mathbf{im}\left[\eta_\theta(\mathbf{s})\right] d\Gamma[\mathbf{s}] \\ &= \left|\vec{\xi}\right|^\alpha \mathbf{im}\left[\Phi_k(\theta)\right].\end{aligned}$$

If $\alpha = 1$, however,

$$\begin{aligned}\mathbf{im}\left[\eta_{\vec{\xi}}(\mathbf{s})\right] &= \mathcal{B}_1 \left\langle \vec{\xi}, \mathbf{s} \right\rangle \log \left|\left\langle \vec{\xi}, \mathbf{s} \right\rangle\right| \\ &= \mathcal{B}_1 \left|\vec{\xi}\right| \langle\theta, \mathbf{s}\rangle \left(\log|\langle\theta, \mathbf{s}\rangle| + \log\left|\vec{\xi}\right|\right) \\ &= \left|\vec{\xi}\right| \left(\mathbf{im}\left[\eta_\theta(\mathbf{s})\right] + \mathcal{B}_1 \langle\theta, \mathbf{s}\rangle \cdot \log\left|\vec{\xi}\right|\right)\end{aligned}$$

Hence,

$$\begin{aligned}\mathbf{im}\left[\Phi_k(\vec{\xi})\right] &= \int_{\mathbb{S}^{D-1}} \mathbf{im}\left[\eta_{\vec{\xi}}(\mathbf{s})\right] d\Gamma[\mathbf{s}] \\ &= \left|\vec{\xi}\right| \int_{\mathbb{S}^{D-1}} \left(\mathbf{im}\left[\eta_\theta(\mathbf{s})\right] + \mathcal{B}_1 \langle\theta, \mathbf{s}\rangle \cdot \log\left|\vec{\xi}\right|\right) d\Gamma[\mathbf{s}] \\ &= \left|\vec{\xi}\right| \left(\mathbf{im}\left[\Phi_k(\theta)\right] + \mathcal{B}_1 \log\left|\vec{\xi}\right| \cdot \int_{\mathbb{S}^{D-1}} \langle\theta, \mathbf{s}\rangle \, d\Gamma[\mathbf{s}]\right) \\ &= \left|\vec{\xi}\right| \left(\mathbf{im}\left[\Phi_k(\theta)\right] + \mathcal{B}_1 \log|\vec{\xi}| \, \langle\theta, c_{\Gamma_k}\rangle\right)\end{aligned}$$

..... □ [Claim 1.2]

Claim 1.3: *If $\alpha \neq 1$, then* $\left|\Phi_1(\vec{\xi}) - \Phi_2(\vec{\xi})\right| \leq \left|\vec{\xi}\right|^\alpha |\Phi_1(\theta) - \Phi_2(\theta)|$
If $\alpha = 1$, then

$$\left|\Phi_1(\vec{\xi}) - \Phi_2(\vec{\xi})\right| \leq \left|\vec{\xi}\right|^\alpha \cdot \left[|\Phi_1(\theta) - \Phi_2(\theta)| + \left|\mathcal{B}_1 \cdot \log|\vec{\xi}|\right| |c_{\Gamma_1} - c_{\Gamma_2}|\right].$$

Proof: By **Claim 1.1**,

$$\begin{aligned}&\Phi_1(\vec{\xi}) - \Phi_2(\vec{\xi}) \\ &= \mathbf{re}\left[\Phi_1(\vec{\xi})\right] - \mathbf{re}\left[\Phi_2(\vec{\xi})\right] + \mathbf{i}\left(\mathbf{im}\left[\Phi_1(\vec{\xi})\right] - \mathbf{im}\left[\Phi_2(\vec{\xi})\right]\right) \\ &= \left|\vec{\xi}\right|^\alpha \left(\mathbf{re}\left[\Phi_1(\theta)\right] - \mathbf{re}\left[\Phi_2(\theta)\right]\right) + \mathbf{i}\left(\mathbf{im}\left[\Phi_1(\vec{\xi})\right] - \mathbf{im}\left[\Phi_2(\vec{\xi})\right]\right).\end{aligned}$$

Now, when $\alpha \neq 1$, use **Claim 1.2** to conclude:

$\Phi_1(\vec{\xi}) - \Phi_2(\vec{\xi})$

$$= \left|\vec{\xi}\right|^{\alpha} \left[\left(\mathbf{re}\left[\Phi_1(\theta)\right] - \mathbf{re}\left[\Phi_2(\theta)\right]\right) + \mathbf{i}\left(\mathbf{im}\left[\Phi_1(\theta)\right] - \mathbf{im}\left[\Phi_2(\theta)\right]\right)\right]$$
$$= \left|\vec{\xi}\right|^{\alpha} \left(\Phi_1(\theta) - \Phi_2(\theta)\right)$$

whereas, when $\alpha = 1$, **Claim 1.2** tells us:

$$\Phi_1(\vec{\xi}) - \Phi_2(\vec{\xi})$$
$$= \left|\vec{\xi}\right| \cdot \left[\left(\mathbf{re}\left[\Phi_1(\theta)\right] - \mathbf{re}\left[\Phi_2(\theta)\right]\right) + \mathbf{i}\left(\mathbf{im}\left[\Phi_1(\theta)\right] - \mathbf{im}\left[\Phi_2(\theta)\right] + \mathcal{B}_1 \log|\vec{\xi}|\,\langle\theta,\ c_{\Gamma_1} - c_{\Gamma_2}\rangle\right)\right]$$
$$= \left|\vec{\xi}\right| \cdot \left[\left(\Phi_1(\theta) - \Phi_2(\theta)\right) + \mathbf{i}\mathcal{B}_1 \log|\vec{\xi}|\,\langle\theta,\ c_{\Gamma_1} - c_{\Gamma_2}\rangle\right]$$

..... □ [Claim 1.3]

Now, recall **Claim 1** of Proposition 4 on page 17: For any $x, y \in \mathbb{C}$, if $\mathbf{re}\left[x\right] < \mathbf{re}\left[y\right]$, then $|\exp(x) - \exp(y)| < \exp\left(\mathbf{re}\left[y\right]\right) \cdot |x - y|$. Hence, by **Claim 1.1**,

$$\left|\chi_1(\vec{\xi}) - \chi_2(\vec{\xi})\right| \leq \exp\left(\mathbf{re}\left[\Phi_{max}(\theta)\right]\left|\vec{\xi}\right|^{\alpha}\right) \cdot \left|\Phi_1(\vec{\xi}) - \Phi_2(\vec{\xi})\right|$$

Now use **Claim 1.3** to substitute the appropriate expression for $\left|\Phi_1(\vec{\xi}) - \Phi_2(\vec{\xi})\right|$, and Claim 1 is proved. ..... □ [Claim 1]

So, when $\alpha \neq 1$, we have:

$$\|\chi_1 - \chi_2\|_1$$
$$= \int_{\mathbb{R}^D} \left|\chi_1(\vec{\xi}) - \chi_2(\vec{\xi})\right| \, d\vec{\xi}$$
$$=_{(1)} \int_0^{\infty} \int_{\mathbb{S}^{D-1}} |\chi_1(r.\theta) - \chi_2(r.\theta)|\, r^{D-1}\, d\theta\, dr$$
$$\leq_{(2)} \int_0^{\infty} \int_{\mathbb{S}^{D-1}} \exp\left(\mathbf{re}\left[\Phi_{max}(\theta)\right] \cdot r^{\alpha}\right) \cdot |\Phi_1(\theta) - \Phi_2(\theta)| \cdot r^{\alpha} \cdot r^{D-1}\, d\theta\, dr$$
$$= \int_{\mathbb{S}^{D-1}} |\Phi_1(\theta) - \Phi_2(\theta)| \cdot \int_0^{\infty} \exp\left(\mathbf{re}\left[\Phi_{max}(\theta)\right] \cdot r^{\alpha}\right) \cdot r^{D+\alpha-1}\, dr\, d\theta \qquad (A)$$

(1) is just changing to spherical coordinates.

(2) follows from **Claim 1**.

But for any $\theta \in \mathbb{S}^{D-1}$,

$$\int_0^{\infty} \exp\left(\mathbf{re}\left[\Phi_{max}(\theta)\right] \cdot r^{\alpha}\right) \cdot r^{D+\alpha-1}\, dr$$

$$
\begin{aligned}
&= \frac{1}{|\mathbf{re}\,[\Phi_{max}(\theta)]|^{1+D/\alpha}} \int_0^\infty \exp(-s^\alpha) \cdot s^{D+\alpha-1}\, ds \\
&= \frac{1}{|\mathbf{re}\,[\Phi_{max}(\theta)]|^{1+D/\alpha}} \cdot \mathcal{C}_\alpha, \qquad\qquad (B)
\end{aligned}
$$

where we make the substitution:

$$
\begin{aligned}
s &= |\mathbf{re}\,[\Phi_{max}(\theta)]|^{1/\alpha}\, r \\
\text{thus,}\quad dr &= |\mathbf{re}\,[\Phi_{max}(\theta)]|^{-1/\alpha}\, ds
\end{aligned}
$$

Thus, substituting (B) into (A), we get:

$$
\|\chi_1 - \chi_2\|_1 \leq \int_{\mathbb{S}^{D-1}} \frac{|\Phi_1(\theta) - \Phi_2(\theta)|}{|\mathbf{re}\,[\Phi_{max}(\theta)]|^{1+D/\alpha}} \cdot \mathcal{C}_\alpha\, d\theta
$$

where $\Phi_{max}(\theta) \in \{\Phi_1(\theta), \Phi_2(\theta)\}$ is chosen at each $\theta \in \mathbb{S}^{D-1}$ so as to maximize $\mathbf{re}\,[\Phi_{max}]$ —or, equivalently, since $\mathbf{re}\,[\Phi_k] < 0$ for $k = 1, 2$, we can say that Φ_{max} is chosen to *minimize* $\left|\mathbf{re}\,[\Phi_{max}]\right|$. But by equation (3.2) , for all $\theta \in \mathbb{S}^{D-1}$,

$$
\begin{aligned}
\left|\mathbf{re}\,[\Phi_{max}](\theta)\right| &\geq \mathcal{K}, \\
\text{hence,}\quad \|\chi_1 - \chi_2\|_1 &\leq \mathcal{C}_\alpha \cdot \int_{\mathbb{S}^{D-1}} \frac{|\Phi_1(\theta) - \Phi_2(\theta)|}{\mathcal{K}^{1+D/\alpha}}\, d\theta \\
&= \frac{\mathcal{C}_\alpha}{\mathcal{K}^{1+D/\alpha}} \cdot \left\| (\Phi_1 - \Phi_2)_{|\mathbb{S}^{D-1}} \right\|_1
\end{aligned}
$$

This proves the theorem when $\alpha \neq 1$.

When $\alpha = 1$, use **Claim 1** to get:

$$
\begin{aligned}
&\|\chi_1 - \chi_2\|_1 \\
&\quad = \int_{\mathbb{R}^D} \left|\chi_1(\vec{\xi}) - \chi_2(\vec{\xi})\right|\, d\vec{\xi} \\
&\quad =_{(1)} \int_0^\infty \int_{\mathbb{S}^{D-1}} |\chi_1(r.\theta) - \chi_2(r.\theta)|\, r^{D-1}\, d\theta\, dr \\
&\quad \leq_{(2)} \int_0^\infty \int_{\mathbb{S}^{D-1}} \exp(\mathbf{re}\,[\Phi_{max}(\theta)]r) \cdot r \cdot \\
&\qquad\qquad \left[\, |\Phi_1(\theta) - \Phi_2(\theta)| \;+\; |\mathcal{B}_1 \cdot \log(r)|\, |c_{\Gamma_1} - c_{\Gamma_2}| \right] d\theta r^{D-1} dr \\
&\quad = \int_{\mathbb{S}^{D-1}} |\Phi_1(\theta) - \Phi_2(\theta)| \left(\int_0^\infty \exp(\mathbf{re}\,[\Phi_{max}(\theta)]r) \cdot r^D\, dr \right) d\theta
\end{aligned}
$$

$$
\begin{aligned}
&+ \int_{\mathbb{S}^{D-1}} |\mathcal{B}_1| \cdot |c_{\Gamma_1} - c_{\Gamma_2}| \cdot \left(\int_0^\infty \exp\left(\mathbf{re}\left[\Phi_{max}(\theta)\right] r\right) \cdot r^D \cdot |\log(r)| \ dr \right) \ d\theta \\
\leq_{(3)} \quad & \frac{\mathcal{C}_1}{\mathcal{K}^{D+1}} \cdot \left\| (\Phi_1 - \Phi_2)_{|\mathbb{S}^{D-1}} \right\|_1 \\
&+ |\mathcal{B}_1| \cdot |c_{\Gamma_1} - c_{\Gamma_2}| \cdot \int_{\mathbb{S}^{D-1}} \left(\int_0^\infty \exp\left(\mathbf{re}\left[\Phi_{max}(\theta)\right] r\right) \cdot r^D \cdot |\log(r)| \ dr \right) \ d\theta.
\end{aligned}
$$

(1) is again a change to spherical coordinates.

(2) By **Claim 1**.

(3) follows exactly as in the $\alpha \neq 1$ case.

$$
\begin{aligned}
\text{Now, for fixed } \theta, \text{ let } \quad P &= \left|\mathbf{re}\left[\Phi_{max}(\theta)\right]\right|, \\
\text{and } \quad r &= P^{-1} s. \\
\text{Thus, } \quad dr &= P^{-1} ds, \\
\text{and } \quad |\log(r)| &= \left|\log(s) - \log(P)\right|.
\end{aligned}
$$

$$
\begin{aligned}
\text{Thus, } \quad & \int_0^\infty \exp\left(-Pr\right) \cdot r^D \cdot |\log(r)| \ dr \\
= \quad & \frac{1}{P^{D+1}} \int_0^\infty \exp\left(-s\right) \cdot s^D \cdot \Big| \log(s) - \log(P) \Big| \ ds \\
\leq \quad & \frac{1}{P^{D+1}} \int_0^\infty \exp\left(-s\right) \cdot s^D \cdot |\log(s)| \ ds \ + \ \frac{|\log(P)|}{P^{D+1}} \int_0^\infty \exp\left(-s\right) \cdot s^D \ ds \\
= \quad & \frac{\mathcal{E}}{P^{D+1}} \ + \ \frac{|\log(P)|}{P^{D+1}} \mathcal{C}_1 \\
\leq_{(1)} \quad & \frac{\mathcal{E}}{\mathcal{K}^{D+1}} \ + \ \frac{|\log(P)|}{P^{D+1}} \mathcal{C}_1 \\
\leq_{(2)} \quad & \frac{\mathcal{E}}{\mathcal{K}^{D-1}} \ + \ \mathcal{C}_1 \cdot \max\left\{ \frac{1}{e \cdot (D+1)}, \ \frac{|\log(\mathcal{K})|}{\mathcal{K}^{D+1}} \right\} \\
= \quad & \mathcal{F}(\mathcal{K}).
\end{aligned}
$$

(1) Recall that $P = \Big|\mathbf{re}\left[\Phi_{max}(\theta)\right]\Big|$. Thus, by equation (3.2), $\mathcal{K} < P$ for any θ.

(2) Let $f(x) = \frac{\log(x)}{x^{D+1}}$. Then $f'(x) = \frac{1-(D+1)\log(x)}{x^{D+2}} = 0$ when $x = e^{1/(D+1)}$, and f takes a maximum at this point. Thus, for any $P > 1$, we have $f(P) > 0$, so that $|f(P)| = f(P) \leq f\left(e^{1/(D+1)}\right) = \frac{1}{e(D+1)}$. On

the other hand, if $P < 1$, but $\mathcal{K} \leq P$, then $0 > f(P) \geq f(\mathcal{K})$, therefore $|f(P)| \leq |f(\mathcal{K})|$.

Hence,

$$\begin{aligned} & |\mathcal{B}_1| \cdot |c_{\Gamma_1} - c_{\Gamma_2}| \cdot \int_{\mathbb{S}^{D-1}} \left(\int_0^\infty \exp\left(|\mathbf{re}\left[\Phi_{max}(\theta)\right]| \, r \right) \cdot r^D \, |\log(r)| \; dr \right) \; d\theta \\ \leq \; & |\mathcal{B}_1| \cdot |c_{\Gamma_1} - c_{\Gamma_2}| \cdot \int_{\mathbb{S}^{D-1}} \mathcal{F}(\mathcal{K}) \; ds \\ = \; & \mathcal{F}(\mathcal{K}) \cdot |\mathcal{B}_1| \, |c_{\Gamma_1} - c_{\Gamma_2}| \cdot \mathbf{Area}\left[\mathbb{S}^{D-1}\right] \\ \leq \; & \mathcal{F}(\mathcal{K}) \cdot |\mathcal{B}_1| \cdot \|\Gamma_1 - \Gamma_2\|_{var} \cdot \mathcal{A}. \end{aligned}$$

□

3.2.3 Convergence of the Log-Fourier Transform

We now have explicit bounds on the $\mathbf{L}^p$ distance between the Fourier Transforms χ_1 and χ_2, in terms of the distance between log-Fourier transforms Φ_1 and Φ_2. Now we seek bounds on this distance, in turn, in terms of the distance between spectral measures Γ_1 and Γ_2.

Lemma 9: *Let $\alpha \in [0, 2)$, and, for any $\vec{\xi} \in \mathbb{R}^D$, let $\eta_{\vec{\xi}}$ be defined as before. Then*

$$\left\| \eta_{\vec{\xi}} \right\|_\infty \; < \; \begin{cases} (1 + |\mathcal{B}_\alpha|) \left| \vec{\xi} \right|^\alpha & if \quad \alpha \neq 1 \\ \left| \vec{\xi} \right| \left(1 + \left| \mathcal{B}_1 \log \left| \vec{\xi} \right| \right| \right) & if \quad \alpha = 1 \end{cases}$$

$$\textit{For any } p \in [1, \infty], \quad \left\| \eta_{\vec{\xi}} \right\|_p \; < \; \begin{cases} (1 + |\mathcal{B}_\alpha|) \left| \vec{\xi} \right|^\alpha \mathcal{A}^{1/p} & if \quad \alpha \neq 1 \\ \left| \vec{\xi} \right| \left(1 + \left| \mathcal{B}_1 \log \left| \vec{\xi} \right| \right| \right) \cdot \mathcal{A}^{1/p} & if \quad \alpha = 1 \end{cases}$$

where $\mathcal{A}$ is the surface area of the $(D-1)$-sphere.

Proof: The bounds in $\mathbf{L}^p$ follow immediately from the $\mathbf{L}^\infty$ bounds, so we will prove these.

Case ($\alpha \neq 1$): For any $\mathbf{s} \in \mathbb{S}^{D-1}$, $\left| \eta_{\vec{\xi}}(\mathbf{s}) \right| \leq \left| \left\langle \vec{\xi}, \mathbf{s} \right\rangle \right|^\alpha + |\mathcal{B}_\alpha| \left| \left\langle \vec{\xi}, \mathbf{s} \right\rangle^{\langle \alpha \rangle} \right| \leq \left| \vec{\xi} \right|^\alpha + |\mathcal{B}_\alpha| \left| \vec{\xi} \right|^\alpha = (1 + |\mathcal{B}_\alpha|) \left| \vec{\xi} \right|^\alpha$.

Case ($\alpha = 1$): For any $\mathbf{s} \in \mathbb{S}^{D-1}$, $\left|\eta_{\vec{\xi}}(\mathbf{s})\right| \leq \left|\left\langle \vec{\xi}, \mathbf{s} \right\rangle\right| + |\mathcal{B}_1| \left|\left\langle \vec{\xi}, \mathbf{s} \right\rangle^{\langle 1 \rangle}\right| \leq \left|\vec{\xi}\right| + |\mathcal{B}_\alpha| \left|\vec{\xi}\right| \log \left|\vec{\xi}\right|$. ——————□

Lemma 10: *Let $\alpha \in [0, 2)$, and let $\chi, \Phi : \mathbb{R}^D \longrightarrow \mathbb{C}$ be the Fourier transform and log-Fourier transform, respectively, of an α-stable probability distribution ρ, with spectral measure Γ, and let $\mathcal{K} = \mathcal{K}(\rho)$.*

$$\textit{For every } \vec{\xi} \in \mathbb{R}^D, \quad \mathbf{re}\left[\Phi(\vec{\xi})\right] < -\mathcal{K} \cdot |\vec{\xi}|^\alpha$$

Proof: Let $\eta_{\vec{\xi}}$ be as previously defined. If $\vec{\xi} \in \mathbb{R}^D$, let $\theta = \frac{\vec{\xi}}{|\vec{\xi}|}$, and note that, for any $\mathbf{s} \in \mathbb{S}^{D-1}$, $\mathbf{re}\left[\eta_{\vec{\xi}}(\mathbf{s})\right] = \left|\vec{\xi}\right|^\alpha \mathbf{re}\left[\eta_\theta(\mathbf{s})\right]$.

$$\begin{aligned} \text{Thus} \quad \mathbf{re}\left[\Phi(\vec{\xi})\right] &=_{(1)} \left|\vec{\xi}\right|^\alpha \mathbf{re}\left[\Phi(\theta)\right], \\ &= \left|\vec{\xi}\right|^\alpha \log|\chi(\theta)| \\ &\leq \left|\vec{\xi}\right|^\alpha \max_{\mathbf{s} \in \mathbb{S}^{D-1}} \log|\chi(\mathbf{s})| \\ &= \left|\vec{\xi}\right|^\alpha (-\mathcal{K}). \end{aligned}$$

(1) By Claim 1.1 of Proposition 8 on page 20. ——————□

One way to make the Fourier transforms of ρ_1 and ρ_2 close is to make Γ_1 and Γ_2 close relative to some norm. Which norm to use depends on the structure of Γ_1 and Γ_2.

Definition 11: *Nice Measure, $\|\bullet\|_{[p]}$*

Let Γ be a measure on $\mathbb{S}^{D-1}$. We say Γ is a **nice measure** *if Γ is a sum of a "purely atomic" component, and a component absolutely continuous with respect to Lebesgue measure. Formally:*

$$\Gamma = g \cdot \mathcal{L}^{bsg} + \sum_{\mathbf{a} \in \mathbb{A}} \gamma_{\mathbf{a}} \delta_{\mathbf{a}},$$

where $g \in \mathbf{L}^1(\mathbb{S}^{D-1}, \mathcal{L}^{bsg})$, where $\mathbb{A} \subset \mathbb{S}^{D-1}$ is some finite set, and where $\{\gamma_{\mathbf{a}} \; ; \; \mathbf{a} \in \mathbb{A}\}$ are real numbers.

If Γ is a **nice measure**, *then for any $q \in [1, \infty]$, we can define the* $[q]$**-norm** *of Γ:*

$$\|\Gamma\|_{[q]} \quad = \quad \mathcal{A}^{1/p} \cdot \|g\|_q \; + \; \sum_{\mathbf{a} \in \mathbb{A}} |\gamma_{\mathbf{a}}| \, .$$

where $p \in [1, \infty]$ is chosen so that $\frac{1}{p} + \frac{1}{q} = 1$, and $\mathcal{A} =$ **Area** $\left[\mathbb{S}^{D-1}\right]$ *(if $p = \infty$, we formally define $\mathcal{A}^{1/\infty} = 1$).*

Remark 12:

- Notice that $\|\Gamma\|_{[1]} \; = \; \|\Gamma\|_{var}$.
- If Γ_1 and Γ_2 are both **nice** measures, then so is $\Gamma_1 - \Gamma_2$. Indeed, suppose that, for $k = 1, 2$, $\Gamma_k = g_k \cdot \mathcal{L}^{bsg} \; + \; \sum_{\mathbf{a} \in \mathbb{A}_k} \gamma_{k,\mathbf{a}} \delta_{\mathbf{a}}$, where $g_1, g_2 \in \mathbf{L}^1(\mathbb{S}^{D-1}, \mathcal{L}^{bsg})$, where $\mathbb{A}_k \subset \mathbb{S}^{D-1}$ are finite sets, and where $\{\gamma_{k,\mathbf{a}} \; ; \; \mathbf{a} \in \mathbb{A}_k\}$ are real values. If we define $\mathbb{A} = \mathbb{A}_1 \cup \mathbb{A}_2$, then we can write $\Gamma_k = g_k \cdot \mathcal{L}^{bsg} \; + \; \sum_{\mathbf{a} \in \mathbb{A}} \gamma_{k,\mathbf{a}} \delta_{\mathbf{a}}$, where we simply define $\gamma_{k,\mathbf{a}} = 0$ for all $\mathbf{a} \in \mathbb{A} \setminus \mathbb{A}_k$. Then

$$\Gamma_1 - \Gamma_2 = (g_1 - g_2) \cdot \mathcal{L}^{bsg} \; + \; \sum_{\mathbf{a} \in \mathbb{A}} (\gamma_{1,\mathbf{a}} - \gamma_{2,\mathbf{a}}) \, \delta_{\mathbf{a}}.$$

- In this case, for any $q \in [1, \infty]$,

$$\|\Gamma_1 - \Gamma_2\|_{[q]} = \mathcal{A}^{1/p} \cdot \|g_1 - g_2\|_q \; + \; \sum_{\mathbf{a} \in \mathbb{A}} |\gamma_{1,\mathbf{a}} - \gamma_{2,\mathbf{a}}|$$

Proposition 13: *(Convergence of Spectral Measure: Consequences)*
Let ρ_1 and ρ_2 be α-stable probability measures on $\mathbb{R}^D$, with spectral measures Γ_1 and Γ_2, and characteristic functions χ_1 and χ_2, respectively.

$$\begin{aligned} \textit{Let } \mathcal{A} &= \textbf{Area}\left[\mathbb{S}^{D-1}\right], \\ \mathcal{D}_\alpha &= \begin{cases} 1 + |\mathcal{B}_\alpha| & if \quad \alpha \neq 1 \\ 1 & if \quad \alpha = 1 \end{cases} \\ \textit{and } \mathcal{K} &= \min\left\{\mathcal{K}(\rho_1), \; \mathcal{K}(\rho_2)\right\}, \end{aligned}$$

and let $\mathcal{C}_\alpha$, $\mathcal{E}$, and $\mathcal{F}(\mathcal{K})$ be defined the same as in Proposition 8 on page 20.

1. *First suppose* Γ_1 *and* Γ_2 *are arbitrary measures. Then:*

   (a) *If* $\alpha \neq 1$, *then* $\|\chi_1 - \chi_2\|_\infty < \frac{(1+|\mathcal{B}_\alpha|)}{\mathcal{K}e} \cdot \|\Gamma_1 - \Gamma_2\|_{var}$.
   If $\alpha = 1$, *then*

   $$\|\chi_1 - \chi_2\|_\infty < \left(\frac{1-|\mathcal{B}_1|\log(\mathcal{K})}{e\cdot\mathcal{K}} + \frac{|\mathcal{B}_1|}{\mathcal{K}}\frac{4}{e^2}\right)\cdot\|\Gamma_1-\Gamma_2\|_{var}\,.$$

   (b) *If* $\alpha \neq 1$, *then* $\|\chi_1 - \chi_2\|_1 < \frac{\mathcal{C}_\alpha\mathcal{AD}_\alpha}{\mathcal{K}^{1+D/\alpha}} \cdot \|\Gamma_1 - \Gamma_2\|_{var}$.
   If $\alpha = 1$, *then*

   $$\|\chi_1 - \chi_2\|_1 < \mathcal{A}\cdot\left[\frac{\mathcal{C}_1\mathcal{D}_1}{\mathcal{K}^{D+1}} + |\mathcal{B}_1|\cdot\mathcal{F}(\mathcal{K})\right]\cdot\|\Gamma_1-\Gamma_2\|_{var}\,.$$

2. *Next suppose that* Γ_1 *and* Γ_2 *are nice measures, with* $\Gamma_k = g_k\cdot\mathcal{L}^{leg} + \sum_{\mathbf{a}\in\mathbb{A}}\gamma_{k,\mathbf{a}}\delta_{\mathbf{a}}$, *where* $g_1, g_2 \in \mathbf{L}^1(\mathbb{S}^{D-1}, \mathcal{L}^{leg})$, *where* $\mathbb{A}\subset\mathbb{S}^{D-1}$ *is some finite set, and where* $\{\gamma_{k,\mathbf{a}}\,;\,\mathbf{a}\in\mathbb{A}\}$ *are nonnegative values (some possibly zero). Then for any* $p\in[1,\infty]$

   (a) *If* $\alpha \neq 1$, *then* $\|\chi_1 - \chi_2\|_\infty < \frac{(1+|\mathcal{B}_\alpha|)}{\mathcal{K}e} \cdot \|\Gamma_1 - \Gamma_2\|_{[p]}$.
   If $\alpha = 1$, *then*

   $$\|\chi_1 - \chi_2\|_\infty < \left(\frac{1-|\mathcal{B}_1|\log(\mathcal{K})}{e\cdot\mathcal{K}} + \frac{|\mathcal{B}_1|}{\mathcal{K}}\frac{4}{e^2}\right)\cdot\|\Gamma_1-\Gamma_2\|_{[p]}\,.$$

   (b) *If* $\alpha \neq 1$, *then* $\|\chi_1 - \chi_2\|_1 < \frac{\mathcal{C}_\alpha\mathcal{AD}_\alpha}{\mathcal{K}^{1+D/\alpha}} \cdot \|\Gamma_1 - \Gamma_2\|_{[p]}$.
   If $\alpha = 1$, *then*

   $$\|\chi_1 - \chi_2\|_1 < \mathcal{A}\cdot\left[\frac{\mathcal{C}_1\mathcal{D}_1}{\mathcal{K}^{D+1}}\|\Gamma_1-\Gamma_2\|_{[p]} + |\mathcal{B}_1|\cdot\mathcal{F}(\mathcal{K})\cdot\|\Gamma_1-\Gamma_2\|_{var}\right].$$

Proof: Let Φ_1 and Φ_2 be the log-Fourier transforms of ρ_1 and ρ_2.

Claim 1: *For any* $\vec{\xi}\in\mathbb{R}^D$,

$$\left|\Phi_1(\vec{\xi}) - \Phi_2(\vec{\xi})\right| < \|\Gamma_1-\Gamma_2\|_{var}\cdot\begin{cases}(1+|\mathcal{B}_\alpha|)\left|\vec{\xi}\right|^\alpha & if \quad \alpha\neq 1\\ \left|\vec{\xi}\right|\left(1+\left|\mathcal{B}_1\log\left|\vec{\xi}\right|\right|\right) & if \quad \alpha = 1\end{cases}$$

Proof:

$$\begin{aligned}
\left|\Phi_1(\vec{\xi}) - \Phi_2(\vec{\xi})\right| &= \left|\left\langle \eta_{\vec{\xi}}, \Gamma_1 \right\rangle - \left\langle \eta_{\vec{\xi}}, \Gamma_2 \right\rangle\right| \\
&= \left|\left\langle \eta_{\vec{\xi}}, \Gamma_1 - \Gamma_2 \right\rangle\right| \\
&\leq \left\|\eta_{\vec{\xi}}\right\|_\infty \|\Gamma_1 - \Gamma_2\|_{var} \\
&\leq_{(1)} \|\Gamma_1 - \Gamma_2\|_{var} \cdot \begin{cases} (1 + |\mathcal{B}_\alpha|) \left|\vec{\xi}\right|^\alpha & if \quad \alpha \neq 1 \\ \left|\vec{\xi}\right| \left(1 + \left|\mathcal{B}_1 \log \left|\vec{\xi}\right|\right|\right) & if \quad \alpha = 1 \end{cases}
\end{aligned}$$

Here, inequality (1) follows from Lemma 9 on page 26. □ `[Claim 1]`

Claim 2: *If Γ_1 and Γ_2 are nice, as described in* **Part 2** *of the theorem, then, for any $\vec{\xi} \in \mathbb{R}^D$,*

$$\left|\Phi_1(\vec{\xi}) - \Phi_2(\vec{\xi})\right| < \|\Gamma_1 - \Gamma_2\|_{[q]} \cdot \begin{cases} (1 + |\mathcal{B}_\alpha|) \left|\vec{\xi}\right|^\alpha & if \quad \alpha \neq 1 \\ \left|\vec{\xi}\right| \left(1 + \left|\mathcal{B}_1 \log \left|\vec{\xi}\right|\right|\right) & if \quad \alpha = 1 \end{cases}$$

Proof: When $\alpha \neq 1$,

$$\begin{aligned}
\left|\Phi_1(\vec{\xi}) - \Phi_2(\vec{\xi})\right| &= \left|\left\langle \eta_{\vec{\xi}}, \Gamma_1 - \Gamma_2 \right\rangle\right| \\
&\leq \left|\left\langle \eta_{\vec{\xi}}, g_1 - g_2 \right\rangle\right| + \sum_{\mathbf{a} \in \mathbb{A}} \left|\eta_{\vec{\xi}}(\mathbf{a})\right| \cdot |\gamma_{\mathbf{a};1} - \gamma_{\mathbf{a};2}| \\
&\leq_{(1)} \left\|\eta_{\vec{\xi}}\right\|_p \cdot \|g_1 - g_2\|_q + \sum_{\mathbf{a} \in \mathbb{A}} \left\|\eta_{\vec{\xi}}\right\|_\infty \cdot |\gamma_{\mathbf{a};1} - \gamma_{\mathbf{a};2}| \\
&\leq_{(2)} (1 + |\mathcal{B}_\alpha|)\, \mathcal{A}^{1/p} \left|\vec{\xi}\right|^\alpha \cdot \|g_1 - g_2\|_q \\
&\qquad + (1 + |\mathcal{B}_\alpha|) \left|\vec{\xi}\right|^\alpha \sum_{\mathbf{a} \in \mathbb{A}} |\gamma_{1,\mathbf{a}} - \gamma_{2,\mathbf{a}}| \cdot \\
&= (1 + |\mathcal{B}_\alpha|) \left[\mathcal{A}^{1/p} \cdot \|g_1 - g_2\|_q + \sum_{\mathbf{a} \in \mathbb{A}} |\gamma_{1,\mathbf{a}} - \gamma_{2,\mathbf{a}}|\right] \left|\vec{\xi}\right|^\alpha \\
&= (1 + |\mathcal{B}_\alpha|) \|\Gamma_1 - \Gamma_2\|_{[q]} \cdot \left|\vec{\xi}\right|^\alpha
\end{aligned}$$

- (1) is Hölder's inequality.
- (2) follows from Lemma 9 on page 26.

When $\alpha = 1$, the proof follows in the same way. ..... □ `[Claim 2]`

Proof of Part 1(a): When $\alpha \neq 1$, apply **Part 1** of Proposition 4 on page 17, with

$$\begin{aligned} \mathcal{C}_0 &= \mathcal{K}, && \text{(By Lemma 10 on page 27)} \\ \text{and } \mathcal{C}_1 &= \|\Gamma_1 - \Gamma_2\|_{var} \left(1 + |\mathcal{B}_\alpha|\right), && \text{(by Claim 1)} \\ \text{to conclude: } \|\chi_1 - \chi_2\|_\infty &< \frac{\mathcal{C}_1}{e \cdot \mathcal{C}_0} = \frac{(1 + |\mathcal{B}_\alpha|)}{e \cdot \mathcal{K}} \|\Gamma_1 - \Gamma_2\|_{var}. \end{aligned}$$

To prove **Part 1(a)** when $\alpha = 1$, apply **Part 2** of Proposition 4 on page 17, with

$$\begin{aligned} \mathcal{C}_0 &= \mathcal{K}, && \text{(By Lemma 10 on page 27)} \\ \mathcal{C}_1 &= \|\Gamma_1 - \Gamma_2\|_{var}, \\ \text{and } \mathcal{C}_2 &= |\mathcal{B}_1| \cdot \|\Gamma_1 - \Gamma_2\|_{var} && \text{(by Claim 1)} \\ \text{to conclude: } \|\chi_1 - \chi_2\|_\infty &< \frac{C_1 - C_2 \log(C_0)}{e \cdot C_0} + \frac{C_2}{C_0} \frac{4}{e^2} \\ &= \left(\frac{1 - |\mathcal{B}_1| \log(\mathcal{K})}{e \cdot \mathcal{K}} + \frac{|\mathcal{B}_1|}{\mathcal{K}} \frac{4}{e^2} \right) \cdot \|\Gamma_1 - \Gamma_2\|_{var}. \end{aligned}$$

Proof of Part 2(a): When $\alpha \neq 1$, again apply Proposition 4 on page 17, now with

$$\begin{aligned} \mathcal{C}_0 &= \mathcal{K} && \text{(By Lemma 10 on page 27)} \\ \text{and } \mathcal{C}_1 &= \|\Gamma_1 - \Gamma_2\|_{[q]} \left(1 + |\mathcal{B}_\alpha|\right) && \text{(by Claim 2)} \\ \text{to conclude: } \|\chi_1 - \chi_2\|_\infty &< \frac{\mathcal{C}_1}{e \cdot \mathcal{C}_0} = \frac{(1 + |\mathcal{B}_\alpha|)}{e \cdot \mathcal{K}} \|\Gamma_1 - \Gamma_2\|_{[p]}. \end{aligned}$$

To prove **Part 2(a)** when $\alpha = 1$, apply **Part 2** of Proposition 4 on page 17, with

$$\begin{aligned} \mathcal{C}_0 &= \mathcal{K}, && \text{By Lemma 10 on page 27} \\ \mathcal{C}_1 &= \|\Gamma_1 - \Gamma_2\|_{[p]}, \\ \text{and } \mathcal{C}_2 &= |\mathcal{B}_1| \cdot \|\Gamma_1 - \Gamma_2\|_{[p]}, && \text{(by Claim 2)} \\ \text{to conclude } \|\chi_1 - \chi_2\|_\infty &< \frac{C_1 - C_2 \log(C_0)}{e \cdot C_0} + \frac{C_2}{C_0} \frac{4}{e^2} \\ &= \left(\frac{1 - |\mathcal{B}_1| \log(\mathcal{K})}{e \cdot \mathcal{K}} + \frac{|\mathcal{B}_1|}{\mathcal{K}} \frac{4}{e^2} \right) \cdot \|\Gamma_1 - \Gamma_2\|_{[p]}. \end{aligned}$$

Proof of Part 1(b): We use Proposition 8 on page 20. By applying **Claim 1** with $\left|\vec{\xi}\right| = 1$, we have:

$$\left\| (\Phi_1 - \Phi_2)_{|\mathbb{S}^{D-1}} \right\|_\infty < \|\Gamma_1 - \Gamma_2\|_{var} \cdot \begin{cases} 1 + |\mathcal{B}_\alpha| & if \quad \alpha \neq 1 \\ 1 & if \quad \alpha = 1 \end{cases}$$

$$= \|\Gamma_1 - \Gamma_2\|_{var} \cdot \mathcal{D}_\alpha.$$

therefore, $\left\|(\Phi_1 - \Phi_2)_{|\mathbb{S}^{D-1}}\right\|_1 < \|\Gamma_1 - \Gamma_2\|_{var} \cdot \mathcal{A} \cdot \mathcal{D}_\alpha$

Thus, when $\alpha \neq 1$, Proposition 8 on page 20 says:

$$\begin{aligned} \|\chi_1 - \chi_2\|_1 &< \frac{\mathcal{C}_\alpha}{\mathcal{K}^{1+D/\alpha}} \cdot \left\|(\Phi_1 - \Phi_2)_{|\mathbb{S}^{D-1}}\right\|_1 \\ &= \frac{\mathcal{C}_\alpha}{\mathcal{K}^{1+D/\alpha}} \cdot \|\Gamma_1 - \Gamma_2\|_{var} \cdot \mathcal{A} \cdot \mathcal{D}_\alpha. \end{aligned}$$

When $\alpha = 1$, Proposition 8 on page 20 says:

$$\begin{aligned} \|\chi_1 - \chi_2\|_1 &< \frac{\mathcal{C}_1}{\mathcal{K}^{D+1}} \cdot \left\|(\Phi_1 - \Phi_2)_{|\mathbb{S}^{D-1}}\right\|_1 + \mathcal{A} \cdot |\mathcal{B}_1| \cdot \mathcal{F}(\mathcal{K}) \cdot \|\Gamma_1 - \Gamma_2\|_{var} \\ &= \frac{\mathcal{C}_1}{\mathcal{K}^{D+1}} \cdot \|\Gamma_1 - \Gamma_2\|_{var} \cdot \mathcal{A} \cdot \mathcal{D}_1. + \mathcal{A} \cdot |\mathcal{B}_1| \cdot \mathcal{F}(\mathcal{K}) \cdot \|\Gamma_1 - \Gamma_2\|_{var} \\ &= \mathcal{A} \cdot \left(\frac{\mathcal{C}_1 \mathcal{D}_1}{\mathcal{K}^{D+1}} + |\mathcal{B}_1| \cdot \mathcal{F}(\mathcal{K}) \right) \cdot \|\Gamma_1 - \Gamma_2\|_{var}. \end{aligned}$$

Proof of Part 2(b): Again we'll use Proposition 8 on page 20. By applying **Claim 2** with $\left|\vec{\xi}\right| = 1$, we again have:

$$\left\|(\Phi_1 - \Phi_2)_{|\mathbb{S}^{D-1}}\right\|_\infty < \|\Gamma_1 - \Gamma_2\|_{[q]} \cdot \mathcal{D}_\alpha,$$

both when $\alpha \neq 1$ and when $\alpha = 1$. Thus,

$$\left\|(\Phi_1 - \Phi_2)_{|\mathbb{S}^{D-1}}\right\|_1 < \mathcal{A} \cdot \|\Gamma_1 - \Gamma_2\|_{[q]} \cdot \mathcal{D}_\alpha.$$

When $\alpha \neq 1$, Proposition 8 on page 20 says:

$$\begin{aligned} \|\chi_1 - \chi_2\|_1 &< \frac{\mathcal{C}_\alpha}{\mathcal{K}^{1+D/\alpha}} \cdot \left\|(\Phi_1 - \Phi_2)_{|\mathbb{S}^{D-1}}\right\|_1 \\ &< \frac{\mathcal{C}_\alpha}{\mathcal{K}^{1+D/\alpha}} \cdot \mathcal{A} \cdot \|\Gamma_1 - \Gamma_2\|_{[q]} \cdot \mathcal{D}_\alpha. \end{aligned}$$

When $\alpha = 1$, Proposition 8 on page 20 says:

$$\begin{aligned} \|\chi_1 - \chi_2\|_1 &< \frac{\mathcal{C}_1}{\mathcal{K}^{D+1}} \cdot \left\|(\Phi_1 - \Phi_2)_{|\mathbb{S}^{D-1}}\right\|_1 + \mathcal{A}\,|\mathcal{B}_1| \cdot \mathcal{F}(\mathcal{K}) \cdot \|\Gamma_1 - \Gamma_2\|_{var} \\ &< \frac{\mathcal{C}_1}{\mathcal{K}^{D+1}} \cdot \mathcal{A} \cdot \|\Gamma_1 - \Gamma_2\|_{[p]} \cdot \mathcal{D}_1 + \mathcal{A}\,|\mathcal{B}_1| \cdot \mathcal{F}(\mathcal{K}) \cdot \|\Gamma_1 - \Gamma_2\|_{var} \end{aligned}$$

□

Corollary 14: *Let $\alpha \in [0,2)$, and, for $k = 1, 2$, let ρ_k be stable probability measures on $\mathbb{R}^D$ with density functions $F_k : \mathbb{R}^D \longrightarrow [0, \infty)$ and spectral measures Γ_k. Let $\mathcal{K} = \min\{\mathcal{K}(\rho_1), \mathcal{K}(\rho_2)\}$ where $\mathcal{K}(\rho_k)$ is as in equation (3.1) on page 19. Then there is a constant $\mathcal{G}_\alpha(\mathcal{K}) > 0$ so that*

1. $\|F_1 - F_2\|_\infty \ < \ \mathcal{G}_\alpha(\mathcal{K}) \cdot \|\Gamma_1 - \Gamma_2\|_{var}$.

2. *If Γ_1, Γ_2 are nice measures and $\alpha \neq 1$, then, for any $p \in [1, \infty]$,*

$$\|F_1 - F_2\|_\infty \ < \ \mathcal{G}_\alpha(\mathcal{K}) \cdot \|\Gamma_1 - \Gamma_2\|_{[p]} .$$

Proof: Set $\mathcal{G}_\alpha(\mathcal{K}) := \dfrac{\mathcal{C}_\alpha \mathcal{AD}_\alpha}{\mathcal{K}^{1+D/\alpha}}$, if $\alpha \neq 1$, and $\mathcal{G}_1(\mathcal{K}) := \dfrac{\mathcal{C}_1 \mathcal{AD}_1}{\mathcal{K}^{1+D}} + |\mathcal{B}_1| \cdot \mathcal{F}(\mathcal{K})$.

Apply **Part 1(b)** and **Part 2(b)** of Theorem 13 on page 28 to translate bounds $\|\Gamma_1 - \Gamma_2\|_*$ (for $* = [p]$ or var) into bounds on $\|\chi_1 - \chi_2\|_1$, where χ_k is the characteristic function of ρ_k. Then use **Part 1** of Lemma 3 on page 16 to translate these into bounds on $\|F_1 - F_2\|_\infty$. □

Corollary 15: *Suppose $\alpha \in [0,2)$, and that $[\rho_k|_{k=1}^\infty]$ is a sequence of α-stable probability measures on $\mathbb{R}^D$, with density functions $[F_k|_{k=1}^\infty]$ and spectral measures $[\Gamma_k|_{k=1}^\infty]$. Let ρ_∞ be some other α-stable measure with density F_∞ and spectral measure Γ_∞. Suppose that $\liminf_{k \in [1 \ldots \infty]} \mathcal{K}(\rho_k) > 0$.*

If: $\left(\lim_{k \to \infty} \|\Gamma_k - \Gamma_\infty\|_{var} \ = \ 0 \right)$

or: $\left(\alpha \neq 1 \ \text{ and } \ \lim_{k \to \infty} \|\Gamma_k - \Gamma_\infty\|_{[p]} \ = \ 0, \ \text{ for some } p \in [1, \infty] \right)$,

$$\text{then, for every } q \in [1, \infty], \quad \lim_{k \to \infty} \|F_k - F_\infty\|_q \ = \ 0.$$

Proof: This follows from the previous Corollary, and from **Part 2** of Lemma 3 on page 16. □

3.3 Spectral Approximation by Atoms

Now suppose that Γ_1 is an arbitrary measure on $\mathbb{S}^{D-1}$, and Γ_2 is some "approximation" of Γ_1 by a *sum of atoms*. From a computational point of view, it would be very convenient to approximate Γ_1 in this fashion. Of course, we can approximate Γ_1 as closely as we like in the weak* topology with atoms; hence, we can apply the results of §3.1. However, Γ_2 will still probably be very far from Γ_1 in total variation norm. For example, if Γ_1 is absolutely continuous with respect to the Lebesgue measure $\mathcal{L}^{\text{leb}}$, then the total variation distance between Γ_1 and Γ_2 will always be $\|\Gamma_1\| + \|\Gamma_2\|$. Thus, we cannot apply the results of §3.2. However, the kernels $\eta_{\vec{\xi}}$ which we are integrating against Γ_1 are very "smooth", so this does not matter too much.

Lemma 16: *Let $\eta_{\vec{\xi}}$ be defined as on page 19, and let $\mathbf{s}, \mathbf{u} \in \mathbb{S}^{D-1}$. Then*

$$\left|\eta_{\vec{\xi}}(\mathbf{s}) - \eta_{\vec{\xi}}(\mathbf{u})\right| \quad < \quad \Delta_\alpha\left(\vec{\xi},\ \mathbf{s}-\mathbf{u}\right)$$

Where,

$$\text{if } \alpha > 1, \text{ then } \quad \Delta_\alpha\left(\vec{\xi},\ \mathbf{s}-\mathbf{u}\right) \quad = \quad \alpha\left(1+|\mathcal{B}_\alpha|\right)\left|\vec{\xi}\right|^\alpha |\mathbf{s}-\mathbf{u}|\,;$$

$$\text{if } \alpha < 1, \text{ then } \quad \Delta_\alpha\left(\vec{\xi},\ \mathbf{s}-\mathbf{u}\right) \quad = \quad \alpha\left(1+|\mathcal{B}_\alpha|\right)\left|\vec{\xi}\right|^\alpha \left(1-\frac{|\mathbf{s}-\mathbf{u}|^2}{4}\right)^{(\alpha-1)/2} \cdot |\mathbf{s}-\mathbf{u}|\,;$$

Proof:

Claim 1: *Suppose $\alpha \neq 1, 2$. Then for any $\theta \in \mathbb{S}^{D-1}$,*

$$\nabla\eta_{\vec{\xi}}(\theta) \quad = \quad \alpha\left(\left\langle\vec{\xi},\theta\right\rangle^{\langle\alpha-1\rangle} \ + \ \mathcal{B}_\alpha\left|\left\langle\vec{\xi},\theta\right\rangle\right|^{\alpha-1}\mathbf{i}\right)\cdot\vec{\xi}$$

Proof: First note that, if $f_\alpha(x) = |x|^\alpha$, and $g_\alpha(x) = x^{\langle\alpha\rangle}$, then $f'_\alpha(x) = g_{(\alpha-1)}$, while $g'_\alpha(x) = f_{(\alpha-1)}$. For any $d \in [1..D]$, apply the Chain Rule to $\eta_{\vec{\xi}}(\bullet) = f_\alpha\left(\left\langle\vec{\xi},\bullet\right\rangle\right) + \mathbf{i}\cdot\mathcal{B}_\alpha g_\alpha\left(\left\langle\vec{\xi},\bullet\right\rangle\right)$ to get:

$$\begin{aligned}\partial_d\ \eta_{\vec{\xi}}(\theta) \quad &= \quad \alpha\left\langle\vec{\xi},\theta\right\rangle^{\langle\alpha-1\rangle}\left(\partial_d\left\langle\vec{\xi},\bullet\right\rangle\right)(\theta) \\ &\qquad + \ \mathcal{B}_\alpha\alpha\left|\left\langle\vec{\xi},\theta\right\rangle\right|^{\alpha-1}\mathbf{i}\left(\partial_d\left\langle\vec{\xi},\bullet\right\rangle\right)(\theta)\end{aligned}$$

$$= \alpha \left(\left\langle \vec{\xi}, \theta \right\rangle^{\langle \alpha-1 \rangle} + \mathcal{B}_\alpha \left| \left\langle \vec{\xi}, \theta \right\rangle \right|^{\alpha-1} \mathbf{i} \right) \cdot \xi_d.$$

$$\text{So, if } \quad \mathbf{e}_d = (\underbrace{0, \ldots, 0}_{d-1}, 1, \ldots, 0),$$

$$\text{then } \quad \nabla \eta_{\vec{\xi}}(\theta) = \sum_{d=1}^{D} \partial_d \, \eta_{\vec{\xi}}(\theta) \cdot \mathbf{e}_d$$

$$= \sum_{d=1}^{D} \alpha \left(\left\langle \vec{\xi}, \theta \right\rangle^{\langle \alpha-1 \rangle} + \mathcal{B}_\alpha \left| \left\langle \vec{\xi}, \theta \right\rangle \right|^{\alpha-1} \mathbf{i} \right) \cdot \xi_d \cdot \mathbf{e}_d$$

$$= \alpha \left(\left\langle \vec{\xi}, \theta \right\rangle^{\langle \alpha-1 \rangle} + \mathcal{B}_\alpha \left| \left\langle \vec{\xi}, \theta \right\rangle \right|^{\alpha-1} \mathbf{i} \right) \cdot \vec{\xi}.$$

..... □ [Claim 1]

If $\mathbf{s}$ and $\mathbf{u}$ are close together, then by the Mean Value Theorem

$$\eta_{\vec{\xi}}(\mathbf{s}) - \eta_{\vec{\xi}}(\mathbf{u}) = \nabla \eta_{\vec{\xi}}(\vec{x}) \cdot (\mathbf{s} - \mathbf{u})$$

for some $\vec{x}$ on the line segment between $\mathbf{u}$ and $\mathbf{s}$. But by **Claim 1**,

$$|\nabla \eta_\xi(\vec{x})| \leq \alpha \left| \left\langle \vec{\xi}, \vec{x} \right\rangle \right|^{\alpha-1} (1 + |\mathcal{B}_\alpha|) \cdot \left| \vec{\xi} \right|$$

$$\leq \alpha \left| \vec{\xi} \right|^{\alpha-1} |\vec{x}|^{\alpha-1} (1 + |\mathcal{B}_\alpha|) \cdot \left| \vec{\xi} \right|$$

Case ($\alpha > 1$)**:** In this case, $\alpha - 1 > 0$, and $|\vec{x}| < 1$, so we can conclude that $|\vec{x}|^{\alpha-1} < 1$, and thus, get $|\nabla \eta_\xi(\vec{x})| < \alpha \left| \vec{\xi} \right|^{\alpha-1} \cdot (1 + |\mathcal{B}_\alpha|) \cdot \left| \vec{\xi} \right|$.

Case ($\alpha < 1$)**:** Now $\alpha - 1 < 0$, so things are more complicated; instead of bounding $|\vec{x}|$ above, we must must bound it below. By construction, $\vec{x} = \lambda \mathbf{s} + (1 - \lambda)\mathbf{u}$ for some $\lambda \in [0, 1]$; and $|\mathbf{u}| = 1 = |\mathbf{s}|$. Thus, $|\vec{x}|$ takes on the minimum value when $\lambda = 1/2$. Hence, $|\vec{x}| > |\mathbf{s} + \mathbf{u}| / 2$. Now apply the Parallelogram Law ([33], Theorem 2.3):

$$|\mathbf{s} + \mathbf{u}|^2 + |\mathbf{s} - \mathbf{u}|^2 = 2 |\mathbf{s}|^2 + 2 |\mathbf{u}|^2 = 4$$

Thus, $|2 \cdot \vec{x}|^2 \geq |\mathbf{s} + \mathbf{u}|^2 = 4 - |\mathbf{s} - \mathbf{u}|^2$, so we conclude that

$$|\vec{x}| \geq \left(1 - \frac{|\mathbf{s} - \mathbf{u}|^2}{4} \right)^{1/2}, \text{ and thus, } |\vec{x}|^{\alpha-1} \leq \left(1 - \frac{|\mathbf{s} - \mathbf{u}|^2}{4} \right)^{(\alpha-1)/2}.$$

□

Theorem 17: *(Approximation by Atoms)*
Suppose $\alpha \neq 1$. Let Γ_1 be a measure on $\mathbb{S}^{D-1}$. Let $1 > \epsilon > 0$, and suppose that $\mathbb{A} \subset \mathbb{S}^{D-1}$ is a finite set, and suppose that $\left\{\mathcal{U}_{\mathbf{a}} \subset \mathbb{S}^{D-1}\,;\ \mathbf{a} \in \mathbb{A}\right\}$ is a partition of $\mathbb{S}^{D-1}$ into disjoint measurable subsets, so that, for all $\mathbf{a} \in \mathbb{A}$,

- $\mathbf{a} \in \mathcal{U}_{\mathbf{a}}$,
- *For all* $\mathbf{u} \in \mathcal{U}_{\mathbf{a}}$, $\quad |\mathbf{u} - \mathbf{a}| < \epsilon$.

Let Γ_2 be the discrete measure defined: $\Gamma_2 = \sum_{\mathbf{a} \in \mathbb{A}} \Gamma_1\left[\mathcal{U}_{\mathbf{a}}\right] \cdot \delta_{\mathbf{a}}$, and suppose that χ_1 and χ_2 are the Fourier Transforms of the α-stable probability measures on $\mathbb{R}^D$ induced by Γ_1 and Γ_2, respectively. Then:

1. $\|\chi_1 - \chi_2\|_\infty < \dfrac{\|\Gamma_1\|}{\mathcal{K} \cdot e} \cdot \mathcal{F}_\alpha \cdot \epsilon$.
2. $\|\chi_1 - \chi_2\|_1 < \dfrac{\|\Gamma_1\|\, \mathcal{C}_\alpha \mathcal{A}}{\mathcal{K}^{1+D/\alpha}} \cdot \mathcal{F}_\alpha \cdot \epsilon$.

$$\begin{aligned}
\text{where } \|\Gamma_1\| &= \Gamma_1\left[\mathbb{S}^{D-1}\right], \\
\mathcal{A} &= \mathbf{Area}\left[\mathbb{S}^{D-1}\right], \\
\mathcal{C}_\alpha &\quad \text{is as in Proposition 8 on page 20,} \\
\mathcal{K} &= \mathcal{K}(\rho_1), \quad \text{(Definition 6 on page 19)} \\
\text{and where } \mathcal{F}_\alpha &= \begin{cases} \alpha\left(1 + |\mathcal{B}_\alpha|\right) & if \quad \alpha > 1 \\ \alpha\left(1 + |\mathcal{B}_\alpha|\right)\left(\frac{4}{3}\right)^{1/2} & if \quad \alpha < 1 \end{cases}
\end{aligned}$$

Proof: Let Φ_1 and Φ_2 be the log-Fourier transforms associated to Γ_1 and Γ_2, respectively.

Claim 1: *For any $\vec{\xi} \in \mathbb{R}^D$,* $\left|\Phi_1(\vec{\xi}) - \Phi_2(\vec{\xi})\right| < \mathcal{F}_\alpha \cdot \epsilon \cdot \left|\vec{\xi}\right|^\alpha \|\Gamma\|$.

Proof: Let $\mathbf{a} \in \mathbb{A}$. Then

$$\begin{aligned}
&\left|\int_{\mathcal{U}_{\mathbf{a}}} \eta_{\vec{\xi}}(\mathbf{u})\ d\Gamma_1[\mathbf{u}] \ - \ \int_{\mathcal{U}_{\mathbf{a}}} \eta_{\vec{\xi}}(\mathbf{u})\ d\Gamma_2[\mathbf{u}]\right| \\
&= \left|\int_{\mathcal{U}_{\mathbf{a}}} \eta_{\vec{\xi}}(\mathbf{u})\ d\Gamma_1[\mathbf{u}] \ - \ \Gamma_1\left[\mathcal{U}_{\mathbf{a}}\right] \cdot \eta_{\vec{\xi}}(\mathbf{a})\right| \\
&= \left|\int_{\mathcal{U}_{\mathbf{a}}} \Big(\eta_{\vec{\xi}}(\mathbf{u}) \ - \ \eta_{\vec{\xi}}(\mathbf{a})\Big)\ d\Gamma_1[\mathbf{u}]\right|
\end{aligned}$$

$$\begin{aligned} &\leq \int_{\mathcal{U}_{\mathbf{a}}} \left| \eta_{\vec{\xi}}(\mathbf{u}) \; - \; \eta_{\vec{\xi}}(\mathbf{a}) \right| \; d\Gamma_1[\mathbf{u}] \\ &\leq_{(1)} \int_{\mathcal{U}_{\mathbf{a}}} \Delta_\alpha(\vec{\xi}, \; \mathbf{u}-\mathbf{a}) \; d\Gamma_1[\mathbf{u}] \end{aligned}$$

where (1) follows from Lemma 16 on page 34. But, when $\alpha > 1$,

$$\begin{aligned} \int_{\mathcal{U}_{\mathbf{a}}} \Delta_\alpha(\vec{\xi}, \; \mathbf{u}-\mathbf{a}) \; d\Gamma_1[\mathbf{u}] &= \int_{\mathcal{U}_{\mathbf{a}}} \alpha \mathcal{F}_\alpha \left|\vec{\xi}\right|^\alpha |\mathbf{u}-\mathbf{a}| \; d\Gamma_1[\mathbf{u}] \\ &\leq \alpha \mathcal{F}_\alpha \left|\vec{\xi}\right|^\alpha \int_{\mathcal{U}_{\mathbf{a}}} \epsilon \; d\Gamma_1[\mathbf{u}] \\ &= \alpha \mathcal{F}_\alpha \left|\vec{\xi}\right|^\alpha \epsilon \cdot \Gamma_1 \left[\mathcal{U}_{\mathbf{a}}\right], \end{aligned}$$

while, when $\alpha < 1$, if $\mathbf{u} \in \mathcal{U}_{\mathbf{a}}$, then

$$\begin{aligned} \Delta_\alpha(\vec{\xi}, \; \mathbf{u}-\mathbf{a}) &= \alpha \left(1 + |\mathcal{B}_\alpha|\right) \left|\vec{\xi}\right|^\alpha \left(1 - \frac{|\mathbf{s}-\mathbf{u}|^2}{4}\right)^{(\alpha-1)/2} \cdot |\mathbf{s}-\mathbf{u}| \\ &\leq \alpha \left(1 + |\mathcal{B}_\alpha|\right) \left|\vec{\xi}\right|^\alpha \left(1 - \frac{\epsilon^2}{4}\right)^{(\alpha-1)/2} \cdot \epsilon \\ &\leq \alpha \left(1 + |\mathcal{B}_\alpha|\right) \left|\vec{\xi}\right|^\alpha \left(1 - \frac{1}{4}\right)^{(\alpha-1)/2} \cdot \epsilon \\ &= \alpha \left(1 + |\mathcal{B}_\alpha|\right) \left|\vec{\xi}\right|^\alpha \left(\frac{3}{4}\right)^{(\alpha-1)/2} \cdot \epsilon \\ &\leq \alpha \left(1 + |\mathcal{B}_\alpha|\right) \left|\vec{\xi}\right|^\alpha \left(\frac{4}{3}\right)^{1/2} \cdot \epsilon \\ &= \mathcal{F}_\alpha \cdot \epsilon \cdot \left|\vec{\xi}\right|^\alpha \end{aligned}$$

$$\begin{aligned} \text{Hence,} \quad \int_{\mathcal{U}_{\mathbf{a}}} \Delta_\alpha(\vec{\xi}, \; \mathbf{u}-\mathbf{a}) \; d\Gamma_1[\mathbf{u}] &\leq \int_{\mathcal{U}_{\mathbf{a}}} \mathcal{F}_\alpha \cdot \epsilon \cdot \left|\vec{\xi}\right|^\alpha \; d\Gamma_1[\mathbf{u}] \\ &= \mathcal{F}_\alpha \cdot \epsilon \cdot \left|\vec{\xi}\right|^\alpha \Gamma_1 \left[\mathcal{U}_{\mathbf{a}}\right] \end{aligned}$$

$$\begin{aligned} \text{Thus,} \quad &\left| \int_{\mathbb{S}^{D-1}} \eta_{\vec{\xi}}(\mathbf{u}) \; d\Gamma_1[\mathbf{u}] \; - \; \int_{\mathbb{S}^{D-1}} \eta_{\vec{\xi}}(\mathbf{u}) \; d\Gamma_2[\mathbf{u}] \right| \\ &= \left| \sum_{\mathbf{a} \in \mathbb{A}} \int_{\mathcal{U}_{\mathbf{a}}} \eta_{\vec{\xi}}(\mathbf{u}) \; d\Gamma_1[\mathbf{u}] \; - \; \sum_{\mathbf{a} \in \mathbb{A}} \int_{\mathcal{U}_{\mathbf{a}}} \eta_{\vec{\xi}}(\mathbf{u}) \; d\Gamma_2[\mathbf{u}] \right| \end{aligned}$$

$$\begin{aligned} &\leq \sum_{\mathbf{a}\in\mathbb{A}} \left| \int_{\mathcal{U}_\mathbf{a}} \eta_{\vec{\xi}}(\mathbf{u})\ d\Gamma_1[\mathbf{u}] \ - \ \int_{\mathcal{U}_\mathbf{a}} \eta_{\vec{\xi}}(\mathbf{u})\ d\Gamma_2[\mathbf{u}] \right| \\ &\leq \sum_{\mathbf{a}\in\mathbb{A}} \mathcal{F}_\alpha \cdot \epsilon \left|\vec{\xi}\right|^\alpha \Gamma_1\left[\mathcal{U}_\mathbf{a}\right] \\ &= \mathcal{F}_\alpha \cdot \epsilon \left|\vec{\xi}\right|^\alpha \|\Gamma_1\| \,. \end{aligned}$$

..... □ [Claim 1]

The prove **Part 1**, use Proposition 4 on page 17, with $C_1 = \mathcal{F}_\alpha \cdot \epsilon \cdot \|\Gamma_1\|$ as in **Claim 1**, and with $C_0 = \mathcal{K}$, as given by Lemma 10 on page 27.

To prove **Part 2**, use Proposition 8 on page 20. By applying **Claim 1** with $\left|\vec{\xi}\right| = 1$, we have:

$$\begin{aligned} \left\| (\Phi_1 - \Phi_2)_{|\mathbb{S}^{D-1}} \right\|_\infty &< \|\Gamma_1\|\, \mathcal{F}_\alpha \cdot \epsilon \\ \text{and thus,} \quad \left\| (\Phi_1 - \Phi_2)_{|\mathbb{S}^{D-1}} \right\|_1 &< \|\Gamma_1\|\, \mathcal{A}\mathcal{F}_\alpha \cdot \epsilon. \end{aligned}$$

□

Remark 18: The set $\mathbb{A}$ of atoms in the previous theorem must be ϵ-dense in $\mathbb{S}^{D-1}$. To do this will require a the set $\mathbb{A}$ to have a cardinality greater than:

$$\frac{2\pi^{1/2}\left(\frac{D-1}{2}\right)!}{\left(\frac{D}{2}-1\right)!} \cdot \epsilon^{1-D}.$$

Proof: If ϵ is small, then an ϵ-sized "disk" on the surface of $\mathbb{S}^{D-1}$ is approximately flat, so its surface area is approximately the $(D-1)$-volume of a $(D-1)$-sphere of radius ϵ; by [48], p. 76, this is

$$\frac{\pi^{(D-1)/2}}{\left(\frac{D-1}{2}\right)!}\epsilon^{D-1}$$

(here, for any real $r \in \mathbb{R}$, $r! = \Gamma(r+1)$ is the Γ-function of r; we are using the notation "$r!$" to avoid confusion with the prior use of Γ as a spectral measure.)

Meanwhile, the surface area of $\mathbb{S}^{D-1}$ is $\dfrac{2\pi^{D/2}}{\left(\frac{D}{2}-1\right)!}$.

Hence, the minimum number of ϵ-disks necessary to cover $\mathbb{S}^{D-1}$ is

$$\frac{2\pi^{1/2}\left(\frac{D-1}{2}\right)!}{\left(\frac{D}{2}-1\right)!} \cdot \epsilon^{1-D}$$

□

This is in fact a very crude lower bound, because it assumes a "perfect" packing of $\mathbb{S}^{D-1}$ by disks, which is of course impossible. The actual cardinality of $\mathbb{A}$ will probably be several times this amount. However, it will be proportional to N^{D-1}.

Theorem 17 is somewhat inconvenient, since it requires one to evaluate Γ on subsets of $\mathbb{S}^{D-1}$ before one can build an approximation, thereby "putting the cart before the horse". However, when the spectral measure is absolutely continuous, and its Radon-Nikodym derivative is Lipschitz, there is a more convenient weak* atomic approximation we can use.

Lemma 19: *Suppose Γ is a measure on $\mathbb{S}^{D-1}$. For all $N \in \mathbb{N}$, let $\mathbb{A}_N \subset \mathbb{S}^{D-1}$ be a finite subset, and, for $k = 0, 1$, and all $N \in \mathbb{N}$, let*

$$\Gamma_k^{[N]} = \sum_{\mathbf{a} \in \mathbb{A}_N} \gamma_{k,\mathbf{a}} \delta_{\mathbf{a}}$$

be a discrete measure, with $\gamma_{k,\mathbf{a}} \in \mathbb{R}$, for all $\mathbf{a} \in \mathbb{A}_N$. Suppose that there is some $C > 0$ so that, for all $N \in \mathbb{N}$ and all $\mathbf{a} \in \mathbb{A}_N$,

$$|\gamma_{0,\mathbf{a}} - \gamma_{1,\mathbf{a}}| < \frac{C}{N \cdot \mathcal{C}_{and}[\mathbb{A}_N]}.$$

Then: $\left(\mathbf{wk}^*\text{-}\lim_{N\to\infty} \Gamma_0^{[N]} = \Gamma\right) \Longrightarrow \left(\mathbf{wk}^*\text{-}\lim_{N\to\infty} \Gamma_1^{[N]} = \Gamma\right)$.

Proof: Let $f : \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$ be continuous, with $\|f\|_\infty = M$. Then for large enough N,

$$\begin{aligned} \left|\left\langle f,\ \Gamma_0^{[N]}\right\rangle - \left\langle f,\ \Gamma_1^{[N]}\right\rangle\right| &= \left|\sum_{\mathbf{a}\in\mathbb{A}_N} f(\mathbf{a})\gamma_{0,\mathbf{a}} - \sum_{\mathbf{a}\in\mathbb{A}_N} f(\mathbf{a})\gamma_{1,\mathbf{a}}\right| \\ &\leq \sum_{\mathbf{a}\in\mathbb{A}_N} |f(\mathbf{a})| \cdot |\gamma_{0,\mathbf{a}} - \gamma_{1,\mathbf{a}}| \end{aligned}$$

$$\begin{aligned} &\leq \sum_{\mathbf{a}\in\mathbb{A}_N} M \cdot \frac{C}{N \cdot \mathcal{C}_{ard}\left[\mathbb{A}_N\right]} \\ &= \mathcal{C}_{ard}\left[\mathbb{A}_N\right] \cdot \frac{M \cdot C}{N \cdot \mathcal{C}_{ard}\left[\mathbb{A}_N\right]} \\ &\leq \frac{M \cdot C}{N} \\ &\xrightarrow[N\to\infty]{} 0. \end{aligned}$$

Thus, for every continuous function,

$$\lim_{n\to\infty} \left\langle f,\ \Gamma_1^{[N]} \right\rangle \ =\ \lim_{n\to\infty} \left\langle f,\ \Gamma_0^{[N]} \right\rangle \ =\ \langle f,\ \Gamma \rangle\,.$$

□

Definition 20: *Mesh Sequence*

A **mesh sequence** *in* $\mathbb{S}^{D-1}$ *is sequence of finite subsets* $\left\{\mathbb{A}_N \subset \mathbb{S}^{D-1}\ ;\ N \in \mathbb{N}\right\}$, *accompanied by a collection of measurable sets* $\{\mathcal{U}_{\mathbf{a}}\ ;\ \mathbf{a} \in \mathbb{A}_N\}$ *for each* $N \in \mathbb{N}$, *and a constant* $C > 0$ *so that, for all* N,

1. $\mathcal{C}_{ard}\left[\mathbb{A}_N\right] \ <\ C \cdot N^{D-1}$
2. $\mathbb{A}_N$ *is* (C/N)*-dense in* $\mathbb{S}^{D-1}$.
3. *For all* $\mathbf{a} \in \mathbb{A}_N$,
   (a) *For all* $\mathbf{u} \in \mathcal{U}_{\mathbf{a}}$, $\mathbf{dist}\left[\mathbf{a}, \mathbf{u}\right] <\ C/N$.
   (b) $\mathcal{L}^{bsg}\left[\mathcal{U}_{\mathbf{a}}\right] \leq \mathcal{A}/N^{D-1}$ *(where* $\mathcal{A} = \mathbf{Area}[\mathbb{S}^{D-1}]$.
4. $\mathbb{S}^{D-1} \ =\ \bigsqcup_{\mathbf{a}\in\mathbb{A}_N} \mathcal{U}_{\mathbf{a}}$.

Lemma 21: *For every* $D \geq 2$, *a mesh sequence exists on* $\mathbb{S}^{D-1}$.

Proof: Let $\mathbb{I}^D = [-1,1] \times \ldots \times [-1,1]$, and let $\mathbb{J} = \partial\mathbb{I}^D$. There is a natural projection $\phi : \mathbb{J} \longrightarrow \mathbb{S}^{D-1}$ where $\phi(\mathbf{x}) = \dfrac{\mathbf{x}}{|\mathbf{x}|}$.

ϕ is clearly a bijective contraction mapping. Thus, a mesh sequence defined on $\mathbb{J}$ will clearly project to a mesh sequence on $\mathbb{S}^{D-1}$. To define a mesh sequence on $\mathbb{J}$, simply represent $\mathbb{J}$ as a union of $2D$ distinct cubes of dimension $D-1$ and edge length 2; defining a mesh sequence on each of these is straightforward. □

Proposition 22: *(Approximation by Atoms for Lipschitz Measures)*
Suppose $\alpha \neq 1$. *Suppose* Γ *is absolutely continuous with respect to Lebesgue measure, with* $d\Gamma = \mathbf{g}\, d\mathcal{L}^{bsg}$, *and suppose that* $\mathbf{g}$ *is a* **Lipschitz** *function. Let* $\left[\left\{\mathbb{A}_N \subset \mathbb{S}^{D-1}\,;\, N \in \mathbb{N}\right\}\,;\, \{\mathcal{U}_{\mathbf{a}}\,;\, \mathbf{a} \in \mathbb{A}_N, N \in \mathbb{N}\}\,;\, C\right]$ *determine a mesh sequence on* $\mathbb{S}^{D-1}$.

Let $\Gamma_1^{[N]} := \sum_{\mathbf{a}\in\mathbb{A}_N} \gamma_{1,\mathbf{a}}\delta_{\mathbf{a}}$, *where,* $\forall \mathbf{a} \in \mathcal{A}$, $\gamma_{1,\mathbf{a}} = \mathbf{g}(\mathbf{a}) \cdot \mathcal{L}^{bsg}[\mathcal{U}_{\mathbf{a}}]$.

Then $\mathbf{wk}^*\!-\!\lim_{N\to\infty} \Gamma_1^{[N]} = \Gamma$.

Proof: Define $\Gamma_0^{[N]} := \sum_{\mathbf{a}\in\mathbb{A}_N} \gamma_{0,\mathbf{a}}\delta_{\mathbf{a}}$, where $\gamma_{0,\mathbf{a}} = \Gamma[\mathcal{U}_{\mathbf{a}}]$. Thus, $\mathbf{wk}^*\!-\!\lim_{N\to\infty} \Gamma_0^{[N]} = \Gamma$.

Claim 1: *There is a constant* $c > 0$ *so that, for any* $N \in \mathbb{N}$ *and* $\mathbf{a} \in \mathbb{A}_N$,

$$|\gamma_{0,\mathbf{a}} - \gamma_{1,\mathbf{a}}| < \frac{c}{N \cdot \mathcal{C}_{ard}\left[\mathbb{A}_N\right]}$$

Proof: Since $\mathbf{g}$ is Lipschitz, there is some constant $L > 0$ so that, for any close enough $\mathbf{x}, \mathbf{y} \in \mathbb{S}^{D-1}$, $\mathbf{g}(\mathbf{x}) - \mathbf{g}(\mathbf{y}) < L \cdot \mathbf{dist}\,[\mathbf{x}, \mathbf{y}]$. Thus,

$$\begin{aligned}
|\gamma_{1,\mathbf{a}} - \gamma_{0,\mathbf{a}}| &= |\mathbf{g}(\mathbf{a}) \cdot \mathcal{L}^{bsg}[\mathcal{U}_{\mathbf{a}}] - \Gamma[\mathcal{U}_{\mathbf{a}}]| \\
&= \left|\int_{\mathcal{U}_{\mathbf{a}}} \mathbf{g}(\mathbf{a})\; d\mathcal{L}^{bsg}[\mathbf{u}] - \int_{\mathcal{U}_{\mathbf{a}}} \mathbf{g}(\mathbf{u})\; d\mathcal{L}^{bsg}[\mathbf{u}]\right| \\
&= \left|\int_{\mathcal{U}_{\mathbf{a}}} \mathbf{g}(\mathbf{a}) - \mathbf{g}(\mathbf{u})\; d\mathcal{L}^{bsg}[\mathbf{u}]\right| \\
&\leq \int_{\mathcal{U}_{\mathbf{a}}} |\mathbf{g}(\mathbf{a}) - \mathbf{g}(\mathbf{u})|\; d\mathcal{L}^{bsg}[\mathbf{u}] \\
&\leq_{(1)} \int_{\mathcal{U}_{\mathbf{a}}} L \cdot \mathbf{dist}\,[\mathbf{a}, \mathbf{u}]\; d\mathcal{L}^{bsg}[\mathbf{u}] \\
&\leq_{(2)} \int_{\mathcal{U}_{\mathbf{a}}} \frac{L \cdot C}{N} d\mathcal{L}^{bsg}[\mathbf{u}] \\
&= \frac{L \cdot C}{N} \mathcal{L}^{bsg}[\mathcal{U}_{\mathbf{a}}] \\
&\leq_{(3)} \frac{L \cdot C \cdot \mathcal{A}}{N^D} \\
&\leq_{(4)} \frac{c}{N \cdot \mathcal{C}_{ard}\left[\mathbb{A}_N\right]}.
\end{aligned}$$

(1) By the Lipschitz condition.

(2) By mesh sequence condition 3(a).

(3) By mesh sequence condition 3(b).

(4) Where $\mathcal{C}_{ard}[\mathbb{A}_N] \leq C \cdot N^{D-1}$ by mesh sequence condition 1, and where $c := L \cdot C^2 \cdot \mathcal{A}$. ..... □ [Claim 1]

Now apply Lemma 19. ——————————————————————————□

Prior Research: A different approach to the problem of approximating a spectral measure by atoms is developed in [174]. The authors show how, for any uniform ϵ error bound on the probability density function, it is possible to approximate the spectral measure of an arbitrary multivariate stable distribution by a finite sum of atoms so as to satisfy this error bound. They prove the same result if we now measure the uniform error when treating the probability measure and its approximation as functionals on the space of Borel subsets of $\mathbb{R}^D$. These results are then utilized in [79], where the authors develop a method for simulation of multivariate stable random variables similar to that described in Appendix B.5; they use the results of [174] to bound the error of this simulation method.

Chapter 4

The Spherical Log-Characteristic Function

4.1 Preliminaries

Let $\mathbf{X}$ be a $D-$dimensional $\alpha-$stable random vector, and assume for simplicity that $\mathbf{X}$ has mean zero, so that we have log-characteristic function:

$$\Phi_{\mathbf{X}}(\vec{\xi}) \quad = \quad - \int_{\mathbb{S}^{D-1}} \left(\left| \left\langle \vec{\xi}, \mathbf{s} \right\rangle \right|^{\alpha} + \mathcal{B}_{\alpha} \left\langle \vec{\xi}, \mathbf{s} \right\rangle^{\langle \alpha \rangle} \mathbf{i} \right) \, d\Gamma[\mathbf{s}]$$

For any $\vec{\xi} \in \mathbb{R}^D$, let $\theta := \frac{\vec{\xi}}{\|\vec{\xi}\|}$ be the unit vector pointing in the same direction as $\vec{\xi}$. We can then rewrite the log-characteristic function:

$$\begin{aligned} \Phi_{\mathbf{X}}(\vec{\xi}) \quad &= \quad - \left\| \vec{\xi} \right\|^{\alpha} \cdot \int_{\mathbb{S}^{D-1}} \left(|\langle \theta, \mathbf{s} \rangle|^{\alpha} + \mathcal{B}_{\alpha} \langle \theta, \mathbf{s} \rangle^{\langle \alpha \rangle} \mathbf{i} \right) \, d\Gamma[\mathbf{s}] \\ &= \quad - \left\| \vec{\xi} \right\|^{\alpha} \cdot g(\theta) \end{aligned}$$

where $\mathbf{g} : \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$ is defined:

$$\mathbf{g}(\theta) \quad = \quad \int_{\mathbb{S}^{D-1}} \left(|\langle \theta, \mathbf{s} \rangle|^{\alpha} + \mathcal{B}_{\alpha} \langle \theta, \mathbf{s} \rangle^{\langle \alpha \rangle} \mathbf{i} \right) \, d\Gamma[\mathbf{s}]$$

$$\text{If we define:} \quad \eta_{\theta}(\mathbf{s}) \quad = \quad |\langle \theta, \mathbf{s} \rangle|^{\alpha} + \mathcal{B}_{\alpha} \langle \theta, \mathbf{s} \rangle^{\langle \alpha \rangle} \mathbf{i}$$

$$\text{then we can rewrite this:} \quad \mathbf{g}(\theta) \quad = \quad \int_{\mathbb{S}^{D-1}} \eta_{\theta}(\mathbf{s}) \, d\Gamma[\mathbf{s}].$$

Abusing notation, we might write, "$\mathbf{g}(\theta) = \eta * \Gamma(\theta)$", where $\eta : \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$ is defined:

$$\eta(\mathbf{s}) = |\langle \mathbf{e}_1, \mathbf{s} \rangle|^{\alpha} + \mathcal{B}_{\alpha} \langle \mathbf{e}_1, \mathbf{s} \rangle^{\langle \alpha \rangle} \mathbf{i}$$

If $D = 2$ or $D = 4$, then $\mathbb{S}^{D-1}$ is a topological group [179], and this "convolution" can be interpreted literally, via the formula:

$$\eta * \Gamma(\theta) = \int_{\mathbb{S}^{D-1}} \eta(\theta.\mathbf{s}^{-1}) \, d\Gamma[\mathbf{s}]$$

In other dimensions, however, $\mathbb{S}^{D-1}$ is not a topological group [6], and therefore, "convolution" *per se* is not well-defined. We can still conceive of $\mathbf{g}$ as a "convolution"; in order to do so, however, we must think of $\mathbb{S}^{D-1}$ as a homogeneous manifold under the action of $\mathbb{SO}^D[\mathbb{R}]$, and "pull back" Γ, $\mathbf{g}$, and η to measures and functions on $\mathbb{SO}^D[\mathbb{R}]$, where we can apply the methods of nonabelian harmonic analysis. This is dealt with in more detail in chapters 5 and 6.

$\mathbf{g}$ is really just the restriction to $\mathbb{S}^{D-1}$ of the log-characteristic function of the distribution. Thus, we will refer to $\mathbf{g}$ as the **spherical log-characteristic function**, or **SLCF**.

4.2 Properties of the SLCF

Clearly, Γ completely determines $\mathbf{g}$. The converse is also true: $\mathbf{g}$ completely describes the characteristic function of $\mathbf{X}$, and thus, the probability distribution of $\mathbf{X}$, and which is associated to a unique spectral measure. Thus, in principal, Γ and $\mathbf{g}$ provide equal information about the structural properties of the multivariate distribution (ie. correlation, symmetries, etc.). However, because of the very "smoothing" nature of the transformation $\Gamma \mapsto \mathbf{g}$, the information in $\mathbf{g}$ is displayed in a more subtle form.

For example, the following facts are not hard to verify:

- $\mathbf{g}$ detects whether $\mathbf{X}$ is **sub-Gaussian** (see Example 81 on page 147). Γ is uniformly distributed on $\mathbb{S}^{D-1}$ if and only if $\mathbf{g}$ is constant.
- Suppose that $\mathbf{X} = \sigma_1 \mathbf{X}_1 \mathbf{e}_1 + \sigma_2 \mathbf{X}_2 \mathbf{e}_2 + \ldots + \sigma_D \mathbf{X}_D \mathbf{e}_D$ is a sum of independent, symmetric, $\alpha-$stable random variables in separate dimensions, so that

$$\Gamma = \sum_{d=1}^{D} \frac{\sigma_d}{2} \left(\delta_{\mathbf{e}_d} + \delta_{(-\mathbf{e}_d)} \right).$$

Then for all $\theta \in \mathbb{S}^{D-1}$, $\mathbf{g}(\theta) = \|\theta\|_\sigma^\alpha$, where $\|\cdot\|_\sigma$ is the special $\ell^\alpha-$norm on $\mathbb{R}^D$ defined:

$$\|\xi_1, \ldots, \xi_D\|_\sigma = \left(\sum_{d=1}^{D} \sigma_d \cdot \xi_d^\alpha \right)^{1/\alpha}$$

4.3 Smoothness of the SLCF

Although Γ may be highly singular, $\mathbf{g}$ is quite "smooth", since it is the result of "convolving" Γ with a very smooth function.

In the case when $D = 2$, we can take advantage of the compact abelian group structure of $\mathbb{S}^1$, and characterize this smoothness by comparing the Fourier series of $\mathbf{g}$ and Γ.

Proposition 23: *Suppose* $\mathbf{X}$ *is an* α*-stable random vector in* $\mathbb{R}^2$*, with spectral measure* Γ *on* $\mathbb{S}^1$*, and let* $\mathbf{g} : \mathbb{S}^1 \longrightarrow \mathbb{C}$ *be the SLCF. Then, for all* $n \in \mathbb{Z}$, $\widehat{\mathbf{g}}_n = \widehat{\eta}_n \cdot \widehat{\Gamma}_n$, *where*[1]

$$\eta(\theta) = |\cos(\theta)|^\alpha + \mathcal{B}_\alpha \cos(\theta)^{\langle\alpha\rangle}\mathbf{i}$$

As $|n| \to \infty$, $|\hat{\eta}_n| \to 0$ *exponentially.*

Proof: It is easy to verify that $\mathbf{g} = \eta * \Gamma$. The multiplication formula follows from elementary harmonic analysis [84],[48],[33]. The Fourier coefficients of η decay exponentially because η is analytic (see [84], Section I.4, p.26). □

Another way to to characterize the "smoothing" nature of the transformation $\Gamma \mapsto \mathbf{g}$ is to observe that, whereas Γ is an arbitrary measure, $\mathbf{g}$ is a continuous function on the sphere. Indeed, if $\alpha > 1$, then $\mathbf{g}$ is even Lipschitz.

Proposition 24: *Suppose* $\mathbf{X}$ *is an* α*-stable random vector in* $\mathbb{R}^D$*, with spectral measure* Γ *on* $\mathbb{S}^{D-1}$*. Let* $G := \Gamma\left[\mathbb{S}^{D-1}\right]$*, and let* $\mathbf{g} : \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$ *be the SLCF.*

For any $\alpha \in (0, 2)$*, the function* $\mathbf{g}$ *is continuous, in a manner depending only on the values of* α *and* G *(and not on the structure of* Γ*). To be specific:*

[1] Here, $\widehat{\mathbf{g}}_n$, $\widehat{\eta}_n$ and $\widehat{\Gamma}_n$ are the nth Fourier coefficients of $\mathbf{g}$, η and Γ respectively, for $n \in \mathbb{Z}$. For example, $\widehat{\mathbf{g}}_n = \int_0^{2\pi} \mathbf{g}(\theta) e^{-n \cdot \theta \mathbf{i}}\, d\theta$, where we identify $\mathbb{S}^1 \cong [0, 2\pi)$ in the obvious way.

1. *If $\alpha \leq 1$, then there is a function $\delta_\alpha : [0,\infty) \longrightarrow [0,\infty)$ so that, for any ϵ, if $\theta_1, \theta_2 \in \mathbb{S}^{D-1}$ are within $\delta_\alpha(\epsilon)$, then $|\mathbf{g}(\theta_1) - \mathbf{g}(\theta_2)| \leq G \cdot \epsilon$.*

2. *If $\alpha > 1$, $\mathbf{g}$ is (locally) Lipschitz[2]: for any $\theta_1, \theta_2 \in \mathbb{S}^{D-1}$, sufficiently close together,*

$$|\mathbf{g}(\theta_1) - \mathbf{g}(\theta_2)| \leq C_\alpha \cdot G \cdot |\theta_1 - \theta_2| + \mathcal{O}(|\theta_1 - \theta_2|^2) \qquad (4.1)$$

where $C_\alpha = \alpha \cdot (1 + \mathcal{B}_\alpha)$.

Proof:

Proof of Part 2: Suppose $\alpha > 1$.

Claim 1: *If θ_1 and θ_2 are sufficiently close, then $\|\eta_{\theta_1} - \eta_{\theta_2}\|_\infty < C_\alpha |\theta_1 - \theta_2|$.*

Proof: Let $\mathbf{s} \in \mathbb{S}^{D-1}$. Notice that

$$\begin{aligned} \Big| |\langle \theta_1, \mathbf{s} \rangle| - |\langle \theta_2, \mathbf{s} \rangle| \Big| &\leq_{(1)} \Big| \langle \theta_1, \mathbf{s} \rangle - \langle \theta_2, \mathbf{s} \rangle \Big| \\ &= \Big| \langle \theta_1 - \theta_2, \mathbf{s} \rangle \Big| \\ &\leq |\theta_1 - \theta_2| \cdot |\mathbf{s}| \\ &= |\theta_1 - \theta_2| \end{aligned}$$

(1) follows from the fact that, for any x and y, $||x| - |y|| < |x - y|$.

Let $f(x) := |x|^\alpha$, so that $f'(x) = \alpha \cdot x^{\alpha-1}$. Thus, for small h, $f(x+h) - f(x) = f'(x) \cdot h + \mathcal{O}(h^2) = \alpha \cdot x^{\alpha-1} h + \mathcal{O}(h^2)$; hence $|f(x+h) - f(x)| \leq \alpha \cdot |x|^{\alpha-1} |h| + \mathcal{O}(h^2)$. Thus,

$$\begin{aligned} |\mathbf{re}\,[\eta_{\theta_1}(\mathbf{s}) - \eta_{\theta_2}(\mathbf{s})]| &= |f(\langle \theta_1, \mathbf{s} \rangle) - f(\langle \theta_2, \mathbf{s} \rangle)| \\ &\leq \alpha \cdot |\langle \theta_1, \mathbf{s} \rangle|^{\alpha-1} \cdot |\theta_1 - \theta_2| + \mathcal{O}\left(|\theta_1 - \theta_2|^2\right) \\ &\leq_{(1)} \alpha \cdot |\theta_1 - \theta_2| + \mathcal{O}\left(|\theta_1 - \theta_2|^2\right) \end{aligned}$$

(1) $\theta_1, \mathbf{s} \in \mathbb{S}^{D-1}$, so $|\langle \theta_1, \mathbf{s} \rangle| < 1$, and thus, $|\langle \theta_1, \mathbf{s} \rangle|^{\alpha-1} < 1$, because $\alpha > 1$.

[2]Equation (4.1) is only a *local* Lipschitz property, yielding a linear bound on $|\mathbf{g}(\theta_1) - \mathbf{g}(\theta_2)|$ for θ_1, θ_2 close together. For our purposes in the proof of Proposition 25, this is sufficient. Actually, since $\mathbb{S}^{D-1}$ is compact, (4.1) is equivalent to the usual Lipschitz condition, albeit with a constant larger than C_α.

By a similar argument, we have:

$$|\mathbf{im}\left[\eta_{\theta_1}(\mathbf{s}) - \eta_{\theta_2}(\mathbf{s})\right]| \leq \alpha \cdot \mathcal{B}_\alpha \left|\theta_1 - \theta_2\right| + \mathcal{O}\left(\left|\theta_1 - \theta_2\right|^2\right)$$

$$\begin{aligned} \text{Thus,} \quad |\eta_{\theta_1}(\mathbf{s}) - \eta_{\theta_2}(\mathbf{s})| &= |\mathbf{re}\left[\eta_{\theta_1}(\mathbf{s}) - \eta_{\theta_2}(\mathbf{s})\right]| + |\mathbf{im}\left[\eta_{\theta_1}(\mathbf{s}) - \eta_{\theta_2}(\mathbf{s})\right]| \\ &\leq \alpha \cdot (1 + \mathcal{B}_\alpha) \left|\theta_1 - \theta_2\right| + \mathcal{O}\left(\left|\theta_1 - \theta_2\right|^2\right). \end{aligned}$$

..... □ [Claim 1]

$$\begin{aligned} \text{Thus,} \quad |\mathbf{g}(\theta_1) - \mathbf{g}(\theta_2)| &= \left| \int_{\mathbb{S}^{D-1}} \left(\eta_{\theta_1}(\mathbf{s}) - \eta_{\theta_2}(\mathbf{s})\right) \, d\Gamma[\mathbf{s}] \right| \\ &\leq \int_{\mathbb{S}^{D-1}} \left\|\eta_{\theta_1} - \eta_{\theta_2}\right\|_\infty \, d\Gamma[\mathbf{s}] \\ &\leq \Gamma[\mathbb{S}^{D-1}] \cdot C_\alpha \cdot \left|\theta_1 - \theta_2\right| + \mathcal{O}\left(\left|\theta_1 - \theta_2\right|^2\right) \quad \text{by Claim 1.} \end{aligned}$$

Proof of Part 1: The proof is similar to that of **Part 2**.

Claim 2: *For any $\epsilon > 0$, there is a $\delta_\alpha(\epsilon) > 0$ so that, if θ_1 and θ_2 are within $\delta_\alpha(\epsilon)$, then $\left\|\eta_{\theta_1} - \eta_{\theta_2}\right\|_\infty < \epsilon$.*

Proof: The function $\eta : \mathbb{S}^{D-1} \times \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$ is uniformly continuous. □ [Claim 2]

The proof now proceeds as before: if θ_1 and θ_2 are within $\delta_\alpha(\epsilon)$, then

$$\begin{aligned} |\mathbf{g}(\theta_1) - \mathbf{g}(\theta_2)| &\leq \int_{\mathbb{S}^{D-1}} \left\|\eta_{\theta_1} - \eta_{\theta_2}\right\|_\infty \, d\Gamma[\mathbf{s}] \\ &\leq \Gamma[\mathbb{S}^{D-1}] \cdot \epsilon \quad \text{by Claim 2.} \end{aligned}$$

□

4.4 Estimating the SLCF from Empirical data

Suppose that $\mathbf{X}_1, \ldots, \mathbf{X}_N$ are N independent samples of the random vector $\mathbf{X}$. Let $\theta \in \mathbb{S}^{D-1}$, and Suppose we estimate $\chi(\theta)$ with the empirical value

$$\chi_N(\theta) = \frac{1}{N} \sum_{n=1}^{N} \exp\left(2\pi \mathbf{i} \left\langle \mathbf{X}_n, \theta \right\rangle\right) \tag{4.2}$$

We can then estimate $\mathbf{g}(\theta)$ by the value $\mathbf{g}_N(\theta) = \log(\chi_N(\theta))$. If we do this for each θ in some suitably fine "mesh" on $\mathbb{S}^{D-1}$, then, since $\mathbf{g}$ is relatively "smooth", we can linearly interpolate the values of $\mathbf{g}$ between these mesh points, to get a good empirical estimate of $\mathbf{g}$ everywhere on $\mathbb{S}^{D-1}$. Call this empirical estimator $\mathbf{g}_N$.

Suppose $\epsilon, \kappa > 0$, and we want $\|\mathbf{g}_N - \mathbf{g}\|_\infty < \epsilon$, with a probability greater than $1 - \kappa$. How big must N be?

Proposition 25: *Let $\mathbf{X}$ be a random vector in $\mathbb{R}^D$, with spectral measure Γ, and let $G = \Gamma[\mathbb{S}^{D-1}]$. Let χ be the characteristic function of $\mathbf{X}$, and let $m > 0$ be such that $\min_{\theta \in \mathbb{S}^{D-1}} |\chi(\theta)| > \frac{1}{m}$.*[3]

Let $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \ldots$, be a collection of independent random samples drawn from $\mathbf{X}$.

Let $\{\theta_1, \ldots, \theta_J\} \subset \mathbb{S}^{D-1}$. Fix $N \in \mathbb{N}$, and, for each $j \in [1..J]$, define

$$\mathbf{g}_N(\theta_j) \;=\; \log\left[\frac{1}{N}\sum_{n=1}^{N} \exp\left(2\pi \mathbf{i}\,\langle \mathbf{X}_n, \theta_j\rangle\right)\right] \tag{4.3}$$

and then define $\mathbf{g}_N : \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$ to be peicewise constant: for any $\mathbf{s} \in \mathbb{S}^{D-1}$, $\mathbf{g}_N(\mathbf{s}) = \mathbf{g}_N(\theta_j)$, where θ_j is the element of Θ closest to $\mathbf{s}$.

Let $\epsilon < \frac{3}{2} m\sqrt{2}$ and let $\kappa > 0$ be "small".

1. *If $N > \frac{40m^2}{\epsilon^2} \log\left[\frac{4J}{\kappa}\right]$, then with probability greater than $1 - \kappa$:*

$$\forall j \in [1..J], \quad |\mathbf{g}_N(\theta_j) - \mathbf{g}(\theta_j)| < \frac{\epsilon}{2} \tag{4.4}$$

2. *Define*

$$\Upsilon = \begin{cases} \dfrac{\epsilon}{2G \cdot C_\alpha} & \text{if } \alpha > 1 \\ \\ \delta_\alpha\left(\dfrac{\epsilon}{2G}\right) & \text{if } \alpha \leq 1 \end{cases},$$

where $\delta_\alpha : [0,\infty) \longrightarrow [0,\infty)$ and C_α are as in Proposition 24 on page 45.

If (4.4) holds, and $\{\theta_1, \ldots, \theta_J\}$ is Υ-dense in $\mathbb{S}^{D-1}$, then $\|\mathbf{g}_N - \mathbf{g}\|_\infty < \epsilon$.

[3] Since χ is everywhere nonzero, and $\mathbb{S}^{D-1}$ is compact, we can always find such an $m > 0$.

3. *To get such an Υ-dense set will require* $J \sim \mathcal{O}\left[\Upsilon^{1-D}\right]$ *mesh points.*

4. *To estimate* $\mathbf{g}$ *with a uniform accuracy of* ϵ *everywhere on* $\mathbb{S}^{D-1}$*, with confidence greater than* $1-\kappa$*, it suffices to have a sample of size*

$$N \geq \begin{cases} \mathcal{O}\left[\frac{40m^2}{\epsilon^2}\log\left(\frac{2^{D+1}G^{D-1}C_\alpha^{D-1}}{\kappa\cdot\epsilon^{D-1}}\right)\right] & \text{if } \alpha > 1 \\ \mathcal{O}\left[\frac{40m^2}{\epsilon^2}\log\left(\frac{4}{\kappa\cdot\delta_\alpha\left(\frac{\epsilon}{2G}\right)^{D-1}}\right)\right] & \text{if } \alpha \leq 1 \end{cases}$$

Proof:

Proof of Part 1: We assume that $|\chi(\theta_j)| \geq \frac{1}{m}$. Recall that, if $x \geq \frac{1}{m}$, then $\frac{d}{dx}\log(x) \leq m$. Thus, the error in the estimation of $\log(x)$ is always smaller than m times the error in the estimation of x.

Thus, the problem of estimating $\Phi_{\mathbf{X}}(\theta_j)$ with accuracy $\frac{\epsilon}{2}$ is reduced to the problem of estimating $\chi(\theta_j)$ with accuracy $\frac{\epsilon}{2m}$. Since it ranges over the unit circle, the random variable $\exp\left(2\pi\mathbf{i}\left\langle\mathbf{X},\theta\right\rangle\right)$ has variance less than 2 always. Thus, the random variable $\chi_N(\theta_j)$ has variance less than $\frac{2}{N}$, and mean $\chi(\theta_j)$, and has magnitude bounded by $1/N$. At this point, we employ:

Bernstein's Inequality[4]**:** *If* $\widetilde{\mathbf{y}}_1,\ldots,\widetilde{\mathbf{y}}_N$ *are independent real random variables with range* $[-R,R]$ *and zero mean, and* $\mathbf{Y}_N := \sum_{n=1}^{N}\widetilde{\mathbf{y}}_n$ *has variance* σ*, then for any* $\Upsilon > 0$, $\mathbf{Prob}\left[|\mathbf{Y}_N| > \Upsilon\right] \leq 2\cdot\exp\left[\frac{-\Upsilon^2}{2\cdot\left(\sigma+\frac{R\cdot\Upsilon}{3}\right)}\right]$ □[175]

In our case, $\widetilde{\mathbf{y}}_n = \frac{1}{N}\mathbf{re}\left[\widetilde{\chi}_n\right]$ or $\frac{1}{N}\mathbf{im}\left[\widetilde{\chi}_n\right]$, where $\widetilde{\chi}_n = \exp\left(2\pi\mathbf{i}\left\langle\mathbf{X}_n,\theta\right\rangle\right)$. Thus, $R=\frac{2}{N}$, $\sigma=\frac{2}{N}$, $\Upsilon=\frac{\epsilon}{2m\sqrt{2}}$, and the theorem says

$$\mathbf{Prob}\left[\left|\mathbf{re}\left[\chi_N(\theta_j)-\chi(\theta_j)\right]\right| > \frac{\epsilon}{2m\sqrt{2}}\right] \leq 2\cdot\exp\left[\frac{-N\cdot\left(\frac{\epsilon}{2m\sqrt{2}}\right)^2}{4+\frac{2\epsilon}{3m\sqrt{2}}}\right] < 2\cdot\exp\left(\frac{-N\cdot\epsilon^2}{5\cdot 8m^2}\right)$$

$$= 2\cdot\exp\left(\frac{-N\cdot\epsilon^2}{40\cdot m^2}\right), \quad (\text{since } \epsilon < \tfrac{3}{2}m\sqrt{2})$$

[4] I'd like to thank Keith Knight for suggesting this use of Bernstein's Inequality.

and likewise for $\mathbf{im}\left[\chi_N(\theta_j) - \chi(\theta_j)\right]$. Thus,

$$\begin{aligned}\mathbf{P}_{rob}\left[\exists j \in [1..J],\ \ |\chi_N(\theta_j) - \chi(\theta_j)| > \frac{\epsilon}{2m}\right] &\leq \sum_{j=1}^{J} \mathbf{P}_{rob}\left[|\chi_N(\theta_j) - \chi(\theta_{<j})| > \frac{\epsilon}{2m}\right] \\ &< 4J \cdot \exp\left(\frac{-N \cdot \epsilon^2}{40m^2}\right).\end{aligned}$$

Hence, if we want the probability of failure to be less than κ, then it suffices to have $N > \frac{-40m^2}{\epsilon^2} \log\left[\frac{\kappa}{4J}\right] = \frac{40m^2}{\epsilon^2} \log\left[\frac{4J}{\kappa}\right]$.

Proof of Part 2: Let $\mathbf{s} \in \mathbb{S}^{D-1}$ be arbitrary, and find an element θ_j so that $\mathbf{dist}\left[\mathbf{s}, \theta_j\right] < \Upsilon$. If $\alpha > 1$ (resp. $\alpha \leq 1$), apply the local Lipschitz property (resp. continuity) of $\mathbf{g}$ described by Proposition 24 to conclude that $|\mathbf{g}(\mathbf{s}) - \mathbf{g}(\theta_j)| < \frac{\epsilon}{2}$. Thus,

$$\begin{aligned}|\mathbf{g}(\mathbf{s}) - \mathbf{g}_N(\mathbf{s})| &\leq |\mathbf{g}(\mathbf{s}) - \mathbf{g}(\theta_j)| + |\mathbf{g}(\theta_j) - \mathbf{g}_N(\mathbf{s})| \\ &\leq \frac{\epsilon}{2} + |\mathbf{g}(\theta_j) - \mathbf{g}_N(\theta_j)| \quad (\mathbf{g}_N(\mathbf{s}) = \mathbf{g}_N(\theta) \text{ by construction}) \\ &\leq \epsilon \qquad (\text{with probability} > 1 - \kappa, \text{ by } \mathbf{Part\ 1}).\end{aligned}$$

Proof of Part 3: This is because $\mathbb{S}^{D-1}$ is $(D-1)$-dimensional.

Proof of Part 4: This follows from combining **Part 1** through **Part 3**.

□

Chapter 5

Estimating spectral measures via Convolutions on Spheres

In this chapter, I develop a method of recovering the spectral measure, Γ, of a stable distribution from its empirical characteristic function by developing a notion of "*convolution*" on spheres, and interpreting the spherical log-characteristic function (section 4.1) as a result of *convolving* Γ with a certain kernel. Recovering Γ from $\mathbf{g}$ then consists of "undoing" this convolution, which can be accomplished through a generalization to spheres of methods from classical harmonic analysis on compact abelian groups.

5.1 Eigenspaces of the Laplacian on Homogeneous Riemannian Manifolds $(*)$

[$(*)$ The discussion of background theory given here loosely follows the theoretical development in [171] chapter 3, section 3. The results in this section are not original, but the proofs presented here are my own.]

Let $\mathcal{M}$ be a compact Riemannian manifold, and let $\mathbb{G}$ be a compact Lie group acting transitively and isometrically on $\mathcal{M}$. For example, consider $\mathcal{M} = \mathbb{S}^{D-1}$, and $\mathbb{G} = \mathbb{SO}^D[\mathbb{R}]$. In fact, for our applications, $\mathbb{S}^{D-1}$ and $\mathbb{SO}^D[\mathbb{R}]$ is actually the only pair we are interested in. It is interesting that much of the following theory can be developed in a more general setting, but the reader may find it helpful to read "$\mathcal{M}$" as "$\mathbb{S}^{D-1}$", and "$\mathbb{G}$" as "$\mathbb{SO}^D[\mathbb{R}]$" throughout what follows.

Let $\mathcal{L}^{lsg}$ be the canonical volume measure[1] induced on $\mathcal{M}$ by its Riemann

[1] see Definition 141 on page 221.

structure. For example, if $\mathcal{M} = \mathbb{S}^{D-1}$, then $\mathcal{L}^{bsg}$ is the usual "surface area" measure on $\mathbb{S}^{D-1}$. $\mathcal{M}$ is compact, so $\mathcal{L}^{bsg}$ is finite —assume $\mathcal{L}^{bsg}$ is normalized to have total mass 1. Let $\mathbf{L}^2(\mathcal{M}) = \mathbf{L}^2(\mathcal{M}, \mathcal{L}^{bsg}; \mathbb{C})$. The action of $\mathbb{G}$ on $\mathcal{M}$ induces a linear $\mathbb{G}$-action on $\mathbf{L}^2(\mathcal{M})$ in the obvious way: if $\phi \in \mathbf{L}^2(\mathcal{M})$ and $g \in \mathbb{G}$, then $g.\phi : \mathcal{M} \longrightarrow \mathbb{C}$ is defined: $g.\phi(m) = \phi(g.m)$.

Let $\mathcal{C}^\infty(\mathcal{M})$ be the space of smooth, complex-valued functions on $\mathcal{M}$. Now, $\mathcal{L}^{bsg}$ is finite, therefore $\mathcal{C}^\infty(\mathcal{M})$ is a linear subspace of $\mathbf{L}^2(\mathcal{M})$ (though not a closed subspace). $\mathbb{G}$ acts smoothly on $\mathcal{M}$, so $\mathcal{C}^\infty(\mathcal{M})$ is $\mathbb{G}$-invariant. We consider the restricted action of $\mathbb{G}$ on $\mathcal{C}^\infty(\mathcal{M})$.

If $\triangle : \mathcal{C}^\infty(\mathcal{M}) \longrightarrow \mathcal{C}^\infty(\mathcal{M})$ is the Laplacian operator[2] (relative to the given Riemannian structure), then the $\triangle$ commutes with the isometric $\mathbb{G}$ action: for all $g \in \mathbb{G}$,

$$\triangle(g.\phi) = g.(\triangle\phi)$$

Let $\Lambda = \{\lambda \in \mathbb{C} \; ; \; -\lambda \text{ is an eigenvalue of } \triangle\}$, and for each $\lambda \in \Lambda$, let $\mathbb{V}_\lambda = \mathbb{V}_\lambda(\mathcal{M}) = \{\phi \in \mathcal{C}^\infty(\mathcal{M}) \; ; \; \triangle\phi = -\lambda\phi\}$ be the corresponding eigenspace. Thus, $\mathbb{V}_\lambda$ is a $\mathbb{G}$-invariant subspace.

Fix $e \in \mathcal{M}$, and define

$$\mathbb{G}_e = \{g \in \mathbb{G} \; ; \; g.e = e\},$$

a compact subgroup of $\mathbb{G}$. The action of $\mathbb{G}$ upon $\mathcal{C}^\infty(\mathcal{M})$ restricts to an action of $\mathbb{G}_e$, and the spaces $\mathbb{V}_\lambda$ remain invariant under this new action.

Example 26: *(The action of $\mathbb{SO}^D[\mathbb{R}]$ on $\mathbb{S}^{D-1}$)*

Suppose $\mathcal{M} = \mathbb{S}^{D-1}$, $\mathbb{G} = \mathbb{SO}^D[\mathbb{R}]$, and $e = \mathbf{e}_1 = (1, 0, \ldots, 0)$. Then $\mathbb{SO}^D[\mathbb{R}]_{\mathbf{e}_1}$ is the set of all orthogonal transformations of $\mathbb{R}^D$ fixing the $\mathbf{e}_1$-axis. In other words, it is the set of all "rotations" of the remaining $(D-1)$ dimensions about this axis; hence, there is a natural isomorphism $\mathbb{SO}^D[\mathbb{R}]_{\mathbf{e}_1} \cong \mathbb{SO}^{D-1}[\mathbb{R}]$. ———————

Definition 27: *Zonal function*

A function $\zeta \in \mathcal{C}^\infty(\mathcal{M})$ is called **zonal** *(relative to $\mathbb{G}$ and the fixed point $e \in \mathcal{M}$) if it is invariant under the action of $\mathbb{G}_e$. Formally, for any $\mathbb{G}_e$-invariant subspace $\mathbb{V} \subset \mathcal{C}^\infty(\mathcal{M})$, define*

$$\mathcal{Z}_e(\mathbb{V}) := \{\zeta \in \mathbb{V} \; ; \; \forall g \in \mathbb{G}_e, \; g.\zeta = \zeta\}$$

[2] See Definition 149 on page 225; for the concrete formulae of Laplacians on spheres of various dimensions, see example 39 on page 62.

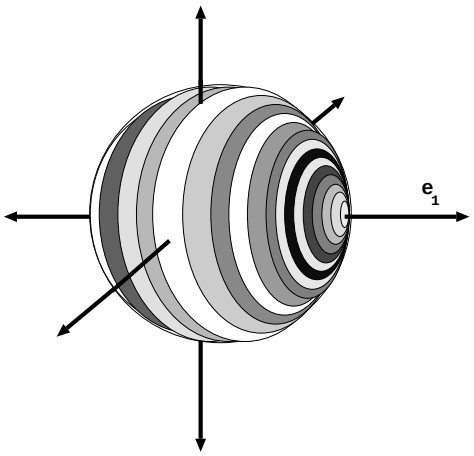

Figure 5.1: A zonal function on $\mathbb{S}^{D-1}$

Example 28: *(Zonal functions on $\mathbb{S}^{D-1}$)*

In the case of $\mathbb{S}^{D-1}$, $\mathbb{SO}^D$ $[\mathbb{R}]$, and $\mathbf{e}_1$, the zonal elements of $\mathcal{C}^\infty(\mathbb{S}^{D-1})$ are smooth functions rotationally invariant about the $\mathbf{e}_1$ axis. Clearly, any such function must be of the form

$$\zeta(\mathbf{x}) \;=\; \zeta_1(x_1)$$

where $\zeta_1 : [-1,1] \longrightarrow \mathbb{C}$, and where $\mathbf{x} = (x_1, x_2, \ldots, x_D)$ is any element of $\mathbb{S}^{D-1}$ (see Figure 5.1). ———————————————

Proposition 29:

1. *If $\mathbb{V} \subset \mathcal{C}(\mathcal{M})$ is a nontrivial $\mathbb{G}$-invariant subspace, then $\mathcal{Z}_e(\mathbb{V})$ is nontrivial.*
2. *If $\dim(\mathcal{Z}_e(\mathbb{V})) \;=\; 1$, then $\mathbb{V}$ is an irreducible $\mathbb{G}$-module*[3].

Proof:

Proof of Part 1:

Claim 1: *$\mathbb{V}$ contains an element ϕ such that $\phi(e) \neq 0$.*

Proof: Since $\mathbb{V}$ is nontrivial, there is some $\psi \in \mathbb{V}$ which is nonzero *somewhere* –say $\psi(x) \neq 0$. Since $\mathbb{G}$ acts transitively on $\mathcal{M}$, find $g \in \mathbb{G}$

[3]See Definition 152 on page 227.

so that $g.e = x$. Thus, if $\phi = g.\psi$, then $\phi(e) = \psi(g.e) = \psi(x) \neq 0$. Since $\mathbb{V}$ is $\mathbb{G}$-invariant, $\phi \in \mathbb{V}$ is the element we seek. □ [Claim 1]

Now, $\mathbb{G}_e$ is a closed subgroup of the compact group $\mathbb{G}$; thus, $\mathbb{G}_e$ is compact, so it has a finite Haar measure $\mathcal{H}^{haar}_{\mathbb{G}_e}$. Define

$$\zeta := \int_{\mathbb{G}_e} g.\phi \; d\mathcal{H}^{haar}_{\mathbb{G}_e}[g]$$

Since $\mathcal{H}^{haar}_{\mathbb{G}_e}$ is finite, this integral is well-defined. Since $\mathbb{V}$ is a closed, $\mathbb{G}$-invariant subspace, the element ζ is in $\mathbb{V}$. Furthermore, since $\zeta(e) = \phi(e)$, and $\phi(e) \neq 0$, we conclude that ζ is nontrivial. Finally, note that ζ is $\mathbb{G}_e$-invariant by construction —in other words, it is zonal.

Proof of Part 2: Suppose $\mathbb{V} = \mathbb{V}_1 \oplus \mathbb{V}_2$, where $\mathbb{V}_1, \mathbb{V}_2$ are $\mathbb{G}$-invariant. Then by **Part 1**, we can construct linearly independent zonal functions $\zeta_1 \in \mathcal{Z}_e(\mathbb{V}_1)$ and $\zeta_2 \in \mathcal{Z}_e(\mathbb{V}_2)$. Since $\zeta_1, \zeta_2 \in \mathcal{Z}_e(\mathbb{V})$, this contradicts the hypothesis that $\dim(\mathcal{Z}_e(\mathbb{V})) = 1$. □

Since $\mathbb{G}_e$ fixes e, the isometric action of $\mathbb{G}_e$ on $\mathcal{M}$ induces a linear, isometric action upon $\mathsf{T}_e\,\mathcal{M}$. If $\vec{v} \in \mathsf{T}_e\,\mathcal{M}$ is the derivative of a path $\gamma : (-\epsilon, \epsilon) \longrightarrow \mathcal{M}$ with $\gamma(0) = e$, then $g.\vec{v}$ is the derivative of the path $(g.\gamma) : (-\epsilon, \epsilon) \longrightarrow \mathcal{M}$.

Definition 30: *Rank One Action*

The manifold $\mathcal{M}$ is of **rank one** *(relative to $\mathbb{G}$ and e) if $\mathbb{G}_e$ acts transitively on the set of unit tangent vectors $\mathsf{T}_e\,\mathcal{M}$.*

For example, it is clear that $\mathbb{S}^{D-1}$ is rank one, relative to $\mathbb{SO}^D[\mathbb{R}]$ and $\mathbf{e}_1$.

The following lemmas are obvious in the case of $\mathbb{S}^{D-1}$, but we include a general proof.

Lemma 31: *For all r, $\mathbb{G}_e$ acts transitively on $\partial\mathbb{B}(e, r)$ in $\mathcal{M}$.*

Proof: The Riemann structure on $\mathcal{M}$ allows us to define an **exponential map**[4] $\exp_e : \mathsf{T}_e\,\mathcal{M} \longrightarrow \mathcal{M}$, where, for any $\vec{v} \in \mathsf{T}_e\,\mathcal{M}$, $\exp_e(\vec{v}) = \gamma_{\vec{v}}(1)$, where $\gamma_{\vec{v}}$ is the unique geodesic passing through e with derivative $\vec{v}$.

Claim 1: *The action of $\mathbb{G}$ commutes with the exponential map: for any $g \in \mathbb{G}$ and $\vec{v} \in \mathsf{T}_e\,\mathcal{M}$,* $\exp_e(g.\vec{v}) = g.\exp_e(\vec{v})$.

[4] see Definition 136 on page 220.

Proof: If $\gamma : (-\epsilon, \epsilon) \longrightarrow \mathcal{M}$ is a geodesic with $\gamma(0) = e$, then for all $g \in \mathbb{G}_e$, the function $g.\gamma : (-\epsilon, \epsilon) \longrightarrow \mathcal{M}$ is also a geodesic (since $\mathbb{G}_e$ acts isometrically), and $g.\gamma(0) = e$ (since $\mathbb{G}_e$ fixes e). Thus, $\gamma'(0)$ and $(g.\gamma)'(0)$ are both in $\mathsf{T}_e\,\mathcal{M}$, and $(g.\gamma)'(0) = g.[\gamma'(0)]$. It follows that, for any $\vec{v} \in \mathsf{T}_e\,\mathcal{M}$, $g.\gamma_{\vec{v}} = \gamma_{(g.\vec{v})}$. Thus, $\exp_e(g.\vec{v}) = \gamma_{(g.\vec{v})}(1) = g.\gamma_{\vec{v}}(1) = g.\exp_e(\vec{v})$. ..... □ [Claim 1]

Now, note that, for any $\vec{v} \in \mathcal{U}$, $\mathbf{dist}\,[\exp_e(\gamma_{\vec{v}}),\ e] = \|\vec{v}\|$. (see Proposition 137 on page 220). By the Hopf-Rinow theorem[5], the exponential map $\exp_e : \mathsf{T}_e\,\mathcal{M} \longrightarrow \mathcal{M}$ is surjective. Thus, for every $R > 0$, if $\mathsf{B}_e\,[0, R]$ is the ball of radius R about 0 in $\mathsf{T}_e\,\mathcal{M}$, then

$$\partial\mathbb{B}\,(e, R) \;=\; \exp_e\,[\partial\mathsf{B}_e\,[0, R]]$$

Now, $\mathbb{G}_e$ acts transitively on $\partial\mathsf{B}_e\,[0, 1]$ in $\mathsf{T}_e\,\mathcal{M}$, and thus, transitively on $\partial\mathsf{B}_e\,[0, R]$ for every $R > 0$. Thus, by commuting with $\exp_e$, we conclude that $\mathbb{G}_e$ must act transitively on $\partial\mathbb{B}\,(e, R)$ in $\mathcal{M}$. ——————□

Corollary 32: *If $\zeta \in \mathcal{Z}_e$, then $\zeta(u)$ is a function only of the distance from u to e.*

Proof: If ζ is zonal (thus, $\mathbb{G}_e$-invariant), then Lemma 31 on the facing page implies ζ must be constant on $\partial\mathbb{B}\,(e, R)$ for every $R > 0$ —in other words, $\zeta(u)$ depends only on the distance between u and e. ————□

Proposition 33: *If $\mathcal{M}$ is of rank one, then each eigenspace $\mathbb{V}_\lambda$ of $\triangle$ is an* **irreducible** *$\mathbb{G}$-module.*

Proof: By Proposition 29 on page 53, it suffices to show that $\dim[\mathcal{Z}_e(\mathbb{V}_\lambda)] = 1$. So, suppose that $\zeta_1, \zeta_2 \in \mathcal{Z}_e(\mathbb{V}_\lambda)$ are linearly independent. By Corollary 32, $\zeta_1(u)$ and $\zeta_2(u)$ are functions only of the distance from u to e. So, for some $u \in \mathcal{M}$ with $\mathbf{dist}\,[u, e] = r$, define $z_1 := \zeta_1(u)$ and $z_2 := \zeta_2(u)$, and let $\zeta := z_2\zeta_1 - z_1\zeta_2$. Thus, ζ is also zonal. We aim to show that ζ is the zero function; thus, ζ_1 and ζ_2 are just scalar multiples of one another.

Claim 1: *ζ is a $(-\lambda)$-eigenfunctions of $\triangle$.*

Proof: ζ is a linear combination of two elements of $\mathbb{V}_\lambda$; hence, it is also in $\mathbb{V}_\lambda$. ..... □ [Claim 1]

[5] Theorem 139 on page 220.

Now, by construction, $\zeta(u) = 0$, and thus, $\zeta \equiv 0$ on $\partial\mathbb{B}^{\mathcal{M}}(e;r)$. Let $\mathcal{C}_0^\infty(\mathbb{B}(u;r)) := \left\{ f \in \mathcal{C}^\infty(\mathbb{B}(u;r)) \ ; \ f_{|\partial\mathbb{B}(u,r)} \equiv 0 \right\}$. Hence, $\zeta \in \mathcal{C}_0^\infty(\mathbb{B}(u;r))$. Now, for any $r > 0$, let λ_r be the smallest nontrivial element of $\Lambda(\mathbb{B}(u;r))$ so that $\mathbb{V}_{(\lambda_r)} \cap \mathcal{C}_0^\infty(\mathbb{B}(u;r)) \neq \emptyset$ —in other words, the smallest eigenvalue of $\triangle_{\mathbb{B}(u;r)}$ admitting the Dirichlet boundary condition $f_{|\partial\mathbb{B}(u,r)} \equiv 0$.

Claim 2: *$\lambda_r \to \infty$ as $r \to 0$.*

Proof: (Sketch) We'll prove this when $\mathcal{M} = \mathbb{R}^D$ and $u = 0$. In this case, there is an isometry $\phi : x \mapsto 2x$ between $\mathbb{B}(0;\ r)$ and $\mathbb{B}(0;\ 2r)$, inducing an isometry $\phi_* : \mathbf{L}^2(\mathbb{B}(0;\ 2r)) \longrightarrow \mathbf{L}^2(\mathbb{B}(0;\ r))$, so that $f \mapsto f \circ \phi$. Clearly, ϕ_* restricts to a map from $\mathcal{C}_0^\infty(\mathbb{B}(0;\ 2r))$ to $\mathcal{C}_0^\infty(\mathbb{B}(0;\ r))$, and for any $f \in \mathcal{C}_0^\infty(\mathbb{B}(0;\ 2r))$, we have $\triangle(\phi_* f) = 4 \cdot \phi_*(\triangle f)$. Thus, every $\triangle$-eigenfunction of $\mathcal{C}_0^\infty(\mathbb{B}(0;\ 2r))$ maps to a $\triangle$-eigenfunction of $\mathcal{C}_0^\infty(\mathbb{B}(0;\ r))$ whose eigenvalue is 4 times the size. Thus, in particular, $\lambda_r = 4 \cdot \lambda_{2r}$. Thus, as $r \to 0$, λ_r increases to infinity with order $1/r^2$. The general case for $\mathbb{B}(u;r) \subset \mathcal{M}$ now follows by approximating the manifold locally by $\mathsf{T}_u\mathcal{M}$. ..... □ [Claim 2]

So, fix λ, and let r get small. If r is made small enough, then, by **Claim 2**, the Dirichlet boundary condition $\zeta_{|\partial\mathbb{B}^{\mathcal{M}}(e;r)} \equiv 0$ forces the smallest eigenvalue of $\triangle$ to be larger in absolute value than λ, contradicting **Claim 1**. □

Example: (Zonal Eigenfunctions on $\mathbb{S}^{D-1}$)

Explicit formulae for the zonal eigenfunctions of the Laplacian on spheres of various dimensions are given by Proposition 44 on page 68.

Now consider $\mathcal{M} = \mathbb{T}^D$, acting transitively on itself by translation, and equipped with the standard equivariant metric. The eigenfunctions of the Laplacian are the periodic functions of the form $\mathcal{E}_{\mathbf{n}}(\mathbf{x}) = \exp(2\pi\mathbf{i} \cdot \langle \mathbf{n}, \mathbf{x} \rangle)$, with $\mathbf{n} \in \widehat{\mathbb{T}^D} \cong \mathbb{Z}^D$, where $\mathbf{x} \in [0,1)^D$ and $[0,1)^D$ is identified with $\mathbb{T}^D$ in the obvious way. These eigenfunctions form an orthonormal basis for $\mathbf{L}^2(\mathbb{T}^D,\ \mathcal{H}^{oar}_{\mathbb{T}^D})$. This has the following generalization to arbitrary homogeneous Riemannian manifolds:

Theorem 34:

1. *If $\lambda_1 \neq \lambda_2$, then the eigenspaces $\mathbb{V}_{\lambda_1}$ and $\mathbb{V}_{\lambda_2}$ are orthogonal as subsets of $\mathbf{L}^2(\mathcal{M})$.*

2. *The eigenspaces of* $\triangle$ *span* $\mathbf{L}^2(\mathcal{M})$. *In other words:* $\mathbf{L}^2(\mathcal{M}) \ = \ \bigoplus_{\lambda \in \Lambda} \mathbb{V}_\lambda$.

3. *Thus, if* $f \in \mathbf{L}^2(\mathcal{M})$, *and, for all* $N \in \mathbb{N}$, *we define* $f_N := \sum_{n=1}^{N} \mathbf{pr}_{\mathbb{V}_\lambda}(f)$, *then* $\|f_N - f\|_2 \xrightarrow[N \to \infty]{} 0$.

4. *Furthermore, if* $f \in \mathcal{C}^\infty(\mathcal{M})$, *then* $\|f_N - f\|_\infty \xrightarrow[N \to \infty]{} 0$.

Proof: For **Part 1**, 2 and 3, see [179], chapter 6, p. 255; or [27], Theorem 3.21, p. 156. Or treat $\triangle$ as an elliptic differential operator, and use [40], §6.5, Theorem 1, p. 335. Alternately, employ the Spectral Theorem for unbounded self-adjoint operators (see [33], chapter X, section 4, p. 319).

For **Part 4**, see [179], chapter 6, p. 256. □

Definition 35: *Equivariant Function*

If $\eta : \mathcal{M} \times \mathcal{M} \longrightarrow \mathbb{C}$, *then say that* η *is a* $\mathbb{G}$**-equivariant** *if, for all* $m, n \in \mathcal{M}$ *and* $g \in \mathbb{G}$,

$$\eta(g.m, g.n) \ = \ \eta(m, n)$$

Since $\mathbb{G}$ *acts isometrically and transitively on* $\mathcal{M}$, *this is equivalent to saying that* $\eta(\mathbf{x}, \mathbf{y})$ *is a function only of the distance* $\mathbf{dist}\,[\mathbf{x}, \mathbf{y}]$.

Example: If the function $\eta^{(\alpha)} : \mathbb{S}^{D-1} \times \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$ is defined:

$$\eta^{(\alpha)}(\mathbf{x}, \mathbf{y}) \ = \ |\langle \mathbf{x}, \mathbf{y} \rangle|^\alpha \ + \ \mathcal{B}_\alpha \cdot \mathbf{i} \, \langle \mathbf{x}, \mathbf{y} \rangle^{\langle \alpha \rangle},$$

then $\eta^{(\alpha)}$ is $\mathbb{SO}^D\,[\mathbb{R}]$-equivariant.

$\mathbb{G}$-equivariant functions are interesting because we can define a sort of **convolution** with them.

Definition 36: *Convolution*

If η *is* $\mathbb{G}$*-equivariant,* $\phi : \mathcal{M} \longrightarrow \mathbb{C}$, *and both are* $\mathcal{L}^{hsg}$*-integrable, then define* $\eta * \phi : \mathcal{M} \longrightarrow \mathbb{C}$ *by*

$$(\eta * \phi)(m) \ = \ \int_{\mathcal{M}} \eta(m, n) \phi(n) \ d\mathcal{L}^{hsg}[n]$$

For example, if Γ is a spectral measure on $\mathbb{S}^{D-1}$, with Radon-Nikodym derivative $\gamma : \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$, then $\eta^{(\alpha)} * \gamma : \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$ is defined

$$\eta^{(\alpha)} * \gamma(\mathbf{x}) \;=\; \int_{\mathbb{S}^{D-1}} \eta^{(\alpha)}(\mathbf{x}, \theta)\gamma(\theta)\; d\mathcal{L}^{bsg}[\theta] \;=\; \int_{\mathbb{S}^{D-1}} \eta^{(\alpha)}(\mathbf{x}, \theta) d\Gamma[\theta] \;=\; \mathbf{g}(\mathbf{x})$$

where **g** is the spherical log-characteristic function introduced in section 4.1 on page 43.

Recall that, in the case of $\mathbb{T}^D$, with the standard equivariant metric, the eigenfunctions of the Laplacian are $\left\{\mathcal{E}_{\mathbf{n}} \;;\; \mathbf{n} \in \mathbb{Z}^D\right\}$. These functions are well-behaved under convolution: classical harmonic analysis tells us that

$$\left(\sum_{\mathbf{n}\in\mathbb{Z}^D} a_{\mathbf{n}}\mathcal{E}_{\mathbf{n}}(\mathbf{x})\right) * \left(\sum_{\mathbf{n}\in\mathbb{Z}^D} b_{\mathbf{n}}\mathcal{E}_{\mathbf{n}}(\mathbf{x})\right) \;=\; \sum_{\mathbf{n}\in\mathbb{Z}^D} (a_{\mathbf{n}} \cdot b_{\mathbf{n}})\, \mathcal{E}_{\mathbf{n}}(\mathbf{x})$$

It turns out that this phenomenon generalizes to arbitrary homogeneous Riemannian manifolds.

Proposition 37: *(Convolution and Eigenfunctions)*
Let $\mathcal{M}$ be a rank-one $\mathbb{G}$-space, and let $\eta : \mathcal{M} \times \mathcal{M} \longrightarrow \mathbb{C}$ be $\mathbb{G}$-equivariant. Fix $\lambda \in \Lambda$; there is a constant $A_\lambda \in \mathbb{C}$ so that, for any $\phi \in \mathbb{V}_\lambda$,

$$\eta * \phi \;=\; A_\lambda \cdot \phi.$$

To compute A_λ, let $\zeta \in \mathcal{Z}_e(\mathbb{V}_\lambda)$ with $\zeta(e) \neq 0$; then $A_\lambda \;:=\; \dfrac{(\eta * \zeta)(e)}{\zeta(e)}$.

Proof: Let $T_\eta : \mathcal{C}^\infty(\mathcal{M}) \longrightarrow \mathcal{C}^\infty(\mathcal{M})$ be defined: $T_\eta(\phi) = \eta * \phi$.

Claim 1: *The operator T_η commutes with the $\mathbb{G}$-action: for all $g \in \mathbb{G}$, $T_\eta[g.\phi] = g.T_\eta[\phi]$.*

Proof: For any $m \in \mathcal{M}$,

$$\begin{aligned} T_\eta[g.\phi](m) &= [\eta * (g.\phi)]\,(m) \\ &= \int_{\mathcal{M}} \eta(m, n)\phi(g.n)\; d\mathcal{L}^{bsg}[n] \\ &= \int_{\mathcal{M}} \eta(m, g^{-1}.n')\phi(n')\; d\mathcal{L}^{bsg}[n'] \qquad (\text{where } n' := g.n) \end{aligned}$$

$$\begin{aligned} &= \int_{\mathcal{M}} \eta(g.m, n')\phi(n')\ d\mathcal{L}^{bsg}[n'] \qquad (\eta \text{ is } \mathbb{G}\text{-equivariant}) \\ &= (\eta * \phi)(g.m) \\ &= g.(\eta * \phi)(m). \end{aligned}$$

..... □ [Claim 1]

Claim 2: *T_η commutes with $\triangle$.*

Proof: For each $y \in \mathcal{M}$, define $\eta_y : \mathcal{M} \longrightarrow \mathbb{C}$ by $\eta_y(x) = \eta(y, x) = \eta(x, y)$. Thus,

$$\begin{aligned} (\eta * \phi)(x) &= \int_{\mathcal{M}} \eta(x, y) \cdot \phi(y)\ d\mathcal{L}^{bsg}[y] \\ &= \int_{\mathcal{M}} \phi(y) \cdot \eta_y(x)\ d\mathcal{L}^{bsg}[y] \\ \text{Hence,}\quad \triangle\,(\eta * \phi)(x) &= \triangle \int_{\mathcal{M}} \phi(y) \cdot \eta_y(x)\ d\mathcal{L}^{bsg}[y] \\ &= \int_{\mathcal{M}} \phi(y) \cdot \triangle\eta_y(x)\ d\mathcal{L}^{bsg}[y] \qquad (*) \end{aligned}$$

because $\triangle$ is a differential operator (see, for example [48], Theorem 2.27, p. 54).

Claim 2.1: $\triangle\eta_y(x) = \triangle\eta_x(y)$.

Proof:

Claim 2.1.1: *There is some $g \in \mathbb{G}$ so that $g.x = y$ and $g.y = x$.*

Proof: By the Hopf-Rinow Theorem[6], there is a geodesic $\gamma : [0, 1] \longrightarrow \mathcal{M}$ so that $\gamma(0) = x$ and $\gamma(1) = y$. Now let $z = \gamma(1/2)$. Then $\mathbf{dist}\,[x, z] = R = \mathbf{dist}\,[y, z]$ for some $R > 0$. Recall that Lemma 31 on page 54 says that $\mathbb{G}_z$ acts transitively on $\mathbb{B}\,(z; R)$. Thus, there is some element $g \in \mathbb{G}_z$ so that $g.x = y$. We want to show that also, $g.y = x$.

Since $g \in \mathbb{G}_z$, we know that $g.z = z$. Also, since $\mathbb{G}$ acts isometrically, the geodesic γ must get sent to another geodesic, $g.\gamma$. But now

$$\begin{array}{ccccc} g.\gamma(0) & = & y & = & \gamma(1) \\ g.\gamma\left(\frac{1}{2}\right) & = & z & = & \gamma\left(\frac{1}{2}\right) \end{array}$$

Hence, we conclude that $g.\gamma : [0, 1] \longrightarrow \mathcal{M}$ must be the geodesic defined

$$g.\gamma(t) \ = \ \gamma(1 - t)$$

[6] Theorem 139 on page 220.

Thus,

$$g.y \;=\; g.\gamma(1) \;=\; \gamma(0) \;=\; x.$$

..... □ [Claim 2.1.1]

Let g be the element provided by this claim.

Claim 2.1.2: $\eta_x = g.\eta_y$

Proof: For any $m \in \mathcal{M}$, $\eta_x(m) = \eta(x,m) = \eta(g.x,\ g.m) = \eta(y,\ g.m) = \eta_y(g.m) = (g.\eta_y)(m)$. ..... □ [Claim 2.1.2]

Thus, $\triangle\eta_x = \triangle(g.\eta_y) = g.(\triangle\eta_y)$.

In particular, $\triangle\eta_x(y) = g.(\triangle\eta_y)(y) = \triangle\eta_y(g.y) = \triangle\eta_y(x)$.

□ [Claim 2.1]

Hence, we can rewrite expression $(*)$ as:

$$\int_{\mathcal{M}} \phi(y) \cdot \triangle\eta_x(y)\ d\mathcal{L}^{bsg}[y]$$

But $\mathcal{M}$ is a manifold without boundary, so $\triangle$ is self-adjoint, by Proposition 150 on page 225. Hence,

$$\begin{aligned}\int_{\mathcal{M}} \phi(y) \cdot \triangle\eta_x(y)\ d\mathcal{L}^{bsg}[y] &= \int_{\mathcal{M}} \triangle\phi(y) \cdot \eta_x(y)\ d\mathcal{L}^{bsg}[y] \\ &= \int_{\mathcal{M}} \eta(x,y) \cdot \triangle\phi(y)\ d\mathcal{L}^{bsg}[y] \\ &= \eta * (\triangle\phi)(x)\end{aligned}$$

..... □ [Claim 2]

It follows from **Claim 2** that T_η must leave invariant all eigenspaces of $\triangle$; in other words, for all $\lambda \in \Lambda$, $\mathbb{V}_\lambda$ is invariant under T_η.

But by **Claim 1**, the restricted map

$$(T_\eta)_| : \mathbb{V}_\lambda \longrightarrow \mathbb{V}_\lambda$$

is then an isomorphism of linear $\mathbb{G}$-modules. Since $\mathbb{G}$ acts **irreducibly** on $\mathbb{V}_\lambda$ (by Proposition 33 on page 55), it follows from Schur's Lemma[7] that T_η must act on $\mathbb{V}_\lambda$ by scalar multiplication: thus, there is some $A_\lambda \in \mathbb{C}$ so that, for all $\phi \in \mathbb{V}_\lambda$,

$$T_\eta(\phi) \;=\; A_\lambda \cdot \phi$$

....in other words, $\eta * \phi = A_\lambda \cdot \phi$. In particular, if $\zeta \in \mathcal{Z}_e(\mathbb{V}_\lambda)$, then $\eta * \zeta = A_\lambda \cdot \zeta$; hence we must have $A_\lambda = \frac{\eta * \zeta(e)}{\zeta(e)}$. _____□

[7] Theorem 153 on page 228.

Corollary 38: *Let $\mathcal{M}$ be a rank-one $\mathbb{G}$-space, Let $\zeta \in \mathcal{Z}_e(\mathbb{V}_\lambda)$ be a zonal eigenfunction, normalized so that $\|\zeta\|_2 = 1$. Define $\mathcal{Z} : \mathcal{M} \times \mathcal{M} \longrightarrow \mathbb{C}$ by*

$$\mathcal{Z}(x, y) \;=\; \bar{\zeta}(g_x^{-1}.y)$$

where $g_x \in \mathbb{G}$ is any element so that $g_x.e = x$. Then $\mathcal{Z}$ is well-defined, independent of the choice of g_x, and is $\mathbb{G}$-equivariant. If we then define $\mathbb{P}_\lambda : \mathbf{L}^2(\mathcal{M}) \longrightarrow \mathbf{L}^2(\mathcal{M})$ by

$$\mathbb{P}_\lambda(\phi) \;=\; \zeta(e) \cdot (\mathcal{Z} * \phi)$$

then $\mathbb{P}_\lambda$ is the **orthogonal projection** *from $\mathbf{L}^2(\mathcal{M})$ onto the eigenspace $\mathbb{V}_\lambda$.*

Proof:

Proof of "Well Defined": If $g_1, g_2 \in \mathbb{G}$ so that $g_1.e = g_2.e = x$, then $g_1^{-1}.g_2.e = e$; thus, $g_1^{-1}.g_2 \in \mathbb{G}_e$.Thus, since ζ is zonal about e,

$$\zeta(g_2^{-1}.y) \;=\; \zeta(g_1^{-1}.g_2.g_2^{-1}.y) \;=\; \zeta(g_1^{-1}.y)$$

Proof of "Equivariant": Let $x, y \in \mathcal{M}$, and $h \in \mathbb{G}$. Note that we can pick $g_{(h.x)} = h.g_x$. Thus,

$$\begin{aligned}\mathcal{Z}(h.x,\ h.y) &= \bar{\zeta}(g_{(h.x)}^{-1}.h.y) \;=\; \bar{\zeta}((h.g_x)^{-1}.h.y) \;=\; \bar{\zeta}(g_x^{-1}.h^{-1}.h.y)\\ &= \bar{\zeta}(g_x^{-1}.y)\\ &= \mathcal{Z}(x,\ y).\end{aligned}$$

Proof of "Orthogonal Projection": Since $\mathbb{P}_\lambda$ is defined by a convolution integral, it is clearly a linear operator. It then suffices to show that $\mathbb{P}_\lambda$ fixes $\mathbb{V}_\lambda$, and annihilates $\mathbb{V}_\lambda^\perp$.

If $\phi \in \mathbb{V}_\lambda$, then by Proposition 37 on page 58,

$$\mathcal{Z} * \phi \;=\; \frac{(\mathcal{Z} * \zeta)(e)}{\zeta(e)} \cdot \phi \qquad \text{thus,} \quad \mathbb{P}_\lambda(\phi) \;=\; (\mathcal{Z} * \zeta)(e) \cdot \phi,$$

so it suffices to show that $(\mathcal{Z} * \zeta)(e) = 1$. But:

$$\begin{aligned}\mathcal{Z} * \zeta(e) &= \int_{\mathcal{M}} \mathcal{Z}(e, y)\zeta(y)\ d\mathcal{L}^{\text{hsg}}[y]\\ &= \int_{\mathcal{M}} \bar{\zeta}(g_e^{-1}.y) \cdot \zeta(y)\ d\mathcal{L}^{\text{hsg}}[y]\\ &= \int_{\mathcal{M}} \bar{\zeta}(y) \cdot \zeta(y)\ d\mathcal{L}^{\text{hsg}}[y] \qquad (\text{since } g_e = \mathbf{Id})\\ &= \|\zeta\|_2^2\\ &= 1 \qquad (\text{by hypothesis})\end{aligned}$$

On the other hand, if $\phi \in \mathbb{V}_\lambda^\perp$, then for all $x \in \mathcal{M}$,

$$\begin{aligned} \mathcal{Z} * \phi(x) &= \int_{\mathcal{M}} \bar{\zeta}(g_x^{-1}.y) \cdot \phi(y)\ d\mathcal{L}^{bsg}[y] \\ &= \int_{\mathcal{M}} \left(g_x^{-1}.\bar{\zeta}\right)(y) \cdot \phi(y)\ d\mathcal{L}^{bsg}[y] \\ &= \langle g_x^{-1}.\bar{\zeta},\ \ \phi \rangle \\ &= 0 \end{aligned}$$

since $g_x^{-1}.\bar{\zeta} \in \mathbb{V}_\lambda \perp \phi$. □

5.2 Spherical Harmonics, and Harmonic Polynomials on $\mathbb{S}^{D-1}$

[The development of background theory given here loosely follows the theoretical development in [171] chapter 4, section 3. Further information on spherical harmonics can be found in chapter II of [13]; chapters 3 and 5 of [59]; chapters 7 and 8 of [96]; §11 and §12 of [168]; and also in [144, 146, 57, 2, 25, 165, 169], and [176].

Theorems with a (∗) are known results, although the proofs presented here are original. All other results are original.]

The material developed so far suggests that identifying the eigenspaces of the Laplacian can greatly simplify computation of (de)convolutions. We will now apply this to convolution on spheres.

Let $\triangle_{\mathbb{R}^D}$ denote the Laplacian on $\mathbb{R}^D$, and $\triangle_{\mathbb{S}^{D-1}}$ the Laplacian on $\mathbb{S}^{D-1}$, Let $\mathbb{R}^D$ have spherical coordinates (r, θ), where $r \in [0, \infty)$, and $\theta \in \mathbb{S}^{D-1}$.

Example 39: (∗) *(Laplacians on Spheres)*

Consider the case $D = 2$. Endow the circle $\mathbb{S}^1$ with the angular coordinate system $\theta \in (0, 2\pi)$, so that any point on $\mathbb{S}^1_* = \mathbb{S}^1 - \{(1, 0)\}$ has coordinates

$$(\cos(\theta),\ \sin(\theta))$$

If $f : \mathbb{S}^1_* \longrightarrow \mathbb{C}$, then, relative to this coordinate system, we have:

$$\triangle_{\mathbb{S}^1} f = \frac{\partial^2 f}{\partial \theta^2}.$$

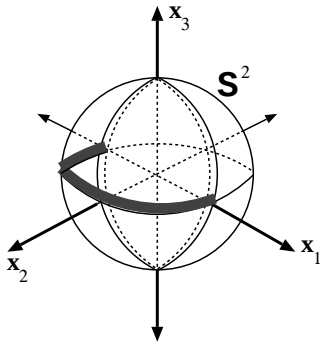

Figure 5.2: $\mathbb{S}^2_* = \mathbb{S}^2 \setminus (\mathbb{R} \times [0,\infty) \times \{0\})$ (shaded area indicates the removed slice)

Now consider the case $D = 3$. Endow the sphere $\mathbb{S}^2$ with the polar coordinate system $(\theta, \phi) \in (0, 2\pi) \times (0, \pi)$, so that any point on $\mathbb{S}^2_* = \mathbb{S}^2 \setminus (\mathbb{R} \times [0,\infty) \times \{0\})$ (see Figure 5.2) has coordinates[8]:

$$[\cos(\phi), \quad \sin(\phi)\cos(\theta), \quad \sin(\phi)\sin(\theta)]$$

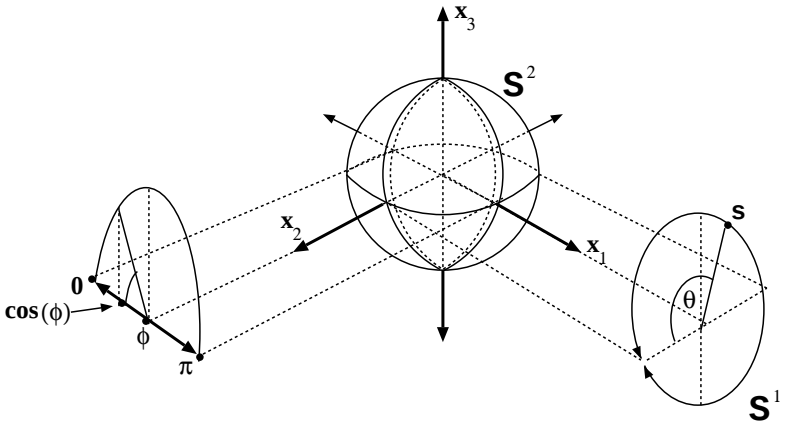

Figure 5.3: The coordinate system $[\cos(\phi), \quad \sin(\phi)\cos(\theta), \quad \sin(\phi)\sin(\theta)] = [\cos(\phi), \quad \sin(\phi) \cdot \mathbf{s}]$

[8]If $\mathbb{S}^2$ is the globe, then the x_1 axis passes through the North Pole, and these are the **Mercator coordinates:** ϕ is **latitude** and θ is **longitude**.

If $f : \mathbb{S}^2_* \longrightarrow \mathbb{C}$, then, relative to this coordinate system, we have:

$$\triangle_{\mathbb{S}^2} f \;=\; \frac{\partial^2 f}{\partial \phi^2} \;+\; \cot(\phi)\frac{\partial f}{\partial \phi} \;+\; \frac{1}{\sin(\phi)^2}\frac{\partial^2 f}{\partial \theta^2}$$

(see, for example, [13], p. 87)

If we reinterpret the (θ, ϕ) coordinate system as diffeomorphism $\mathbb{S}^1_* \times (0, \pi) \;\cong\; \mathbb{S}^2_*$ (see Figure 5.3), then we can rewrite this equation:

$$\triangle_{\mathbb{S}^2} f \;=\; \frac{\partial^2 f}{\partial \phi^2} \;+\; \cot(\phi)\frac{\partial f}{\partial \phi} \;+\; \frac{1}{\sin(\phi)^2}\,\triangle_{\mathbb{S}^1}\, f.$$

More generally, define $\mathbb{S}^D_* \;=\; \mathbb{S}^D \backslash \big(\mathbb{R}^{D-1} \times [0, \infty) \times \{0\}\big)$, and then define the diffeomorphism

$$\begin{array}{rcl} \mathbb{S}^{D-1}_* \times (0, \pi) & \longrightarrow & \mathbb{S}^D_* \\ (\mathbf{s}, \phi) & \mapsto & [\cos(\phi);\;\; \sin(\phi) \cdot \mathbf{s}] \end{array}$$

Then we have the following inductive formula:

$$\triangle_{\mathbb{S}^D} f \;=\; \frac{\partial^2 f}{\partial \phi^2} \;+\; (D-1)\cot(\phi)\frac{\partial f}{\partial \phi} \;+\; \frac{1}{\sin(\phi)^2}\,\triangle_{\mathbb{S}^{D-1}}\, f. \qquad (5.1)$$

(see [168], p. 212) ——————————————————————□

If $F : \mathbb{R}^D \longrightarrow \mathbb{C}$ is such that $\triangle F \equiv 0$, then F is called **harmonic**.

The eigenfunctions of the Laplacian on the sphere $\mathbb{S}^{D-1}$ are called **spherical harmonics**. We will now attempt to characterize these eigenfunctions.

Proposition 40: $(*)$ *Relative to the aforementioned spherical coordinate system on* $\mathbb{R}^D$,

$$\triangle_{\mathbb{R}^D} \;=\; \frac{\partial^2}{\partial r^2} \;+\; \frac{D-1}{r}\frac{\partial}{\partial r} \;+\; \frac{1}{r^2}\,\triangle_{\mathbb{S}^{D-1}}.$$

Proof: This is a straightforward computation. ——————————————————————□

Now, $\triangle_{\mathbb{S}^{D-1}}$ is an self-adjoint operator (Proposition 150 on page 225), thus, normal. Hence, employing the Spectral Theorem for unbounded normal operators and the associated functional calculus (see, for example [33], chapter **X**, §4), we can well-define the operator

$$\left(\frac{(D-2)^2}{4} - \triangle_{\mathbb{S}^{D-1}}\right)^{1/2}.$$

Corollary 41: $(*)$ *Suppose that* $f : \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$ *is smooth, and define* $F : \mathbb{R}^D \longrightarrow \mathbb{C}$ *by:*

$$F(e^t,\ \theta) \ = \ \exp\left[t \cdot \left(\frac{2-D}{2} + \left(\frac{(D-2)^2}{4} - \triangle_{\mathbb{S}^{D-1}}\right)^{1/2}\right)\right](f)(\theta).$$

Then F *is harmonic, and* $F_{|\mathbb{S}^{D-1}} \equiv f$.

Proof: This follows from Proposition 40. $\square$

Corollary 42: $(*)$ *Let* $-\lambda$ *be an eigenvalue of* $\triangle_{\mathbb{S}^{D-1}}$, *and define:*

$$N_\lambda \ = \ \left[\frac{2-D}{2} + \left(\frac{(D-2)^2}{4} + \lambda\right)^{1/2}\right] \tag{5.2}$$

1. *If* $F : \mathbb{R}^D \longrightarrow \mathbb{C}$ *is harmonic and a homogeneous polynomial of degree* N_λ, *then* $f := F_{|\mathbb{S}^{D-1}}$ *is an eigenfunction of* $\triangle_{\mathbb{S}^{D-1}}$, *with eigenvalue* $-\lambda$.

2. *Conversely, if* $f \in \mathcal{C}^\infty(\mathbb{S}^{D-1})$ *is an eigenfunction of* $\triangle_{\mathbb{S}^{D-1}}$ *with eigenvalue* $-\lambda$, *and* $F : \mathbb{R}^D \longrightarrow \mathbb{C}$ *is defined as in Corollary 41, then* F *is a* **homogeneous polynomial** *of degree* N_λ, *and is a harmonic function satisfying Dirichlet boundary condition* $F_{|\mathbb{S}^{D-1}} \equiv f$.

3. N_λ *must be an integer. Thus, the only allowable eigenvalues of* $\triangle_{\mathbb{S}^{D-1}}$ *are those* $-\lambda$ *such that* $\left(\frac{(D-2)^2}{4} + \lambda\right)^{\frac{1}{2}}$ *is an integer (if* D *is* even*) or* half-*integer (if* D *is* odd*). In particular, this means*

   - *if* D *is even, then all eigenvalues of* $\triangle_{\mathbb{S}^{D-1}}$ *are integers.*
   - *if* D *is odd, then eigenvalues of* $\triangle_{\mathbb{S}^{D-1}}$ *are half-integers.*

Proof:

Proof of Part 1: If N_λ is defined in the manner shown, then it is a straightforward computation to show that

$$N_\lambda \cdot (N_\lambda + D - 2) \ = \ \lambda. \tag{5.3}$$

Now, if F is homogeneous of degree N_λ, then F has the form

$$F(r, \theta) \ = \ r^{N_\lambda} \cdot f(\theta)$$

for some $f : \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$. But if F is harmonic, then

$$\begin{aligned} 0 &= \triangle_{\mathbb{R}^D} F \\ &= \frac{\partial^2}{\partial r^2} F \quad + \quad \frac{D-1}{r} \frac{\partial}{\partial r} F \quad + \quad \frac{1}{r^2} \triangle_{\mathbb{S}^{D-1}} F \\ &= N_\lambda \cdot (N_\lambda - 1) r^{N_\lambda - 2} \cdot f(\theta) \quad + \quad \frac{D-1}{r} \cdot N_\lambda \cdot r^{N_\lambda - 1} \cdot f(\theta) \\ &\qquad + \quad \frac{1}{r^2} r^{N_\lambda} \cdot \triangle_{\mathbb{S}^{D-1}} f(\theta) \\ &= r^{N_\lambda - 2} \cdot \left[N_\lambda (N_\lambda + D - 2) \cdot f(\theta) \quad + \quad \triangle_{\mathbb{S}^{D-1}} f(\theta) \right], \end{aligned}$$

$$\begin{aligned} \text{hence, } 0 &= N_\lambda (N_\lambda + D - 2) \cdot f(\theta) \quad + \quad \triangle_{\mathbb{S}^{D-1}} f(\theta) \\ \text{thus, } \triangle_{\mathbb{S}^{D-1}} f &= -N_\lambda (N_\lambda + D - 2) \cdot f \\ &= -\lambda \cdot f. \end{aligned}$$

Proof of Part 2: Suppose $\triangle_{\mathbb{S}^{D-1}} f = -\lambda f$. Then

$$\begin{aligned} F(r,\ \theta) &= \exp\left[\ln(r) \cdot \left(\frac{2-D}{2} + \left(\frac{(D-2)^2}{4} - \triangle_{\mathbb{S}^{D-1}} \right)^{1/2} \right) \right] (f)(\theta) \\ &= \exp\left[\ln(r) \cdot \left(\frac{2-D}{2} + \left(\frac{(D-2)^2}{4} + \lambda \right)^{1/2} \right) \right] (f)(\theta) \\ &= \exp\left[\ln(r) N_\lambda \right] \cdot f(\theta) \\ &= r^{N_\lambda} \cdot f(\theta) \end{aligned}$$

Thus, F is homogeneous of degree N_λ. It remains to show that F is actually a polynomial.

But by the previous Corollary, F is harmonic. Thus, F is real-analytic on $\mathbb{R}^D$ (see, for example, [40], §2.2, Theorem 10, p. 31). In particular, F must be real-analytic around the origin, which means the exponent N_λ must be a (non negative) integer. (This establishes **Part 3**).

Being analytic, F has a Taylor-series expansion around zero. But since F is homogeneous of degree N_λ, this means all nontrivial terms in the Taylor series have degree exactly N_λ —in other words, F is a homogeneous polynomial of degree N_λ. ——————————□

We have thus clearly identified what all the eigenvalues and corresponding eigenspaces must look like. In order for these results to be non vacuous, however, we must show that these eigenspaces are nontrivial.

Notation: Let $\mathcal{H}^N(\mathbb{R}^D)$ denote the set of all **harmonic homogeneous polynomials** on $\mathbb{R}^D$ of degree N. Let $\mathcal{H}^N(\mathbb{S}^{D-1})$ denote the set of all eigenfunctions of the Laplacian operator $\triangle_{\mathbb{S}^{D-1}}$ on $\mathbb{S}^{D-1}$, having eigenvalue λ_N, where the relationship between λ_N and N is as in formulae (5.2) and (5.3). The previous Corollary thus shows that

$$\begin{aligned} \mathbb{V}_{\lambda_N}(\mathbb{S}^{D-1}) &= \mathcal{H}^N(\mathbb{S}^{D-1}) \\ &= \left\{ F_{|\mathbb{S}^{D-1}} \; ; \; F \in \mathcal{H}^N(\mathbb{R}^D) \right\}, \\ \text{and} \;\; \mathcal{H}^N(\mathbb{R}^D) &= \left\{ F(r.\theta) = r^N \cdot f(\theta) \; ; \; f \in \mathcal{H}^N(\mathbb{S}^{D-1}) \right\}. \end{aligned}$$

We next show that these sets are nontrivial.

Proposition 43: $(*)$

Let $\mathbf{c} = (c_1, \ldots, c_D) \in \mathbb{C}^D$, *with* $\sum_{d=1}^{D} c_d^2 = 0$, *and define* $p_{\mathbf{c}}^N : \mathbb{R}^D \longrightarrow \mathbb{C}$ *by*

$$p_{\mathbf{c}}^N(\mathbf{x}) = (c_1 x_1 + c_2 x_2 + \ldots + c_D x_D)^N = \langle \mathbf{c}, \mathbf{x} \rangle^N .$$

Then:

1. *For any such* $\mathbf{c}$, *the polynomial* $p_{\mathbf{c}}^N$ *is in* $\mathcal{H}^N(\mathbb{R}^D)$.
2. $\mathcal{H}^N(\mathbb{R}^D)$ *is the linear span of all polynomials of this form.*

Proof:

Proof of Part 1: $p_{\mathbf{c}}^N$ is clearly homogeneous of degree N. It is a straightforward computation to show that $\triangle p_{\mathbf{c}}^N \equiv 0$.

Proof of Part 2: The span of the polynomials $\{p_{\mathbf{c}}^N\}$ is clearly a linear subspace of $\mathcal{H}^N(\mathbb{R}^D)$. Since $\mathbb{V}_\lambda = \mathcal{H}^{N_\lambda}(\mathbb{S}^{D-1})$, and since $\mathbb{SO}^D[\mathbb{R}]$ acts irreducibly on $\mathbb{V}_\lambda$ (by Proposition 33 on page 55), it suffices to show that this span is *invariant* under the action of $\mathbb{SO}^D[\mathbb{R}]$. But for any $g \in \mathbb{SO}^D[\mathbb{R}]$, and any $\mathbf{c} \in \mathbb{C}^D$ and $\mathbf{x} \in \mathbb{R}^D$,

$$\begin{aligned} \left(g.p_{\mathbf{c}}^N\right)(\mathbf{x}) &= p_{\mathbf{c}}^N(g.\mathbf{x}) \\ &= \langle \mathbf{c}, \; g.\mathbf{x} \rangle^N \\ &= \left\langle g^{-1}.\mathbf{c}, \; \mathbf{x} \right\rangle^N \qquad \text{(since } g \text{ is orthogonal)} \\ &= g_{(g^{-1}.\mathbf{c})}^N(\mathbf{x}) \end{aligned}$$

In other words, $g.p_{\mathbf{c}}^N = g_{(g^{-1}.\mathbf{c})}^N$; hence, the set $\{p_{\mathbf{c}}^N\}$ itself is actually invariant under the action of $\mathbb{SO}^D[\mathbb{R}]$. □

Now, Proposition 38 on page 61 provides a method for computing the orthogonal projection onto the eigenspace $\mathcal{H}^N\left(\mathbb{S}^{D-1}\right)$ by convolution with a **zonal function** in this space. Hence, to explicitly compute orthogonal projections, we need an explicit formula for the zonal functions.

We will take as our fixed point $\mathbf{e}_1 = (1, 0, \ldots, 0) \in \mathbb{R}^D$. Thus, $\mathbb{SO}^D[\mathbb{R}]_{\mathbf{e}_1}$ is the group of rotations of $\mathbb{R}^D$ about the $\mathbf{e}_1$ axis. The zonal functions (relative to $\mathbb{SO}^D[\mathbb{R}]$ and $\mathbf{e}_1$) are those which are invariant under these rotations — ie. functions which depend only upon the first coordinate:

$$\zeta(\mathbf{x}) \;=\; \zeta_1(x_1)$$

for some $\zeta_1 : [-1, 1] \longrightarrow \mathbb{C}$.

If ζ is *also* an eigenfunction of $\triangle_{\mathbb{S}^{D-1}}$, then the previous results say that ζ is the restriction to $\mathbb{S}^{D-1}$ of a harmonic homogeneous polynomial $\mathcal{Z} : \mathbb{R}^D \longrightarrow \mathbb{C}$. Of course, the fact that ζ is a function only of x_1 on the sphere does *not* mean that $\mathcal{Z}$ is a function only of x_1 everywhere on $\mathbb{R}^D$. However, it *does* mean that we should be able to write ζ as a polynomial in x_1 only, as long as we confine ourselves to the sphere. And indeed, this is the case.

Proposition 44: $(*)$ *(Zonal Eigenfunctions of $\triangle$ on $\mathbb{S}^{D-1}$)*
Let $N \in \mathbb{N}$. Let ζ_N be a corresponding eigenfunction, and assume that ζ_N is zonal (relative to $\mathbb{SO}^D[\mathbb{R}]$ and $\mathbf{e}_1$).

Case $D = 2$**:** *Modulo multiplication by some normalizing constant,*

$$\zeta_N(\theta) \;=\; \cos(N \cdot \theta)$$

where we use the coordinate system $(0, 2\pi) \ni \theta \twoheadrightarrow (\cos(\theta),\ \sin(\theta)) \in \mathbb{S}^1$. If we write ζ_N in terms of Cartesian coordinates $\mathbf{x} = (x_1, x_2)$ on $\mathbb{R}^2$, we get the **Čebyšev polynomials:**

$$\zeta_N(\mathbf{x}) \quad = \quad 2^{(N-1)}x^N \;+\; \sum_{n=1}^{\lfloor N\frac{}{2}\rfloor}(-1)^n 2^{(N-1-2n)}\frac{N}{n}\binom{N-n-1}{n-1}x_1^{(N-2n)}. \tag{5.4}$$

Case $D = 3$**:** *Modulo multiplication by some constant, ζ_N is a* **Legendre Polynomial:**

$$\zeta_N(\mathbf{x}) \quad = \quad \sum_{n=0}^{\lfloor N/2\rfloor}(-1)^n 2^{N-2n}\frac{\Gamma\left[\frac{1}{2}+N-n\right]}{\Gamma\left[\frac{1}{2}\right]\cdot n!\cdot(N-2n)!}\cdot x_1^{N-2n}$$

General Case: *Assume that ζ_N is of unit norm. Then ζ_N is a normalized* **Gegenbauer polynomial***:*

$$\begin{aligned}
\zeta_N(\mathbf{x}) &= \frac{1}{K_N^{(\nu)}}\, C_N^{(\nu)}(x_1) \\
\textit{where}\ \ C_N^{(\nu)}(x) &= \sum_{n=0}^{\lfloor N/2 \rfloor} (-1)^n 2^{N-2n} \cdot c_{N;n}^{(\nu)} \cdot x^{N-2n} \\
\textit{with}\ \ c_{N;n}^{(\nu)} &= \frac{\Gamma\left(\nu + (N-n)\right)}{\Gamma\left(\nu\right) \cdot n! \cdot (N-2n)!} \\
\textit{and where}\ \ \left(K_N^{(\nu)}\right)^2 &= \int_{\mathbb{S}^{D-1}} \left| C_N^{(\nu)}(x_1) \right|^2 d\mathbf{x} \\
&= \frac{2 \cdot \pi^{(D-1)/2}}{\Gamma(\nu)^2} \cdot \sum_{k=0}^{2\cdot\lfloor N/2 \rfloor} (-1)^k \cdot 2^{2N-2k} \cdot \frac{\Gamma\left(N-k+\frac{1}{2}\right)}{\Gamma\left(N-k+\frac{D}{2}\right)} \\
&\qquad \cdot \left(\sum_{n=0}^{k} c_{N;n}^{(\nu)} c_{N;(k-n)}^{(\nu)} \right),
\end{aligned}$$

where $\nu := \frac{D-2}{2}$.

Proof:

Proof of Case $D = 2$: It is clear from the definition of the Laplacian on $\mathbb{S}^1$ (see Theorem 39) that the function ζ_N is an eigenfunction of $\triangle_{\mathbb{S}^1}$. The subgroup of $\mathbb{SO}^2[\mathbb{R}]$ fixing $\mathbf{e}_1$ is just the two-element group of maps $(x_1, x_2) \mapsto (x_1, \pm x_2)$; since the function ζ_N is symmetric relative to the x_2 variable, it is zonal relative to these maps.

The formula (5.4) is then just a standard trigonometric identity, where we identify $x_1 = \cos(\theta)$; see, for example [66], §1.33 #3, p. 27.

Proof of Case $D = 3$: This is just the Gegenbauer polynomial when $D = 3$. For a direct proof, see, for example [13], Theorem 1, §2.1, p. 90, where there is unfortunately an error in the definition of the Legendre functions —see [168], §1, p.2, for a correct definition.

Proof of General Case: Since $\zeta_N(\mathbf{x})$ depends only on the first coordinate x_1, it is clearly zonal, relative to $\mathbb{SO}^D[\mathbb{R}]$ and $\mathbf{e}_1$. It remains to show ζ_N is an eigenfunction of $\triangle_{\mathbb{S}^{D-1}}$, with eigenvalue $-\lambda_N$ defined in formula (5.3).

We will use the coordinate system on $\mathbb{S}^{D-1}$ described in Example 39 on page 62, and the formula (5.1) of that example for the Laplacian. Relative

to this coordinate system, we have the identity

$$x_1 \;=\; \cos(\phi),$$

and, for any $M \in \mathbb{N}$,

$$\begin{aligned} \frac{\partial}{\partial\phi}\cos(\phi)^M &= -M\sin(\phi)\cos(\phi)^{M-1} \\ \text{and}\quad \frac{\partial^2}{\partial\phi^2}\cos(\phi)^M &= -M^2\cos(\phi)^M \;+\; M(M-1)\cos(\phi)^{M-2} \end{aligned}$$

so that

$$\begin{aligned} \triangle_{\mathbb{S}^{D-1}} x_1^M \;=\;& \frac{\partial^2}{\partial\phi^2}\cos(\phi)^M \;+\; (D-2)\cot(\phi)\frac{\partial}{\partial\phi}\cos(\phi)^M \\ & + \frac{1}{\sin(\phi)^2}\triangle_{\mathbb{S}^{D-2}}\cos(\phi)^M \\ =_{(1)}\;& -M^2\cos(\phi)^M \;+\; M(M-1)\cos(\phi)^{M-2} \\ & + (D-2)\frac{\cos(\phi)}{\sin\phi}(-M)\sin(\phi)\cos(\phi)^{M-1} \\ =_{(2)}\;& -\left(M^2+2M\nu\right)\cos(\phi)^M \;+\; M(M-1)\cos(\phi)^{M-2} \\ =\;& -\left(M^2+2M\nu\right)x_1^M \;+\; M(M-1)x_1^{M-2} \end{aligned}$$

(1) the differential operator $\triangle_{\mathbb{S}^{D-2}}$ is independent of the variable ϕ, and therefore trivial for this function.

(2) Since $D-2 \;=\; 2\nu$.

$$\begin{aligned} \text{Now, if}\quad \widetilde{C}_N^\nu(x_1) &= \Gamma(\nu)\cdot C_N^\nu(x_1) \\ &= \sum_{n=0}^{\lfloor N/2\rfloor} c_N^n \cdot x_1^{N-2n}, \\ \text{where}\quad c_N^n &:= (-1)^n 2^{N-2n}\frac{\Gamma\left[\nu+(N-n)\right]}{n!\cdot(N-2n)!}, \end{aligned}$$

then $\quad \triangle_{\mathbb{S}^{D-1}}\widetilde{C}_N^\nu(x_1)$

$$\begin{aligned} =\;& \sum_{n=0}^{\lfloor N/2\rfloor} c_N^n \triangle_{\mathbb{S}^{D-1}} x_1^{N-2n} \\ =\;& \sum_{n=0}^{\lfloor N/2\rfloor} c_N^n \Big[-\left((N-2n)^2+2(N-2n)\nu\right)x_1^{N-2n} \\ & + (N-2n)(N-2n-1)x_1^{N-2n-2}\Big] \end{aligned}$$

Note that, when $n = \lfloor N/2 \rfloor$, then the second term in the summand is trivial, because either $N - 2n = 0$ or $N - 2n - 1 = 0$. Reindexing the second term in the summand for all $n < \lfloor N/2 \rfloor$, we can rewrite this expression as

$$\begin{aligned} & \sum_{n=1}^{\lfloor N/2 \rfloor} \Big[-c_N^n \Big((N-2n)^2 + 2(N-2n)\nu \Big) x_1^{N-2n} \\ & \qquad + \; c_N^{n-1}(N-2n+2)(N-2n+1) \Big] x_1^{N-2n} \\ & \qquad -c_N^0 \Big(N^2 - 2N\nu \Big) x_1^N \\ = & \sum_{n=0}^{\lfloor N/2 \rfloor} -c_N^n \lambda_N^{(n)} x_1^{N-2n} \quad - \quad c_N^0 \lambda_N^{(0)} x_1^N \end{aligned}$$

where

$$\begin{aligned} \lambda_N^{(n)} &= \Big((N-2n)^2 + 2(N-2n)\nu \Big) \; - \; \frac{c_N^{n-1}}{c_N^n}(N-2n+2)(N-2n+1) \\ &= \Big((N-2n)^2 + 2(N-2n)\nu \Big) \; - \\ &\quad -1 \cdot 2^2 \cdot \frac{\Gamma(\nu+N-n+1)}{\Gamma(\nu+N-n)} \frac{n!}{(n-1)!} \frac{(N-2n)!}{(N-2n+2)!}(N-2n+2)(N-2n+1) \\ &= \Big((N-2n)^2 + 2(N-2n)\nu \Big) \; + \; 4(\nu+N-n)\cdot n \\ &= N^2 - 4nN + 2N\nu - 4n\nu + 4n^2 \quad + \quad 4n\nu + 4nN - 4n^2 \\ &= N^2 + 2N\nu \\ &= N(N+D-2) \quad (\text{ since } \nu = \tfrac{D-2}{2}.) \\ &= \lambda_N \end{aligned}$$

where λ_N is as in equation (5.3), and similarly,

$$\begin{aligned} \lambda_N^{(0)} &= N^2 - 2N\nu \\ &= \lambda_N \end{aligned}$$

We have thus showed that $\lambda_N^{(n)} \;=\; \lambda_N$ for all n, which implies that $\widetilde{C}_N^\nu$ is an eigenfunction with eigenvalue $-\lambda_N$; thus, so is C_N^ν

The Norm of the Gegenbauer Polynomial: Fix ν and N, and, for all $n = 0...\lfloor N/2 \rfloor$, define

$$c_n \;=\; (-1)^n \cdot 2^{N-2n} \cdot \frac{\Gamma\left(\nu + (N-n)\right)}{\Gamma(\nu) \cdot n! \cdot (N-2n)!}$$

$$\text{thus,} \quad \left(C_N^{(\nu)}(x)\right)^2 = \left(\sum_{n=0}^{\lfloor N/2 \rfloor} c_n \cdot x^{N-2n}\right)^2$$
$$= \sum_{n,m=0}^{\lfloor N/2 \rfloor} c_n \cdot c_m \cdot x^{2N-2n-2m}$$

Meanwhile, by Exercise 2.56 (p.77) of [48], for any k,

$$\int_{\mathbb{S}^{D-1}} x_1^{2N-2k}\, d\mathbf{x} = \frac{2\Gamma\left(\frac{2(N-k)+1}{2}\right)\Gamma(1/2)^{D-1}}{\Gamma\left(\frac{2(N-k)+D}{2}\right)}$$
$$= 2\pi^{(D-1)/2}\frac{\Gamma(N-k+1/2)}{\Gamma(N-k+D/2)}$$

Thus,

$$\left(K_N^{(\nu)}\right)^2 = \int_{\mathbb{S}^{D-1}} \left(C_N^{(\nu)}(x_1)\right)^2 d\mathbf{x}$$
$$= \sum_{n,m=0}^{\lfloor N/2 \rfloor} c_n \cdot c_m \int_{\mathbb{S}^{D-1}} x_1^{2N-2n-2m}\, d\mathbf{x}$$
$$= \sum_{k=0}^{2\cdot\lfloor N/2 \rfloor} \cdot \left(\sum_{n=0}^{k} c_n \cdot c_{n-k}\right) \cdot \int_{\mathbb{S}^{D-1}} x_1^{2N-2k}\, d\mathbf{x}$$
$$= \sum_{k=0}^{2\cdot\lfloor N/2 \rfloor} 2\pi^{(D-1)/2}\frac{\Gamma(N-k+1/2)}{\Gamma(N-k+D/2)} \cdot \left(\sum_{n=0}^{k} c_n \cdot c_{n-k}\right)$$

which, when simplified, is the formula of the theorem statement. ____□

The awful expression given in Proposition 44 for the norm of the Gegenbauer polynomial does not appear to admit further simplification, even when attacked with `MAPLE`. The expression for the Gegenbauer polynomials is far from transparent; some intuition about their behaviour can be gleaned from the plots provided in Appendix I on page 241.

5.3 Spherical Fourier Series in $\mathbf{L}^2$ and $\mathcal{C}^\infty$

Theorem 45: *(Spherical Fourier Analysis)*

For all $n \in \mathbb{N}$, let $\zeta_n : \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$ be the zonal harmonic polynomials defined by Proposition 44, and then define $\mathcal{Z}_n : \mathbb{S}^{D-1} \times \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$ by $\mathcal{Z}_n(\mathbf{x}, \mathbf{y}) = \zeta_n(\mathbf{e}_1) \cdot \zeta_n(\langle \mathbf{x}, \mathbf{y} \rangle)$. Then $\mathcal{Z}_n$ is rotationally equivariant.

Now, suppose $\gamma \in \mathbf{L}^2(\mathbb{S}^{D-1}; \mathbb{C})$. If we define $\check{\hat{\gamma}}_n := \mathcal{Z}_n * \gamma$ then $\check{\hat{\gamma}}_n \in \mathcal{H}^n\left(\mathbb{S}^{D-1}\right)$, and γ has the orthogonal decomposition:

$$\gamma = \sum_{n=1}^{\infty} \check{\hat{\gamma}}_n. \tag{5.5}$$

Thus, if, for all $N \in \mathbb{N}$, we define $\gamma_N := \sum_{n=1}^{N} \check{\hat{\gamma}}_n$, then $\|\gamma_N - \gamma\|_2 \xrightarrow[N \to \infty]{} 0$. Furthermore, if $\gamma \in \mathcal{C}^\infty(\mathcal{M})$, then $\|\gamma_N - \gamma\|_\infty \xrightarrow[N \to \infty]{} 0$.

Proof: This follows from Theorem 34 on page 56, and Corollary 38 on page 61, using the zonal functions provided by Proposition 44 on page 68. □

Definition 46: *Spherical Fourier Coefficients*

If $\gamma \in \mathbf{L}^2(\mathbb{S}^{D-1}; \mathbb{C})$, then the **spherical Fourier Coefficients** *of γ are the functions $\check{\hat{\gamma}}_n := \mathcal{Z}_n * \gamma$, for $n \in \mathbb{N}$.*

The **spherical Fourier series** *for γ is then the orthogonal decomposition $\gamma = \sum_{n=1}^{\infty} \check{\hat{\gamma}}_n$.*

Remark: Notice that the spherical Fourier "coefficients" are themselves *functions*, not numbers. These functions can be written as linear combinations of the elements provided by Proposition 43 on page 67

Corollary 47: *((De)convolution on Spheres)*

Suppose $\eta : \mathbb{S}^{D-1} \times \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$ is rotationally equivariant, and suppose that $\mathbf{g} := \eta * \gamma$. Then for all $n \in \mathbb{N}$, $\check{\hat{\mathbf{g}}}_n = A_n \cdot \check{\hat{\gamma}}_n$, where we define $A_n := \dfrac{(\eta * \zeta_n)(\mathbf{e}_1)}{\zeta_n(\mathbf{e}_1)}$.

Conversely, suppose that γ is unknown, but we know η and $\mathbf{g}$. We can reconstruct γ via the formula: $\gamma = \sum_{n=1}^{\infty} \frac{1}{A_n} \check{\hat{\mathbf{g}}}_n$.

Proof: This follows from the previous theorem, and Proposition 37 on page 58. □

Example 48: *(Spherical Fourier series on $\mathbb{S}^1$)*

Let for $N \in \mathbb{N}$, let $\zeta_N : \mathbb{S}^1 \longrightarrow \mathbb{C}$ be as in **Part 1** of Proposition 44 on page 68:

$$\zeta_N(\theta) \quad = \quad \cos(N\theta) \quad = \quad \frac{1}{2}\left(\mathcal{E}_N(\theta) + \mathcal{E}_{(-N)}(\theta)\right)$$

where we identify $\mathbb{S}^1 \stackrel{1}{=} [0, 2\pi)$, and define $\mathcal{E}_K(\theta) := \exp(K\theta \cdot \mathbf{i})$. Let $\mathcal{Z}_N : \mathbb{S}^1 \times \mathbb{S}^1 \longrightarrow \mathbb{C}$ be defined from ζ_N as in Theorem 47 on the preceding page. Then, for any $\gamma : \mathbb{S}^1 \longrightarrow \mathbb{R}$,

$$\begin{aligned}
\check{\gamma}_N &= \mathcal{Z}_N * \gamma \\
&=_{(1)} \gamma * \zeta_N \\
&= \frac{1}{2}\left(\gamma * \mathcal{E}_N + \gamma * \mathcal{E}_{(-N)}\right) \\
&=_{(2)} \frac{1}{2}\left(\widehat{\gamma}(N) \cdot \mathcal{E}_N + \widehat{\gamma}(-N) \cdot \mathcal{E}_{(-N)}\right) \\
&=_{(3)} \frac{1}{2}\left(\widehat{\gamma}(N) \cdot \mathcal{E}_N + \overline{\widehat{\gamma}(N) \cdot \mathcal{E}_N}\right) \\
&= \mathbf{re}\left[\widehat{\gamma}(N) \cdot \mathcal{E}_N\right].
\end{aligned}$$

(1) where the convolution is now meant in the "usual" sense on the group $\mathbb{S}^1 = \mathbb{T}^1$.

(2) here, $\widehat{\gamma}$ is the (classical) Fourier transform of γ as a function on the circle.

(3) because γ is real-valued.

Now, if we write $\widehat{\gamma}(N) = r_N \exp(\phi_N \cdot \mathbf{i})$, where $r_N \in [0, \infty)$ and $\phi_N \in [0, 2\pi)$, then, for any $\theta \in \mathbb{S}^1 \cong [0, 2\pi)$, we have:

$$\begin{aligned}
\check{\gamma}_N(\theta) &= \mathbf{re}\left[r_N \cdot \exp(\phi_N \mathbf{i}) \cdot \mathcal{E}_N(\theta)\right] \\
&= r_N \cdot \mathbf{re}\left[\exp(\phi_N \mathbf{i}) \cdot \exp(N \cdot \theta \cdot \mathbf{i})\right] \\
&= r_N \cdot \mathbf{re}\left[\exp\left(N \cdot \left(\theta + \frac{\phi_N}{N}\right) \cdot \mathbf{i}\right)\right] \\
&= r_N \cdot \mathbf{re}\left[\mathcal{E}_N\left(\theta + \frac{\phi_N}{N}\right)\right] \\
&= r_N \cdot \zeta_N\left(\theta + \frac{\phi_N}{N}\right).
\end{aligned}$$

In other words, convolving ζ_N by γ is equivalent to multiplying the magnitude of ζ_N by r_N, and rotating the phase by ϕ_N/N.

For any function $f : \mathbb{S}^1 \longrightarrow \mathbb{C}$, and $\phi \in \mathbb{S}^1$, we define the function $\phi_* f : \mathbb{S}^1 \longrightarrow \mathbb{C}$ by $\phi_* f(\theta) = f(\theta + \phi)$. Then the equation (5.5) on page 73 can be written:

$$\gamma = \sum_{n=1}^{\infty} r_N \cdot \left(\frac{\phi_N}{N}\right)_* \zeta_N.$$

5.4 Asymptotic Decay and Convergence Rates

In classical harmonic analysis, the infinitesimal properties of a function f are in many ways reflected in the asymptotic behaviour of its Fourier transform, and vice versa. Generally, the smoother f is, the more rapidly $\widehat{f}$ decays near infinity. Conversely, if f is very "jaggy", undifferentiable, or discontinuous, then $\widehat{f}$ decays slowly or not at all near infinity, reflecting a concentration of the "energy" of f in high frequency Fourier components.

Hence, when approximating f by a partial Fourier sum, the more jaggy f is, the more slowly the sum converges, and the more terms we must include to ensure ourselves of a good approximation.

A similar phenomenon manifests when approximating a functions $\gamma : \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$ by a spherical Fourier series. By relating the decay rate of the spherical Fourier series to the smoothness of γ, we will be able to estimate the error introduced by approximating γ with a partial spherical Fourier sum. Intuitively, the more "singular" γ becomes, the greater the proportion of its energy we will find in the "high frequency" eigenspaces —that is, those associated with large magnitude eigenvalues.

Formally, say that a sequence of functions $\left[\check{\widehat{\gamma}}_n|_{n=1}^{\infty}\right]$ is of **order** less than $\mathcal{O}(n^{-\alpha})$ if $\lim_{n\to\infty} \frac{\left\|\check{\widehat{\gamma}}_n\right\|_2}{n^{\alpha}} = 0.$

Theorem 49: *Let $\gamma : \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$, and suppose that γ is continuously $2M$-differentiable. Then the sequence $\left[\check{\widehat{\gamma}}_n|_{n=1}^{\infty}\right]$ is of order less than or equal to $\mathcal{O}(n^{-(2M+1)})$*

Proof: First suppose that γ is twice continuously differentiable. Thus, using formula (5.1) on page 64, we can apply $\triangle_{\mathbb{S}^{D-1}}$ to γ. Let $\alpha = \triangle_{\mathbb{S}^{D-1}}\gamma$. Since α is a continuous function, it is in $\mathbf{L}^2(\mathbb{S}^{D-1})$, and we can compute

the spherical Fourier coefficients $\check{\hat{\alpha}}_n = \mathcal{Z}_n * \alpha$, for all n, and conclude: $\alpha = \sum_{n=1}^{\infty} \check{\hat{\alpha}}_n$.

In particular, since this sum converges absolutely in $\mathbf{L}^2(\mathbb{S}^{D-1})$, we know that the sequence $\left[\check{\hat{\alpha}}_n|_{n=1}^{\infty}\right]$ is of order less than $\mathcal{O}(n^{-1})$.

By construction, we know that $\check{\hat{\gamma}}_n = \mathcal{Z}_n * \gamma$ is an eigenfunction of $\triangle_{\mathbb{S}^{D-1}}$, with eigenvalue $\lambda_n = n(n+D-2)$. By Claim 2 on page 59 of Proposition 37 on page 58, the Laplacian operator commutes with convolution operators. Thus,

$$\begin{aligned} n(n+D-2)\check{\hat{\gamma}}_n &= \triangle_{\mathbb{S}^{D-1}} \check{\hat{\gamma}}_n \\ &= \triangle_{\mathbb{S}^{D-1}} (\mathcal{Z}_n * \gamma) \\ &= \mathcal{Z}_n * (\triangle_{\mathbb{S}^{D-1}} \gamma) \\ &= \mathcal{Z}_n * \alpha \\ &= \check{\hat{\alpha}}_n \end{aligned}$$

Since this is true for all n, we conclude that $\left[\check{\hat{\gamma}}_n|_{n=1}^{\infty}\right]$ is of order less than $\mathcal{O}\left(\frac{1}{n(n+D-2)}\right) \cdot \mathcal{O}(n^{-1}) = \mathcal{O}(n^{-3})$.

Proceed inductively to prove the general case. □

Corollary 50: *(Application to Spectral Measures)*

Let $\alpha \in [0,2)$, $\alpha \neq 1$, and suppose ρ is an α-stable probability measure on $\mathbb{R}^D$ with density function $F : \mathbb{R}^D \longrightarrow [0,\infty)$ and spectral measure Γ, and spherical log-characteristic function $\mathbf{g}$, with $\min_{\mathbf{x}\in\mathbb{S}^{D-1}} \mathbf{g}(\mathbf{x}) > 0$. Suppose that Γ is absolutely continuous relative to $\mathcal{L}^{bsg}$, and that $d\Gamma = \gamma \, d\mathcal{L}^{bsg}$, with $\gamma \in \mathbf{L}^2(\mathbb{S}^{D-1};\ \mathcal{L}^{bsg})$.

For all $N \in \mathbb{N}$, let $\gamma^{[N]} = \sum_{n=1}^{N} \check{\hat{\gamma}}_n$, let $\Gamma^{[N]} = \gamma^{[N]} \mathcal{L}^{bsg}$, and let $\rho^{[N]}$ be the corresponding α-stable probability measure, with density function $F^{[N]} : \mathbb{R}^D \longrightarrow [0,\infty)$. Then:

1. *For all $p \in [1,\infty]$, $\lim_{n\to\infty} \left\| F - F^{[n]} \right\|_p = 0$.*
2. *If $\gamma \in \mathcal{C}^{2M}(\mathbb{S}^{D-1})$, then $\left\| F - F^{[n]} \right\|_p$ is of order less than $\mathcal{O}\left(n^{-2M}\right)$.*

Proof: We will apply Corollaries 14 and 15 on page 33. First we must establish:

Claim 1: $\liminf_{n\in[1...\infty]} \mathcal{K}\left(\rho^{[n]}\right) > 0$, *where* $\mathcal{K}$ *is as in Definition 6 on page 19, and where* $\rho^{[\infty]} = \rho$.

Proof: By hypothesis, $\mathcal{K}(\rho) = \min_{\mathbf{x}\in\mathbb{S}^{D-1}} \mathbf{g}(\mathbf{x}) > 0$. We will show that $\liminf_{n\to\infty} \mathcal{K}\left(\rho^{[n]}\right) > \mathcal{K}(\rho)/2$.

Let $G^{[\infty]} := \Gamma[\mathbb{S}^{D-1}]$, and, for all $n > 0$, let $G^{[n]} := \Gamma^{[n]}[\mathbb{S}^{D-1}]$. Let $\widetilde{G} = 2 \cdot G^{[\infty]}$.

Claim 1.1: *There is some* $N \in \mathbb{N}$ *so that, for all* $n \in [N...\infty]$, $G^{[n]} < \widetilde{G}$.

Proof: Let $\mathbb{1}$ be the unit function on $\mathbb{S}^{D-1}$. Thus, $G^{[\infty]} = \langle \gamma,\ \mathbb{1} \rangle$ and, for all $n \in \mathbb{N}$, $G^{[n]} = \left\langle \gamma^{[n]},\ \mathbb{1} \right\rangle$. Since $\gamma^{[n]} \xrightarrow[n\to\infty]{} \gamma$ in $\mathbf{L}^2$, we conclude that $G^{[n]} = \left\langle \gamma^{[n]},\ \mathbb{1} \right\rangle \xrightarrow[n\to\infty]{} \langle \gamma,\ \mathbb{1} \rangle = G^{[\infty]}$. Thus, if n is large enough, then $G^{[n]} < 2 \cdot G^{[\infty]}$. ..... □ [Claim 1.1]

Let $\mathbf{g}^{[n]} = \eta^{(\alpha)} * \Gamma^{[n]}$. Thus, $\mathcal{K}\left(\rho^{[n]}\right) = \min_{\mathbf{x}\in\mathbb{S}^{D-1}} \mathbf{g}^{[n]}(\mathbf{x})$.

Claim 1.2: *There is some finite subset* $\Theta \subset \mathbb{S}^{D-1}$ *so that, if* $\mathbf{g}^{[n]} \xrightarrow[n\to\infty]{} \mathbf{g}$ *for all* $\theta \in \Theta$, *then* $\mathcal{K}\left(\rho^{[n]}\right) > \mathcal{K}(\rho)/2$.

Proof: We will employ the smoothness properties of $\mathbf{g}$ and $\mathbf{g}^{[n]}$ described by Proposition 24 on page 45. Let $\epsilon = \dfrac{\mathcal{K}(\rho)}{6}$, and then let

$$\Upsilon = \begin{cases} \dfrac{\epsilon}{C_\alpha \cdot \widetilde{G}} & \text{if } \alpha > 1 \\ \\ \delta_\alpha\left(\epsilon/\widetilde{G}\right) & \text{if } \alpha \leq 1 \end{cases}$$

(where $\delta_\alpha(\bullet)$ and C_α are as in Proposition 24). Let Θ be a finite, Υ-dense subset of $\mathbb{S}^{D-1}$.

Let N be as in Claim 1.1. Assume $n > N$, and suppose $\left|\mathbf{g}^{[n]}(\theta) - \mathbf{g}(\theta)\right| < \epsilon$ for all $\theta \in \Theta$. Then for any $\mathbf{s} \in \mathbb{S}^{D-1}$, find a nearby $\theta \in \Theta$ and use Proposition 24 to conclude: $\left|\mathbf{g}^{[n]}(\mathbf{s}) - \mathbf{g}(\mathbf{s})\right| \leq \left|\mathbf{g}^{[n]}(\mathbf{s}) - \mathbf{g}^{[n]}(\theta)\right| + \left|\mathbf{g}^{[n]}(\theta) - \mathbf{g}(\theta)\right| + |\mathbf{g}(\theta) - \mathbf{g}(\mathbf{s})| < 3\epsilon = \dfrac{\mathcal{K}(\rho)}{2}$. Thus, $\left\|\mathbf{g}^{[n]} - \mathbf{g}\right\|_\infty < \dfrac{\mathcal{K}(\rho)}{2}$; this implies that $\mathcal{K}\left(\rho^{[n]}\right) > \dfrac{\mathcal{K}(\rho)}{2}$. ..... □ [Claim 1.2]

It remains to show:

Claim 1.3: *For all* $\theta \in \Theta$, $\mathbf{g}^{[n]}(\theta) \xrightarrow[n\to\infty]{} \mathbf{g}(\theta)$.

Proof: Let $\eta_\theta^{(\alpha)} \in \mathbf{L}^2\left(\mathbb{S}^{D-1}\right)$ be defined: $\eta_\theta^{(\alpha)}(\mathbf{s}) = \eta^{(\alpha)}\left(\langle\theta, \mathbf{s}\rangle\right)$. Then: $\mathbf{g}^{[n]}(\theta) = \int_{\mathbb{S}^{D-1}} \eta_\theta^{(\alpha)}\, d\Gamma^{[n]}[\mathbf{s}] = \left\langle \eta_\theta^{(\alpha)},\ \gamma^{[n]} \right\rangle \xrightarrow[n\to\infty]{}$ $\left\langle \eta_\theta^{(\alpha)},\ \gamma \right\rangle = \mathbf{g}(\theta)$, because $\gamma^{[n]} \xrightarrow[n\to\infty]{} \gamma$ in $\mathbf{L}^2$. .. □ [Claim 1.3]

..... □ [Claim 1]

Proof of Part 1: Note first that $\left\|\Gamma - \Gamma^{[n]}\right\|_{[2]} = \mathcal{A}^{1/2}\left\|\gamma - \gamma^{[n]}\right\|_2$ (where $\|\bullet\|_{[2]}$ is the norm from Definition 11 on page 27, and $\mathcal{A} = \mathbf{Area}\left[\mathbb{S}^{D-1}\right]$), and that, by Proposition 45 on page 72, $\left\|\gamma - \gamma^{[n]}\right\|_2 \xrightarrow[N\to\infty]{} 0$. Now apply Corollary 15.

Proof of Part 2: Use Theorem 49 to conclude that $\left\|\gamma - \gamma^{[n]}\right\|_2$ is of order less than $\mathcal{O}\left(n^{-2M}\right)$. Thus, $\left\|\Gamma - \Gamma^{[n]}\right\|_{[2]}$ is also of order less than $\mathcal{O}\left(n^{-2M}\right)$. Thus, applying Corollary 14, we conclude that $\left\|F - F^{[n]}\right\|_p$ is of order less than $\mathcal{O}\left(n^{-2M}\right)$. □

5.5 Spherical Fourier Series for Arbitrary Measures

Although the classical theory of spherical harmonics was developed in $\mathbf{L}^2$, it is relatively easy to extend these ideas to arbitrary Borel measures on $\mathbb{S}^{D-1}$. Suppose $\Gamma \in \mathcal{M}_{\mathcal{EAS}}\left[\mathbb{S}^{D-1}; \mathbb{C}\right]$ is a complex-valued measure.

Definition 51: *Convolution*

If $\mathcal{Z} : \mathbb{S}^{D-1} \times \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$ *is a rotationally equivariant function, then define* $\mathcal{Z} * \Gamma : \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$ *by:*

$$(\mathcal{Z} * \Gamma)(\mathbf{s}) = \int_{\mathbb{S}^{D-1}} \mathcal{Z}(\theta, \mathbf{s})\, d\Gamma[\theta]$$

Definition 52: *Spherical Fourier Coefficients of a Measure*

For all $n \in \mathbb{N}$, let $\zeta_n : \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$ be the zonal spherical harmonics of Proposition 44, and define $\mathcal{Z}_n : \mathbb{S}^{D-1} \times \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$ by $\mathcal{Z}_n(\mathbf{x}, \mathbf{y}) = \zeta_n(\mathbf{e}_1) \cdot \zeta_n(\langle \mathbf{x}, \mathbf{y} \rangle)$. Then define the n**th spherical Fourier coefficient** *of Γ:*

$$\check{\hat{\Gamma}}_n \;:=\; \mathcal{Z}_n * \Gamma.$$

Lemma 53: *If $\phi \in \mathcal{C}\left(\mathbb{S}^{D-1}; \mathbb{C}\right)$, then, for all $n \in \mathbb{N}$,*

$$\int_{\mathbb{S}^{D-1}} \phi \; d\check{\hat{\Gamma}}_n \;=\; \int_{\mathbb{S}^{D-1}} \check{\hat{\phi}}_n \; d\Gamma$$

Proof:

$$\int_{\mathbb{S}^{D-1}} \phi(\mathbf{s}) \; d\check{\hat{\Gamma}}_n[\mathbf{s}] \;=\; \int_{\mathbb{S}^{D-1}} \int_{\mathbb{S}^{D-1}} \phi(\mathbf{s}) \mathcal{Z}(\theta, \mathbf{s}) \; d\Gamma[\theta] \;=\; \int_{\mathbb{S}^{D-1}} \check{\hat{\phi}}_n(\theta) \; d\Gamma[\theta].$$

□

Proposition 54: *For all $N \in \mathbb{N}$, define $\Gamma_N := \sum_{n=1}^{N} \check{\hat{\Gamma}}_n$. Then $\Gamma_N \xrightarrow[N \to \infty]{}$ Γ in the weak* topology*

Proof: Since $\mathcal{C}^\infty\left[\mathbb{S}^{D-1}\right]$ is dense in $\mathcal{C}\left[\mathbb{S}^{D-1}\right]$ in the uniform norm, it is sufficient to prove that, for all $\phi \in \mathcal{C}^\infty\left[\mathbb{S}^{D-1}\right]$, $\int_{\mathbb{S}^{D-1}} \phi(\mathbf{s}) \; d\Gamma_N$ converges to $\int_{\mathbb{S}^{D-1}} \phi(\mathbf{s}) \; d\Gamma$ as $N \to \infty$. But

$$\begin{aligned}
\int_{\mathbb{S}^{D-1}} \phi(\mathbf{s}) \; d\Gamma_N \;&=\; \sum_{n=1}^{N} \int_{\mathbb{S}^{D-1}} \phi(\mathbf{s}) \; d\check{\hat{\Gamma}}_n \\
&=\; \sum_{n=1}^{N} \int_{\mathbb{S}^{D-1}} \check{\hat{\phi}}_n(\mathbf{s}) \; d\Gamma \quad \text{(by Lemma 53)} \\
&=\; \int_{\mathbb{S}^{D-1}} \phi_N \; d\Gamma,
\end{aligned}$$

where $\phi_N = \sum_{n=1}^{N} \check{\hat{\phi}}_n$. By Theorem 45 on page 72, $\phi_N \xrightarrow[N \to \infty]{} \phi$ in the uniform norm so that $\int_{\mathbb{S}^{D-1}} \phi_N \; d\Gamma \xrightarrow[N \to \infty]{} \int_{\mathbb{S}^{D-1}} \phi \; d\Gamma$. ——————□

Corollary 55: *(Application to Spectral Measures)*
Let $\alpha \in [0, 2)$, and let ρ be an α-stable probability measure on $\mathbb{R}^D$ with spectral measure Γ.

For all $N \in \mathbb{N}$, let $\Gamma^{[N]} = \sum_{n=1}^{N} \check{\hat{\Gamma}}_n$, and let $\rho^{[N]}$ be the corresponding α-stable probability measure. Then $\rho_N \xrightarrow[N\to\infty]{} \rho$ in the weak topology on $\mathcal{MEAS}\left[\mathbb{R}^D\right]$.*

Proof: This follows immediately from Proposition 54 and Proposition 1 on page 14. □

Chapter 6

Estimating spectral measures via Convolution on Special Orthogonal Groups

In this chapter, I sketch a method of recovering the spectral measure, Γ, of a stable distribution from its empirical characteristic function by "pulling back" the spectral measure Γ and the spherical log-characteristic function $\mathbf{g}$ (section 4.1 on page 43) to the Lie group $\mathbb{SO}^D[\mathbb{R}]$, which acts transitively on the sphere $\mathbb{S}^{D-1}$. Once inside the Lie group, we can write $\mathbf{g}$ as the result of *convolving* Γ with a certain kernel on $\mathbb{SO}^D[\mathbb{R}]$. Recovering Γ from $\mathbf{g}$ then consists of "undoing" this convolution, which can be accomplished through methods from noncommutative harmonic analysis.

This chapter assumes some familiarity with the representation theory of compact groups. Background is provided in Appendix H.

6.1 Pulling back from $\mathbb{S}^{D-1}$ to $\mathbb{SO}^D[\mathbb{R}]$.

The compact Lie group $\mathbb{SO}^D[\mathbb{R}]$ has a natural structure as a **smooth fibre bundle**[1] over $\mathbb{S}^{D-1}$, with generic fibre $\mathbb{SO}^{D-1}[\mathbb{R}]$. Fix $\mathbf{e} \in \mathbb{S}^{D-1}$, and define

$$\begin{array}{rcl} \mathbf{P} : \mathbb{SO}^D[\mathbb{R}] & \longrightarrow & \mathbb{S}^{D-1} \\ g & \mapsto & g(\mathbf{e}) \end{array}$$

Suppose $\mathcal{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_D\}$ is an orthonormal basis for $\mathbb{R}^D$, with $\mathbf{e}_1 = \mathbf{e}$. Given the value of $\mathbf{P}(g) = g(\mathbf{e}_1)$, the values of $\{g(\mathbf{e}_2), \ldots, g(\mathbf{e}_D)\}$ completely

[1] See [32] §5.5 or [179], §3.65 for more on $\mathbb{S}^{D-1}$ as a $\mathbb{SO}^D[\mathbb{R}]$-manifold. The definitive reference on fibre bundles in general is [65].

determine g; For a fixed value of $\mathbf{P}(g) = \mathbf{x}$, the $(D-1)$-tuple of vectors $\{g(\mathbf{e}_2), \ldots, g(\mathbf{e}_D)\}$ can range freely over any orthonormal basis of $\mathbf{x}^\perp$. If we fix a specific orthonormal basis of $\mathbf{x}^\perp$, then all other orthonormal bases can be seen as images of this one under orthogonal transforms; hence, there is a diffeomorphic correspondence between the elements of $\mathbf{P}^{-1}\{\mathbf{x}\}$ and the elements of $\mathbb{SO}\left[\mathbf{x}^\perp\right] \cong \mathbb{SO}^{D-1}\left[\mathbb{R}^D\right]$.

This shows that the generic fibre of the map $\mathbf{P} : \mathbb{SO}^D[\mathbb{R}] \longrightarrow \mathbb{S}^{D-1}$ is diffeomorphic to $\mathbb{SO}^{D-1}\left[\mathbb{R}^D\right]$. It is not difficult to show that the fibre bundle structure is smooth.

Let $\mathcal{H}^{aar}_{\mathbb{SO}^D[\mathbb{R}]}$ be the **Haar measure** on $\mathbb{SO}^D[\mathbb{R}]$, and $\mathcal{L}^{bsg}$ the Lebesgue Measure on $\mathbb{S}^{D-1}$ (normalized to have total mass one). Then the projection $\mathbf{P} : \mathbb{SO}^D[\mathbb{R}] \longrightarrow \mathbb{S}^{D-1}$ is measurable and measure-preserving, and the aforementioned fibre bundle structure provides us with a natural **disintegration**[2] of measures:

$$\mathcal{H}^{aar}_{\mathbb{SO}^D[\mathbb{R}]} = \int_{\mathbb{S}^{D-1}} \mathcal{H}^{aar}_{\mathbf{x}} \, d\mathcal{L}^{bsg}[\mathbf{x}]$$

where $\mathcal{H}^{aar}_{\mathbf{x}}$ is the "fibre" Haar measure on $\mathbf{P}^{-1}\{\mathbf{x}\} \cong \mathbb{SO}^{D-1}\left[\mathbb{R}^D\right]$.

We can generalize this construction. If Γ is *any* measure on $\mathbb{S}^{D-1}$, then define the measure $\widetilde{\Gamma}$ on $\mathbb{SO}^D[\mathbb{R}]$ via the disintegration

$$\widetilde{\Gamma} = \int_{\mathbb{S}^{D-1}} \mathcal{H}^{aar}_{\mathbf{x}} \, d\Gamma[\mathbf{x}]$$

If $\eta : \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$ is any function, then define $\widetilde{\eta} = \eta \circ \mathbf{P} : \mathbb{SO}^D[\mathbb{R}] \longrightarrow \mathbb{C}$.

Example 56:

Suppose $\mathbf{e} = (1, 0, \ldots, 0)$, and that that $\eta^{(\alpha)} : \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$ is defined

$$\eta^{(\alpha)}(\mathbf{x}) = |\langle \mathbf{x}, \mathbf{e} \rangle|^\alpha + \mathcal{B}_\alpha \mathbf{i} \langle \mathbf{x}, \mathbf{e} \rangle^{\langle \alpha \rangle}.$$

Then $\widetilde{\eta}^{(\alpha)} : \mathbb{SO}^D[\mathbb{R}] \longrightarrow \mathbb{C}$ is the function

$$\widetilde{\eta}^{(\alpha)}(g) = |g_{11}|^\alpha + \mathcal{B}_\alpha \cdot \mathbf{i} \cdot g_{11}^{\langle \alpha \rangle}.$$

where $g \in \mathbb{SO}^D[\mathbb{R}]$ is represented by the matrix $\begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1D} \\ g_{21} & g_{22} & \cdots & g_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ g_{D1} & g_{D2} & \cdots & g_{DD} \end{bmatrix}$.

[2] See [159] or [50] for good introductions to measure disintegration.

Lemma 57: *For any $\eta : \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$ and any measure Γ on $\mathbb{S}^{D-1}$,*

$$\int_{\mathbb{S}^{D-1}} \eta \; d\Gamma \quad = \quad \int_{\mathbb{SO}^D[\mathbb{R}]} \widetilde{\eta} \; d\widetilde{\Gamma}$$

Proof:

$$\begin{aligned}
\int_{\mathbb{SO}^D[\mathbb{R}]} \widetilde{\eta} \; d\widetilde{\Gamma} &= \int_{\mathbb{S}^{D-1}} \int_{\mathbf{P}^{-1}\{\mathbf{x}\}} \widetilde{\eta}[g] \; d\mathcal{H}^{aar}_{\mathbf{x}}[g] \; d\Gamma[\mathbf{x}] \\
&= \int_{\mathbb{S}^{D-1}} \int_{\mathbf{P}^{-1}\{\mathbf{x}\}} \eta[\mathbf{P}(g)] \; d\mathcal{H}^{aar}_{\mathbf{x}}[g] \; d\Gamma[\mathbf{x}] \\
&= \int_{\mathbb{S}^{D-1}} \int_{\mathbf{P}^{-1}\{\mathbf{x}\}} \eta(x) \; d\mathcal{H}^{aar}_{\mathbf{x}} \; d\Gamma[\mathbf{x}] \\
&= \int_{\mathbb{S}^{D-1}} \eta(x) \left(\int_{\mathbf{P}^{-1}\{\mathbf{x}\}} d\mathcal{H}^{aar}_{\mathbf{x}} \right) d\Gamma[\mathbf{x}] \\
&= \int_{\mathbb{S}^{D-1}} \eta(x) \; d\Gamma[\mathbf{x}]
\end{aligned}$$

Because $\|\mathcal{H}^{aar}_{\mathbf{x}}\| = 1$ for all $\mathbf{x} \in \mathbb{S}^{D-1}$. □

Now we reintroduce a concept important from Chapter 5.

Definition 58: *Zonal Function*

A function $\eta : \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$ is **zonal about** $\mathbf{e} \in \mathbb{S}^{D-1}$ *if, for any $T \in \mathbb{SO}^D[\mathbb{R}]$ such that $T(\mathbf{e}) = \mathbf{e}$,*

$$\eta \circ T \; = \; \eta$$

If this is the case, then, for any $\mathbf{x} \in \mathbb{S}^{D-1}$, define $\eta_{\mathbf{x}}$ to be equal to $\eta \circ T_{\mathbf{x}}^{-1}$, where $T_{\mathbf{x}} \in \mathbb{SO}^D[\mathbb{R}]$ is any transform so that $T(\mathbf{e}) = \mathbf{x}$ (ie. $T \in \mathbf{P}^{-1}\{\mathbf{x}\}$).

Recall that η is zonal about $\mathbf{e}$ if and only if, for any $\mathbf{x} \in \mathbb{S}^{D-1}$, $\eta(\mathbf{x})$ is a function only of the distance from $\mathbf{x}$ to $\mathbf{e}$.

Lemma 59: *If η is zonal about $\mathbf{e}$, then $\eta_{\mathbf{x}}$ is well-defined, independent of the choice of $T_{\mathbf{x}}$.*

Proof: Suppose $T_1, T_2 \in \mathbb{SO}^D[\mathbb{R}]$, so that $T_1(\mathbf{e}) = \mathbf{x} = T_2(\mathbf{x})$. Then $T_1^{-1} \circ T_2(\mathbf{e}) = \mathbf{e}$; hence, since η is zonal, $\eta \circ T_1^{-1} \circ T_2 = \eta$, hence $\eta \circ T_1^{-1} = \eta \circ T_2^{-1}$. □

Example 60:

For any $\alpha \in [0,2]$, $\eta^{(\alpha)} : \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$ is zonal.

For any $\theta \in \mathbb{S}^{D-1}$, $\eta_\theta^{(\alpha)}$ is the function:

$$\eta_\theta^{(\alpha)}(\mathbf{x}) \quad = \quad |\langle \mathbf{x}, \theta \rangle|^\alpha \; + \; \mathcal{B}_\alpha \cdot \mathbf{i} \cdot \langle \mathbf{x}, \theta \rangle^{\langle \alpha \rangle} .$$

Recall the following from Chapter 5:

Definition 61: *Convolution (on a sphere)*

If Γ is any measure on $\mathbb{S}^{D-1}$, and $\eta : \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$ is zonal, define $\Gamma * \eta : \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$ by $\forall \mathbf{x} \in \mathbb{S}^{D-1}, \quad (\Gamma * \eta)(\mathbf{x}) \; = \; \langle \eta_{\mathbf{x}}, \; \Gamma \rangle \; = \; \int_{\mathbb{S}^{D-1}} \eta_{\mathbf{x}} \; d\Gamma.$

Example 62:

If $\eta^{(\alpha)}$ is as in the previous example, and Γ is a spectral measure on $\mathbb{S}^{D-1}$, then $\Gamma * \eta^{(\alpha)}$ is the spherical log-characteristic function $\mathbf{g}$ (§ 4.1 on page 43). ---

Definition 63: *Convolution (on $\mathbb{SO}^D[\mathbb{R}]$)*

If $\widetilde{\eta} : \mathbb{SO}^D[\mathbb{R}] \longrightarrow \mathbb{C}$, and $\widetilde{\Gamma}$ is a measure on $\mathbb{SO}^D[\mathbb{R}]$, we define the **convolution** $\widetilde{\Gamma} * \widetilde{\eta} : \mathbb{SO}^D[\mathbb{R}] \longrightarrow \mathbb{C}$*:*

$$\forall g \in \mathbb{SO}^D[\mathbb{R}], \quad (\widetilde{\Gamma} * \widetilde{\eta})(g) \; = \; \int_{\mathbb{SO}^D[\mathbb{R}]} \widetilde{\eta}(h^{-1} \cdot g) \; d\Gamma[h].$$

It can be shown (Lemma 166 on page 233) that this somewhat peculiar formula is equivalent to the "usual" definition of convolution on a compact group.

We would like to show that the two convolutions are "the same" in some sense. To do this we need η to satisfy an additional requirement

Lemma 64: *If $\eta : \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$ is zonal about $\mathbf{e}$, then it is* **reflectionally symmetric** *about $\mathbf{e}$: for any $T \in \mathbb{SO}^D[\mathbb{R}]$,*

$$\eta \circ T(\mathbf{e}) \;=\; \eta \circ T^{-1}(\mathbf{e})$$

Proof: If η is zonal, then, for any $\mathbf{x}$, $\eta(\mathbf{x}) = f(\mathbf{dist}\,[\mathbf{x},\mathbf{e}])$ for some function $f : [0,\infty) \longrightarrow \mathbb{C}$. Thus, $\eta \circ T(\mathbf{e}) = f(\mathbf{dist}\,[\mathbf{e}, T(\mathbf{e})]) = f(\mathbf{dist}\,[\mathbf{e}, T^{-1}(\mathbf{e})]) = \eta \circ T^{-1}(\mathbf{e})$. □

For example, $\eta^{(\alpha)}(\mathbf{x}) = |\langle \mathbf{x}, \mathbf{e}\rangle|^{\alpha} + \mathcal{B}_\alpha \cdot \mathbf{i} \cdot \langle \mathbf{x}, \mathbf{e}\rangle^{\langle\alpha\rangle}$ is reflectionally symmetric.

Lemma 65: *If $\eta : \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$ is zonal about $\mathbf{e}$, and Γ is some measure on $\mathbb{S}^{D-1}$, then*

$$\widetilde{\Gamma * \eta} \;=\; \widetilde{\Gamma} * \widetilde{\eta}$$

Proof:

Claim 1: *For any $\mathbf{x}, \mathbf{y} \in \mathbb{S}^{D-1}$, $\eta_{\mathbf{x}}(\mathbf{y}) = \eta_{\mathbf{y}}(\mathbf{x})$.*

Proof:

$$\begin{aligned}
\eta_{\mathbf{x}}(\mathbf{y}) &= \eta \circ T_{\mathbf{x}}^{-1}(\mathbf{y}) \\
&= \eta \circ T_{\mathbf{x}}^{-1} \circ T_{\mathbf{y}}(\mathbf{e}) \\
&=_{(1)} \eta \circ \left(T_{\mathbf{x}}^{-1} \circ T_{\mathbf{y}}\right)^{-1}(\mathbf{e}) \\
&= \eta \circ T_{\mathbf{y}}^{-1} \circ T_{\mathbf{x}}(\mathbf{e}) \\
&= \eta \circ T_{\mathbf{y}}^{-1}(\mathbf{x}),
\end{aligned}$$

(1) by the previous lemma. ..... □ [Claim 1]

Hence, by Claim 1, we can rewrite

$$\Gamma * \eta(\mathbf{x}) \;=\; \int_{\mathbb{S}^{D-1}} \eta_{\mathbf{x}}(\mathbf{y})\, d\Gamma[\mathbf{y}] \;=\; \int_{\mathbb{S}^{D-1}} \eta_{\mathbf{y}}(\mathbf{x})\, d\Gamma[\mathbf{y}] \;=\; \int_{\mathbb{S}^{D-1}} \eta \circ T_{\mathbf{y}}^{-1}(\mathbf{x})\, d\Gamma[\mathbf{y}]$$

Thus, for any $g \in \mathbb{SO}^D[\mathbb{R}]$,

$$\widetilde{\Gamma * \eta}(g) \;=\; (\Gamma * \eta)\,[\mathbf{P}(g)]$$

$$
\begin{aligned}
&= (\Gamma * \eta)[g(\mathbf{e})] \\
&= \int_{\mathbb{S}^{D-1}} \eta \circ T_{\mathbf{y}}^{-1}(g(\mathbf{e}))\, d\Gamma[\mathbf{y}] \\
&= \int_{\mathbb{S}^{D-1}} \eta \circ \left(T_{\mathbf{y}}^{-1} \circ g\right)(\mathbf{e})\, d\Gamma[\mathbf{y}]. \\
=_{(1)} &\int_{\mathbb{S}^{D-1}} \int_{\mathbf{P}^{-1}\{\mathbf{y}\}} \eta\left[(h^{-1} \cdot g)(\mathbf{e})\right]\, d\mathcal{H}_{\mathbf{y}}^{oar}[h]\, d\Gamma[\mathbf{y}] \\
=_{(2)} &\int_{\mathbb{SO}^D[\mathbb{R}]} \eta\left[(h^{-1} \cdot g)(\mathbf{e})\right]\, d\widetilde{\Gamma}[h] \\
&= \int_{\mathbb{SO}^D[\mathbb{R}]} \eta \circ \mathbf{P}(h^{-1} \cdot g)\, d\widetilde{\Gamma}[h] \\
&= \int_{\mathbb{SO}^D[\mathbb{R}]} \widetilde{\eta}(h^{-1} \cdot g)\, d\widetilde{\Gamma}[h] \\
&= \widetilde{\Gamma} * \widetilde{\eta}(g).
\end{aligned}
$$

(1) For all $h \in \mathbf{P}^{-1}\{\mathbf{y}\}$, we have $h.\mathbf{e} = \mathbf{y}$ by definition, so any $h \in \mathbf{P}^{-1}\{\mathbf{y}\}$ could be $T_{\mathbf{y}}$. It does not matter which $h \in \mathbf{P}^{-1}\{\mathbf{y}\}$ we pick.

(2) This is the definition of $\widetilde{\Gamma}$.

□

6.2 Deconvolution on $\mathbb{SO}^D[\mathbb{R}]$

At this point, we can employ the following method to reconstruct the spectral measure Γ.

1. Pull back the spherical log-characteristic function (section 4.1 on page 43) $\mathbf{g} : \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$ to a function $\widetilde{\mathbf{g}} : \mathbb{SO}^D[\mathbb{R}] \longrightarrow \mathbb{C}$.

2. Pull back the kernel $\eta^{(\alpha)} : \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$ to a function $\widetilde{\eta}^{(\alpha)} : \mathbb{SO}^D[\mathbb{R}] \longrightarrow \mathbb{C}$.

3. Let $\widetilde{\Gamma}$ be the (unknown) pullback of the (unknown) spectral measure Γ to a measure on $\mathbb{SO}^D[\mathbb{R}]$. Since $\mathbf{g} = \eta^{(\alpha)} * \Gamma$,, we have

$$\widetilde{\mathbf{g}} = \widetilde{\eta}^{(\alpha)} * \widetilde{\Gamma},$$

by Lemma 65 on the preceding page.

4. Now we look at the **matrix-valued Fourier coefficients** (Definitions 163 on page 232 and 169 on page 234) of $\widetilde{\eta}$, $\widetilde{\mathbf{g}}$, and $\widetilde{\Gamma}$. Suppose $\widetilde{\eta}^{(\alpha)}$ has matrix-valued Fourier coefficients $\left\{\widehat{\eta}_\phi^{(\alpha)} \ ; \ \phi \in \mathfrak{U}[\mathbb{G}]\right\}$, $\widetilde{\mathbf{g}}$ has matrix-valued Fourier coefficients $\{\widehat{\mathbf{g}}_\phi \ ; \ \phi \in \mathfrak{U}[\mathbb{G}]\}$, and $\widetilde{\Gamma}$ has matrix-valued Fourier coefficients $\left\{\widehat{\Gamma}_\phi \ ; \ \phi \in \mathfrak{U}[\mathbb{G}]\right\}$. Then, by Corollary 172 on page 235, for any $\phi \in \mathfrak{U}[\mathbb{G}]$, if $\widehat{\eta}_\phi^{(\alpha)}$ is an **invertible** matrix, then:

$$\widehat{\Gamma}_\phi \ = \ \widehat{g}_\phi \cdot \left(\widehat{\eta}_\phi^{(\alpha)}\right)^{-1}.$$

5. By the Peter-Weyl theorem (Theorem 162 on page 231), the one-dimensional components of the matrix-valued functions determining the **irreducible unitary representations** (Definition 154 on page 229) of $\mathbb{SO}^D[\mathbb{R}]$ form an orthonormal basis for $\mathbf{L}^2\left(\mathbb{SO}^D[\mathbb{R}]\right)$. Thus, in principal, knowledge of the Fourier coefficients of $\widetilde{\Gamma}$ completely determines it.

6. Given $\widetilde{\Gamma}$, we can reconstruct Γ by projection:

$$\Gamma = \ \mathbf{P}^* \widetilde{\Gamma}$$

where $\mathbf{P} : \mathbb{SO}^D[\mathbb{R}] \longrightarrow \mathbb{S}^{D-1}$ is the function defined at the beginning of this chapter.

Although this method is good in theory, there is a serious practical difficulty: we need to have an explicit enumeration of the irreducible unitary representations of $\mathbb{SO}^D[\mathbb{R}]$, and explicit expressions for the component functions. We need these explicit expressions in the first place to actually compute the Fourier coefficients of $\widetilde{\eta}$ and $\mathbf{g}$, and in the second place, to reconstruct $\widetilde{\Gamma}$ from it's Fourier coefficients by means of an infinite summation of orthogonal components.

A complete and explicit enumeration of the irreducible unitary representations of $\mathbb{SO}^2[\mathbb{R}]$ is known, and can be found in any book on nonabelian harmonic analysis (for example, [172] or [180]). In higher dimensions, however, the picture is not so clear. Thus, for practical purposes, the method of spherical Fourier analysis, developed in Chapter 5, is probably preferable.

Prior Work: Prior applications of harmonic analysis on $\mathbb{SO}^D[\mathbb{R}]$ have primarily concentrated on the case when $D = 3$, with concrete applications; see for example [83] or [178]. In particular, Healy, Hendriks, and Kim [58]

combine harmonics on $\mathbb{SO}^3[\mathbb{R}]$ with spherical harmonics on $\mathbb{S}^2$ to develop a method of "deconvolution" for probability distributions on $\mathbb{S}^2$, which they propose as a mechanism of "denoising" a $\mathbb{S}^2$-valued "signal" being corrupted by "random rotations" in the form of $\mathbb{SO}^3[\mathbb{R}]$-valued noise.

Part II

Other Statistical Methods

Chapter 7

Normal Rank Correlation Analysis

7.1 Preliminaries

The statistical theory of multivariate normal distributions is extremely well-developed. Unfortunately, the same methods do not generally apply in the non-Gaussian regime. *Normal Rank Correlation* (NRC) analysis [60] is a "trick" for making non-Gaussian distributions amenable to Gaussian analytic methods. The method is as follows:

1. Given any absolutely continuous, univariate probability distribution ρ on $\mathbb{R}$, there exists a nondecreasing, measurable bijection $f : \mathbb{R} \longrightarrow \mathbb{R}$ so that the image measure[1] $f^*\rho$ is a normal distribution with mean 0 and variance 1. In other words, f defines a "change of coordinate systems" so that, relative to the new coordinate system, the distribution ρ "appears" Gaussian.

   The construction of f is actually quite simple. Let $C : \mathbb{R} \longrightarrow [0,1]$ be the cumulative distribution function of ρ:

   $$C(x) \;=\; \int_{-\infty}^{x} \rho(t)\; dt.$$

   Thus, the probability distribution $C^*\rho$ is simply the uniform distribution on $[0,1]$. Similarly, if $D : \mathbb{R} \longrightarrow [0,1]$ is the cumulative distribution of $\mathcal{N}[0,1]$, and $\mathcal{L}^{bsg}$ is the uniform distribution on $[0,1]$, then $\left(D^{-1}\right)^* \mathcal{L}^{bsg} \;=\; \mathcal{N}[0;1]$.

[1] Here, if $S \subset \mathbb{R}$ is measurable, then $f^*\mu[S] = \mu\left[f^{-1}(S)\right]$.

Combining these facts, we conclude:

$$\left(D^{-1}\right)^* C^* \rho \;=\; \frown[0;1]$$

hence, we define $f := \left(D^{-1}\right) \circ C$.

2. Now suppose ρ is a multivariate probability distribution on $\mathbb{R}^N$, with univariate marginals $\rho_1, \rho_2, \ldots, \rho_N$ (note that ρ is *not* necessarily the product of its marginals). We can find functions $f_1, \ldots, f_N : \mathbb{R} \longrightarrow \mathbb{R}$ so that, for all $n \in [1..N]$, $f_n^* \rho_n$ has distribution $\frown[0;1]$. Define $F : \mathbb{R}^N \longrightarrow \mathbb{R}^N$ by

$$F(x_1, \ldots, x_N) \;=\; (f_1(x_1),\; \ldots,\; f_N(x_N))$$

and consider the distribution $\mu = F^* \rho$.

3. Each univariate marginal of μ is normal. Of course, this does not means that μ itself is a multivariate normal distribution, merely that it is a N-fold **coupling** of univariate normal distributions. However, we can hope that, if ρ is some "reasonable" distribution, then μ is reasonably "close" to a multivariate normal distribution —close enough that we can apply Gaussian statistical methods to μ and get useful information.

4. In particular, we hope that the **covariance matrix** of μ can tell us something about the correlations between the different dimensions of the original distribution ρ.

Of course, this methodology rests upon the assumption that μ is "close to normal" in some sense[2]. If this assumption is false, then the whole methodology falls apart. Many Gaussian statistical methods depend upon the fact that the covariance matrices of multivariate normal distributions transform in a particularly nice way under linear changes of coordinates; this is crucial, for example, for Principal Component Analysis. The distribution μ may not have the same kinds of symmetry properties, and therefore an analogous "principal components analysis" of μ may prove misleading.

There are many ways of coupling univariate normal distributions together to form a multivariate distribution that is quite far from normal. However, the multivariate normal represents the coupling with maximum

[2]For example, in total variation norm.

entropy ([173], Example 11.2.8, p.270), and thus, seems natural from the perspective of "maximum-entropy" statistical methodologies[3].

Our goal in this chapter is to test, empirically, whether normal rank correlation analysis is a useful methodology for the analysis of multivariate stable distributions. Of course, in general, a spectral measure of a stable distribution is an "infinite-dimensional" datastructure, whereas a correlation matrix has only $N(N+1)/2$ degrees of freedom. Hence, we expect the latter to be a rather inadequate representation of the former. The question is:

> *To what extent does NRC analysis reflect the correlation structure of a multivariate stable distribution?*

7.2 Experimental Method:

1. Large datasets of pseudo-random vectors are generated, according to particular multivariate stable distributions with specified spectral measures.

2. NRC analysis is performed on these datasets, to construct covariance matrices.

3. These covariance matrices are compared to the correlation structure of the original stable distribution.

To perform step 1, I developed a C software package capable of generating stably distributed pseudo-random vectors with a specified spectral measure, in any number of dimensions (see Section B.5 on page 151 for an explanation of the methodology). To perform step 2, I made use of a C++ software library, designed by Gustavo Comezana of the University of Toronto RiskLab, which performed NRC analysis on the datasets and produced covariance matrices as output.

Step 3 is vaguely formulated; part of the problem was to determine just what sort of "comparison" is appropriate. I settled for a simple methodology, wherein four families of datasets were generated, two with very high correlation, and the other two with zero correlation. The hope was that the

[3] "Maximum-entropy" methods were first developed by Jaynes [73, 71, 69], and have been applied to statistical thermodynamics [72, 70], the theory of random space-filling patterns [90, 170, 123, 34], and geophysics [22], among other things. See Chapter 11 of [173] for a good introduction.

NRC matrices would clearly exhibit high correlation in the first two families, and very low correlation in the second two families.

In each of the four families, datasets were generated in dimensions $D = 2, 3, 4$ and 5. Within each dimension, stable data was generated for stability exponents ranging from $\alpha = 2.0$ (ie. Gaussian) down to $\alpha = 0.4$, in decrements of size 0.1, excluding $\alpha = 1.0$. Thus, each family contained 14 datasets in each of 4 dimensions[4]. Each dataset contained 10 000 randomly generated vectors drawn from the relevant distribution.

The four families used were the following:

High Correlation:

- **Purely Atomic Spectral Measure:** In dimension D, atoms $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_D \in \mathbb{S}^{D-1}$ with weights $w_1, w_2, \ldots, w_D > 0$ were chosen so that, when arranged as a matrix of row vectors,

$$\begin{bmatrix} \longleftarrow & w_1 \cdot \mathbf{a}_1 & \longrightarrow \\ \longleftarrow & w_2 \cdot \mathbf{a}_2 & \longrightarrow \\ \longleftarrow & w_3 \cdot \mathbf{a}_3 & \longrightarrow \\ \vdots & \vdots & \vdots \\ \longleftarrow & w_D \cdot \mathbf{a}_D & \longrightarrow \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ 0 & 1 & 1 & \ldots & 1 \\ 0 & 0 & 1 & \ldots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix}.$$

- **Subgaussian:** In dimension D, a **subgaussian** stable distribution (see Example 81 on page 147) was generated. The underlying Gaussian had principle components $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_D$ such that, when arranged as a matrix of row vectors,

$$\begin{bmatrix} \longleftarrow & \mathbf{x}_1 & \longrightarrow \\ \longleftarrow & \mathbf{x}_2 & \longrightarrow \\ \longleftarrow & \mathbf{x}_3 & \longrightarrow \\ \vdots & \vdots & \vdots \\ \longleftarrow & \mathbf{x}_D & \longrightarrow \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ 0 & 1 & 1 & \ldots & 1 \\ 0 & 0 & 1 & \ldots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix}.$$

[4]The cases $\alpha = 1.0$ and $\alpha < 0.4$ were excluded simply because the NRC algorithm crashed or hung when processing these datasets, probably due to the large number of extreme values. For the same reason, a few other data points are missing; however, the trends in the data are obvious even with these gaps.

Low Correlation:

- **Purely Atomic Spectral Measure:** In dimension D, we used mutually orthogonal atoms $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_D \in \mathbb{S}^{D-1}$ with unit weights. Thus, when arranged as a matrix of row vectors,

$$\begin{bmatrix} \longleftarrow & w_1 \cdot \mathbf{a}_1 & \longrightarrow \\ \longleftarrow & w_2 \cdot \mathbf{a}_2 & \longrightarrow \\ \longleftarrow & w_3 \cdot \mathbf{a}_3 & \longrightarrow \\ \vdots & \vdots & \vdots \\ \longleftarrow & w_D \cdot \mathbf{a}_D & \longrightarrow \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix}.$$

- **Subgaussian:** In dimension D, the underlying Gaussian had mutually orthogonal principle components $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_D$. That is, when arranged as a matrix of row vectors,

$$\begin{bmatrix} \longleftarrow & \mathbf{x}_1 & \longrightarrow \\ \longleftarrow & \mathbf{x}_2 & \longrightarrow \\ \longleftarrow & \mathbf{x}_3 & \longrightarrow \\ \vdots & \vdots & \vdots \\ \longleftarrow & \mathbf{x}_D & \longrightarrow \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix}.$$

When $\alpha = 2.0$, the subgaussian and purely atomic datasets in each family are drawn from identical normal distributions. For $\alpha < 2.0$, however, subgaussian distributions are the "opposite" of purely atomic ones: the spectral measure of a subgaussian is absolutely continuous relative to the Lebesgue measure. Thus, the subgaussian stable datasets should be qualitatively different from the purely atomic spectral datasets; one goal of the experiment was to see if this qualitative difference was reflected in the NRC analysis.

7.3 Plots

For each family, and each dimenson, the different components of the NRC matrix were plotted as functions of the exponent α. In the plots which follow, α ranges from 0.3 to 2.0 along the horizontal axis. The covariance matrix of a multivariate Gaussian is symmetric, and the diagonal entries (the variances) are all equal to one, due to the normalization performed by NRC analysis. Hence, only the superdiagonal entries are plotted.

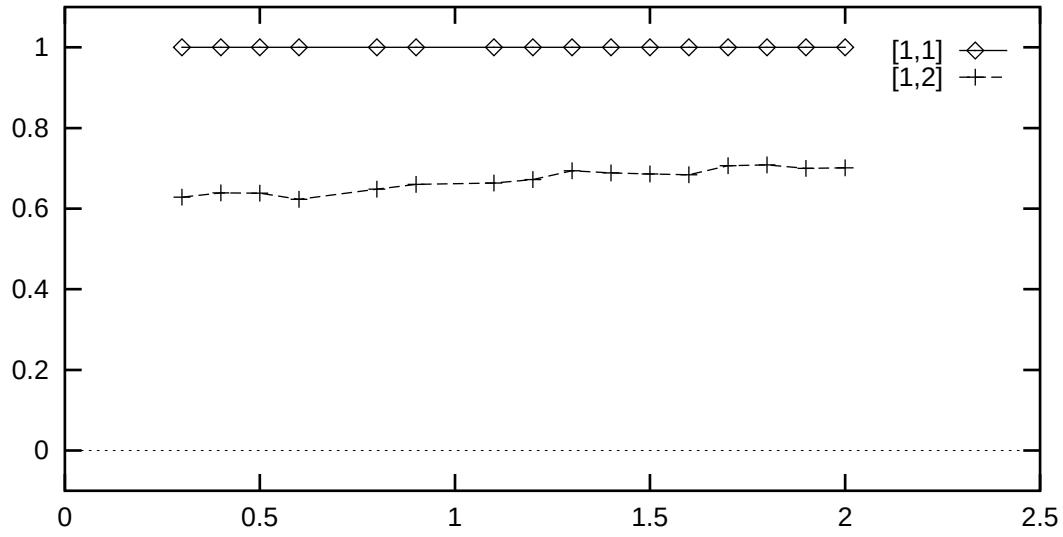

Figure 7.1: *Dimension: 2.....Atomic: High Correlation*

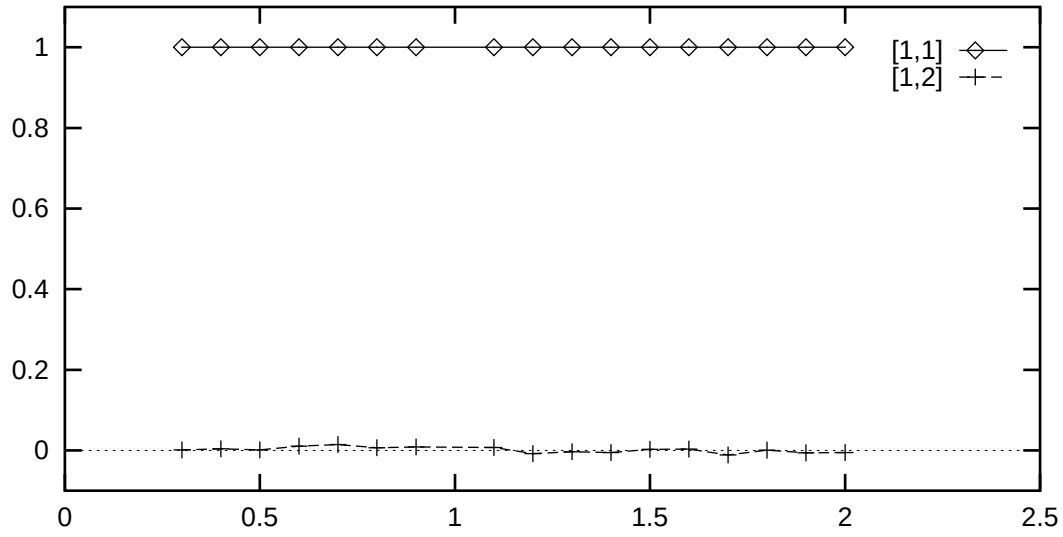

Figure 7.2: *Dimension: 2.....Atomic: Zero Correlation*

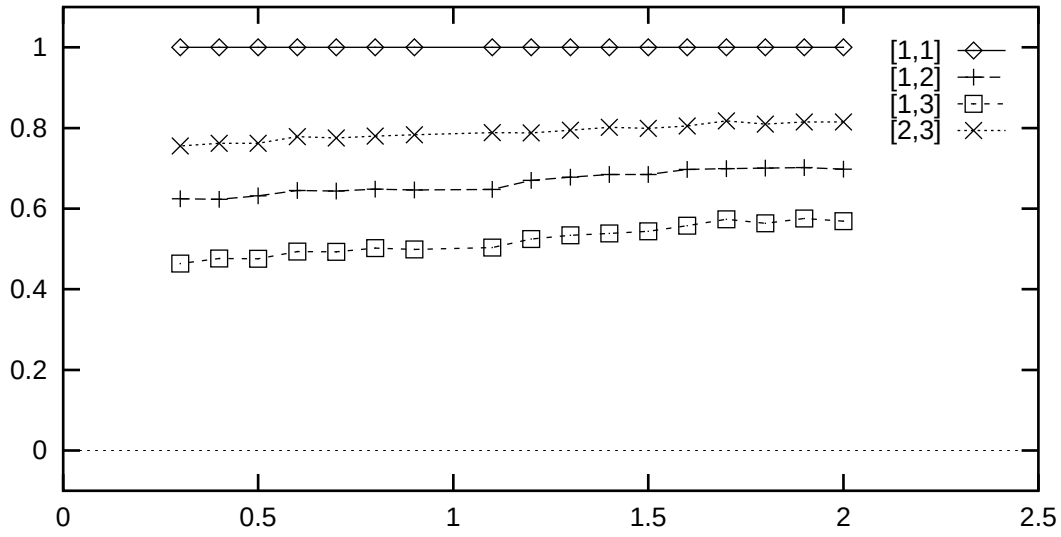

Figure 7.3: *Dimension: 3.....Atomic: High Correlation*

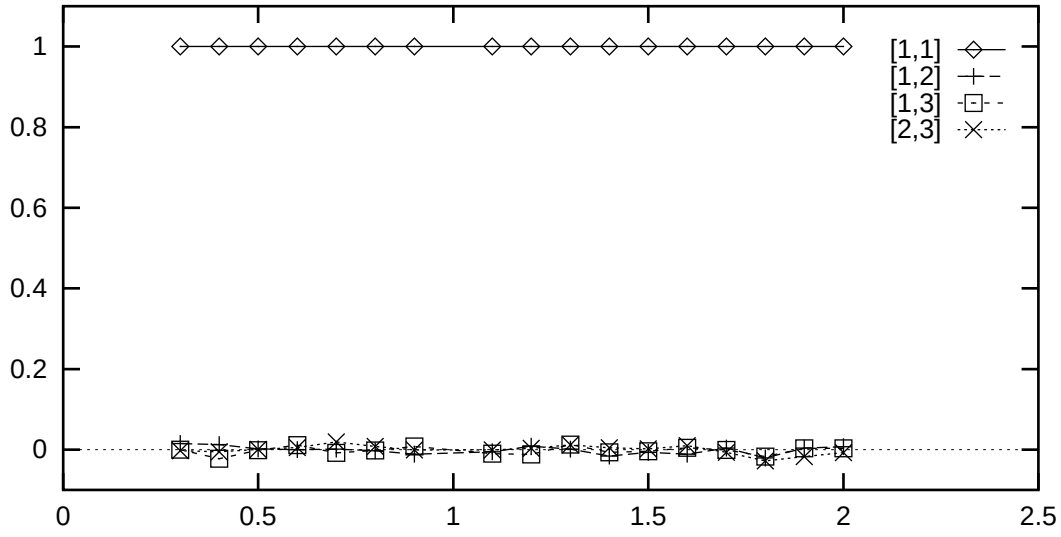

Figure 7.4: *Dimension: 3.....Atomic: Zero Correlation*

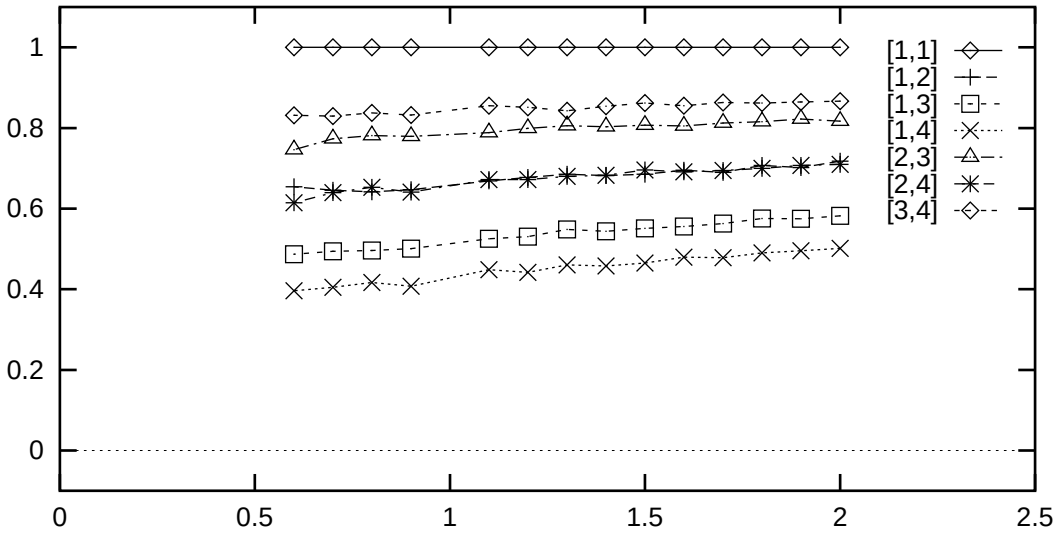

Figure 7.5: *Dimension: 4.....Atomic: High Correlation*

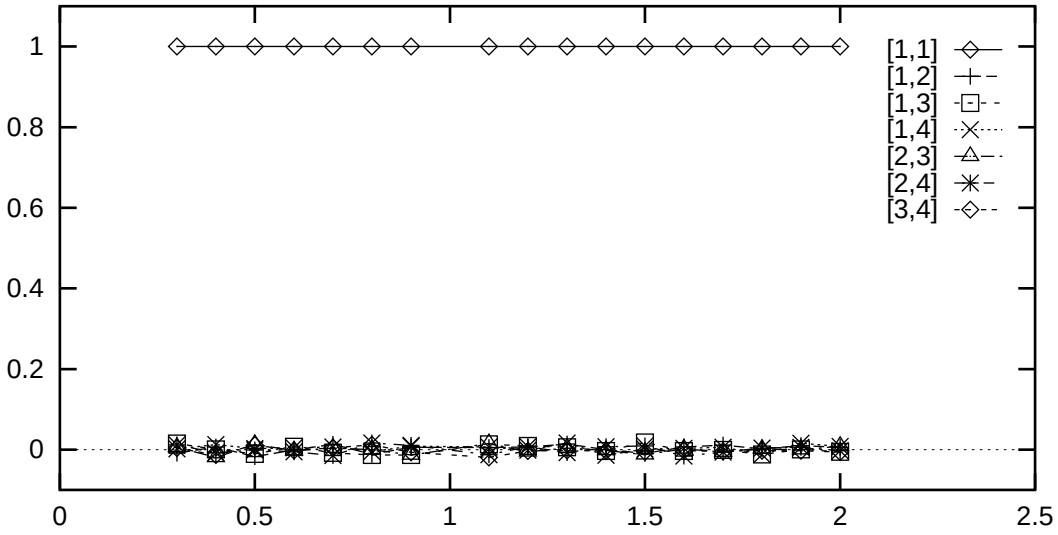

Figure 7.6: *Dimension: 4.....Atomic: Zero Correlation*

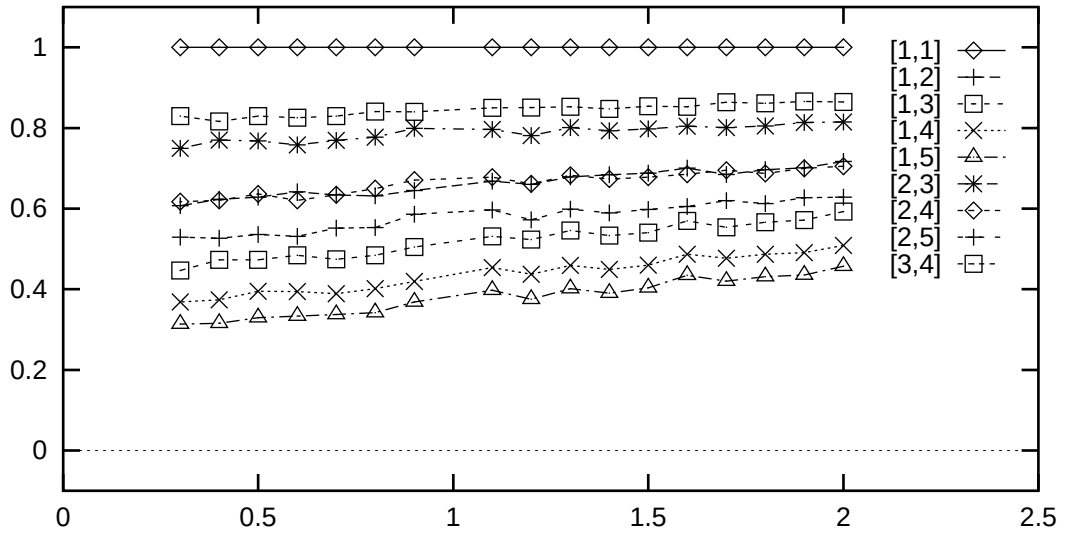

Figure 7.7: *Dimension: 5.....Atomic: High Correlation*

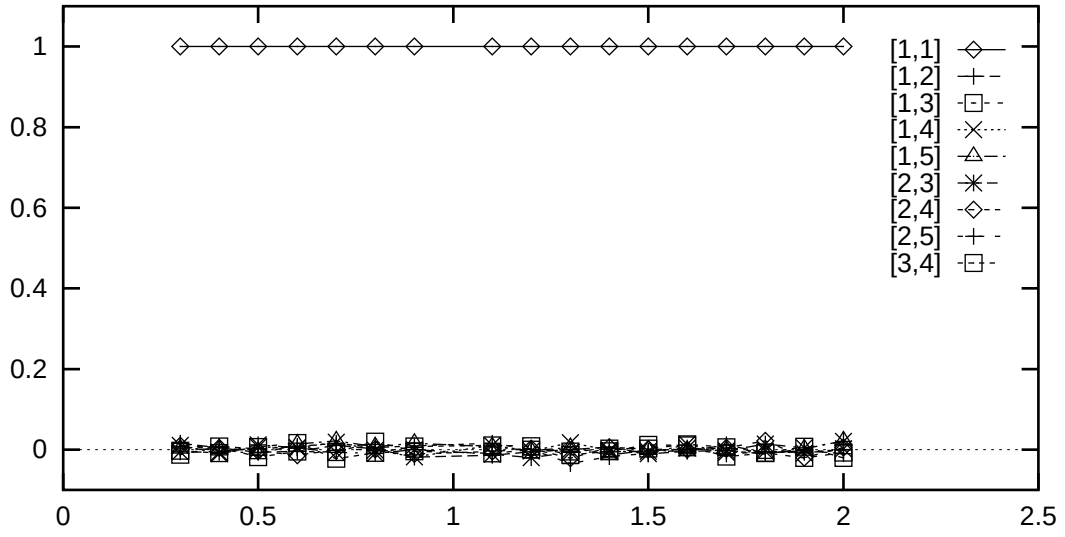

Figure 7.8: *Dimension: 5.....Atomic: Zero Correlation*

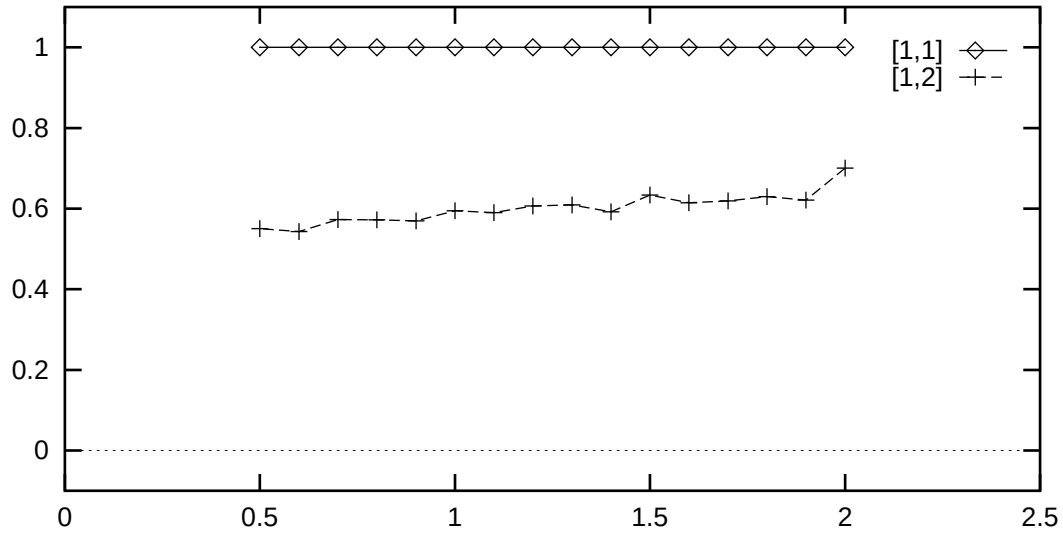

Figure 7.9: *Dimension: 2.....SubGaussian: High Correlation*

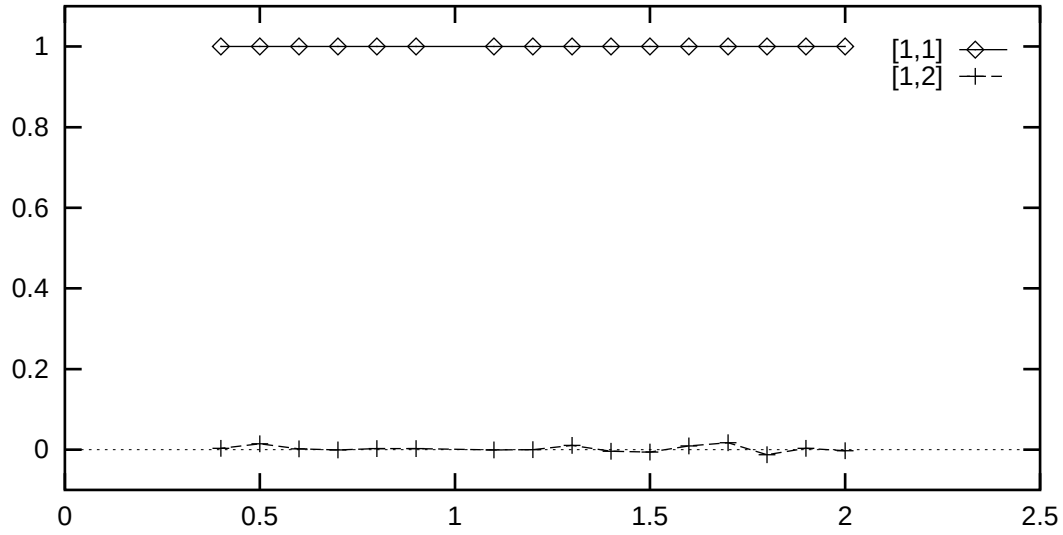

Figure 7.10: *Dimension: 2.....SubGaussian: Zero Correlation*

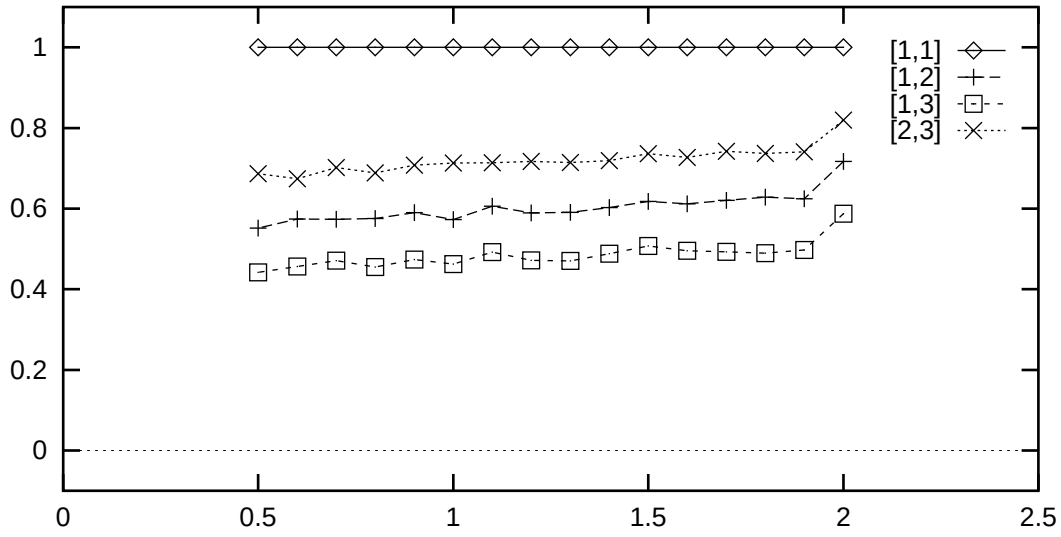

Figure 7.11: *Dimension: 3.....SubGaussian: High Correlation*

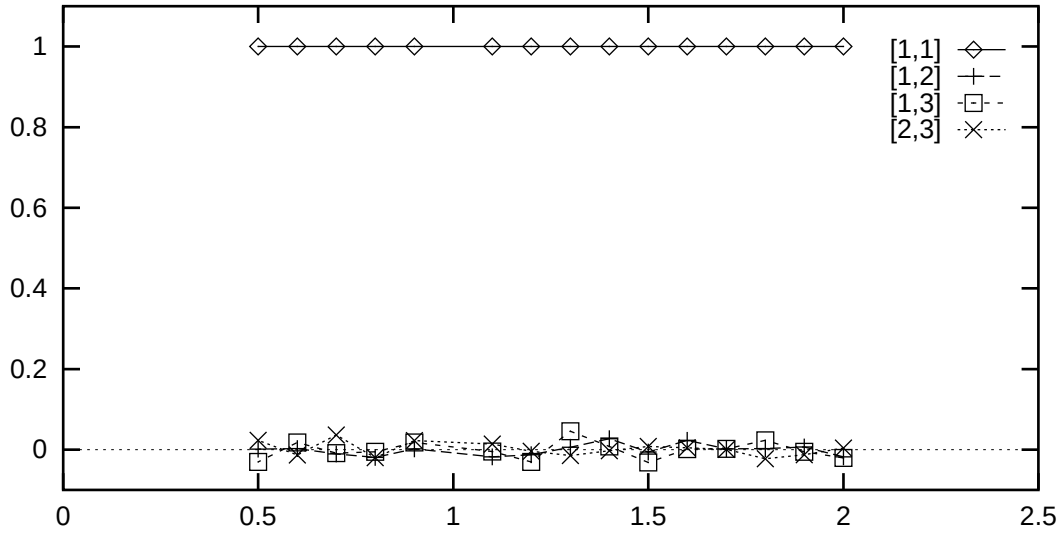

Figure 7.12: *Dimension: 3.....SubGaussian: Zero Correlation*

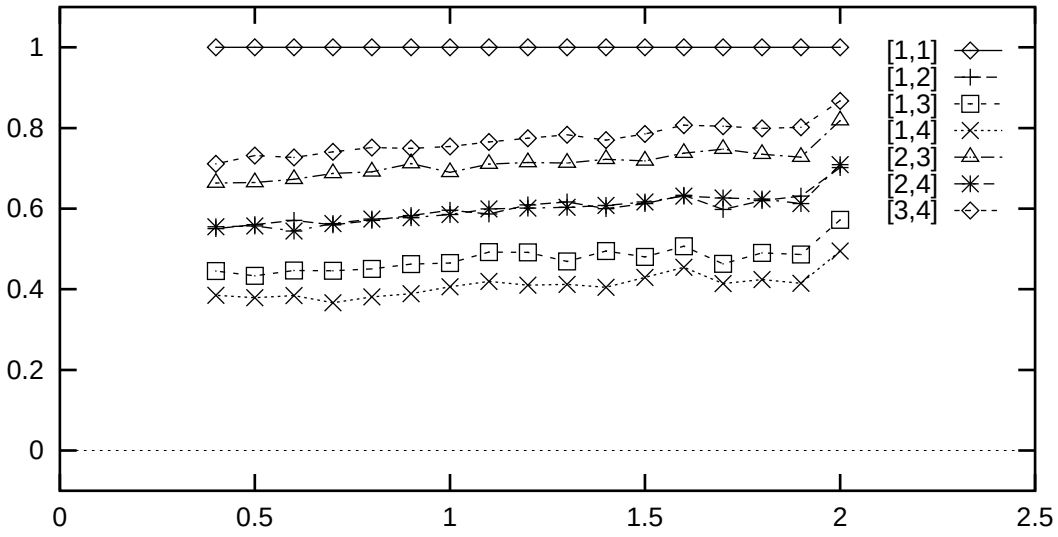

Figure 7.13: *Dimension: 4.....SubGaussian: High Correlation*

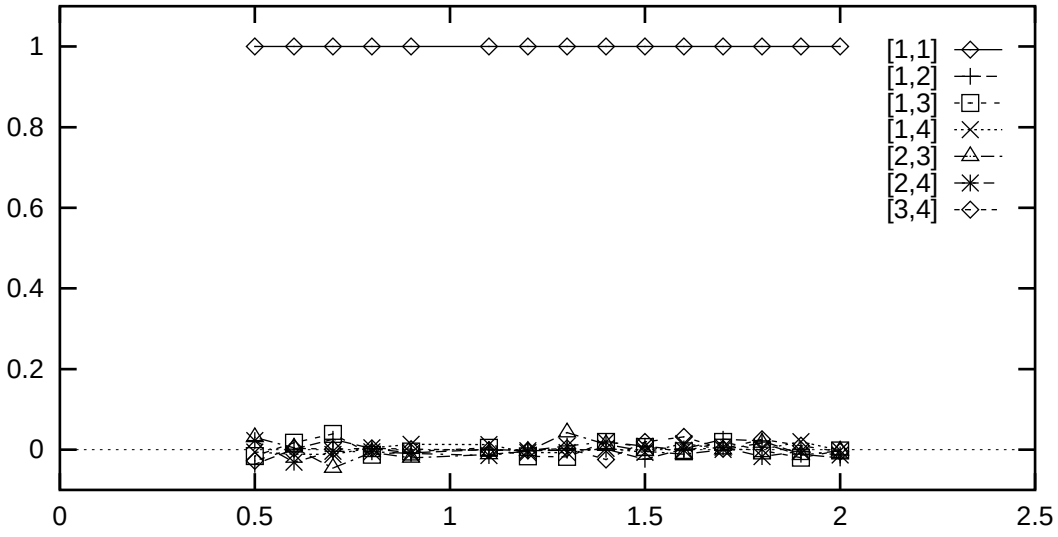

Figure 7.14: *Dimension: 4.....SubGaussian: Zero Correlation*

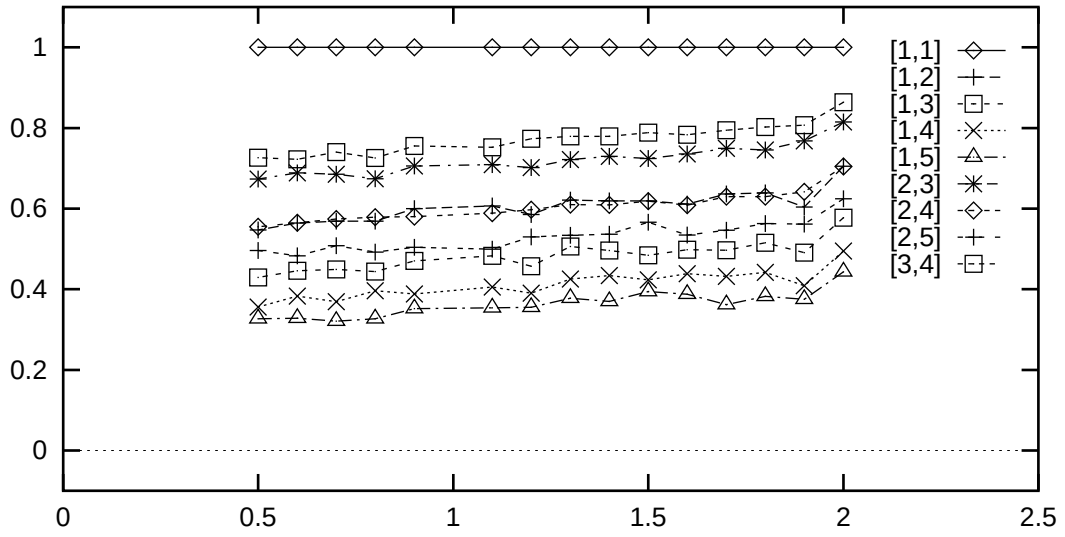

Figure 7.15: *Dimension: 5.....SubGaussian: High Correlation*

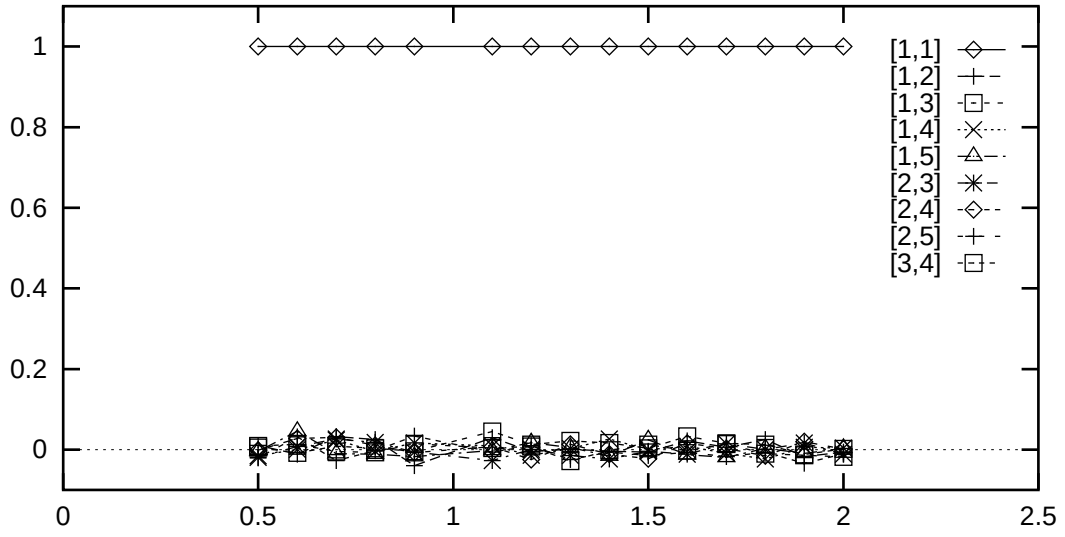

Figure 7.16: *Dimension: 5.....SubGaussian: Zero Correlation*

7.4 Observations

When $\alpha = 2.0$, the datasets really *are* Gaussian; hence, the NRC analysis is a trivial operation, and the NRC matrix should be identical with the covariance matrix of the original data. Hence, for each plot, $\alpha = 2.0$ is the "control" case, against which we compare other values.

Note that, like real-world empirical data, the data from these computer experiments is noisy, and the plots are bumpy. Theoretically, were an infinite number of datapoints generated, and an infinite-precision NRC analysis performed, we would plot perfectly smooth lines. Since only 10 000 points were generated for each dataset, however, there is a small amount of "experimental error". Nevertheless, the trends in the data are clear.

The following qualitative observations are apparent from inspection of these figures:

- In all cases, the correlation coefficients for $\alpha < 2.0$ are comparable in magnitude to the corresponding coefficients when $\alpha = 2.0$.
- In particular, if a certain coefficient is "large" when $\alpha = 2.0$ (indicating large correlation), then it is comparably large for all $\alpha < 2.0$, meaning that NRC analysis detects this correlation.

  Conversely, if a certain coefficient is zero when $\alpha = 2.0$ (indicating zero correlation), then it is very small (basically zero, plus experimental error) for all $\alpha < 2.0$, meaning that NRC analysis detects this absence of correlation.
- As α decreases, all nonzero correlation coefficients gradually and monotonically decrease. In other words, as the stable distribution gets farther away from a Gaussian, the degree of correlation between coordinates measured by NRC analysis slowly dwindles. Nonetheless, this effect is quite gradual.
- This dwindling of correlation coefficients is slightly more pronounced in the subgaussian family than in the pure-atomic spectral measure family. This is the only visible qualitative difference between the subgaussian and pure-atomic families.

Questions: These results indicate that NRC analysis is capable of detecting correlation between the different coordinates of a multivariate α-stable distribution, but it is unclear exactly what sort of correlation information NRC is extracting. In particular, what is the relationship between the NRC

matrix components and other measures of dependency, such as **covariation**, **codifference**, Kendall's τ parameter, or **James orthogonality**?

Chapter 8

Mixed Stable Distributions

Empirically measured stability exponents from real data seldom take on "mathematically canonical" values like 1, 2, or 1/2, which would suggest some mathematical principal forcing the exponent towards a particular value. Instead, they seem to have arbitrary values like 1.83 or 1.72 (see [122], [121], [181], [107], [112]). Indeed, there is wide variation in the measured values of these exponents even within particular market sectors.

This suggests that different stable stochastic phenomena (for example, different financial markets) may exhibit different exponents, and that a the stability exponent of a single phenomena (say, market sector) may evolve over time. More troubling, however, is the possibility that the values we measure may in fact represent a *mixture* of stability exponents. A single observable is usually a linear combination of many unobservable factors; if these unobservables are stable random variables with *different* exponents, then the exponent we measure will somehow be a combination of all of these. How can we separate out the different effects? The calibration tools we have developed so far were directed at a "pure" stable distribution; can we adapt these tools to a "mixture" of distributions?

Heuristic definition of mixed stable random vector: The general "mixed stable random vector" can be written heuristically as an "integral":

$$\mathbf{X} = \int_0^2 \mathbf{X}_\alpha \, d\alpha$$

where, for all $\alpha \in (0, 2]$, $\mathbf{X}_\alpha$ is an α-stable random vector, and all these random vectors are mutually independent. Suppose that $\mathbf{X}_\alpha$ had spectral measure Γ_α. Then we could write this as a stochastic integral (see Example 98 on page 188):

$$\mathbf{X} \;=\; \int_0^2 \int_{\mathbb{S}^{D-1}} \mathbf{X}_\alpha(\theta) \cdot \theta \; d\Gamma_\alpha[\theta] \; d\alpha$$

where, for all $\alpha \in (0,2]$ and $\theta \in \mathbb{S}^{D-1}$, $\mathbf{X}_\alpha(\theta)$ are independent random scalars with distribution $\curvearrowright^{\alpha}[1;1;0]$.

Let $\mathbb{S}^{D-1}(0,2]$ be the closed ball around 0 of radius 2, with the origin removed. Then we could define the measure Γ on $\mathbb{S}^{D-1}(0,2]$, in spherical coordinates, by:

$$\Gamma \;=\; \int_0^2 \int_{\mathbb{S}^{D-1}} \delta_\alpha \otimes \Gamma_\alpha[\theta] \; d\alpha$$

and then write:

$$\mathbf{X} \;=\; \int_{\mathbb{S}^{D-1}(0,2]} \mathbf{X}_\alpha(\theta)\theta \; d\Gamma[\alpha,\theta]$$

Γ is then the **mixed spectral measure** of $\mathbf{X}$.
Now we can formalize this intuition.

Definition 66: *Mixed Stable Distribution*

A **mixed stable probability distribution** *in* $\mathbb{R}^D$ *is a probability measure* ρ*, so that, if* Φ *is the log-characteristic function of* ρ*, then* Φ *is of the form:*

$$\Phi(x) \;=\; -\int_{\mathbb{S}^{D-1}(0,2]} \eta^{(\alpha)}(\vec{\xi},\theta) \; d\Gamma[\alpha,\theta]$$

where $\eta^{(\bullet)} : \mathbb{S}^{D-1}(0,2] \times \mathbb{S}^{D-1} \longrightarrow \mathbb{C}$ *is defined so that, for all* $\alpha, \theta, \vec{\xi}$,

$$\eta^{(\alpha)}(\vec{\xi},\theta) \;=\; \left|\left\langle \vec{\xi},\theta \right\rangle\right|^{\alpha} + \mathcal{B}_\alpha \mathbf{i} \left\langle \vec{\xi},\theta \right\rangle^{\langle\alpha\rangle}$$

The **Estimation problem** for a mixed stable random vector is this:

Given enough random samples drawn from distribution ρ*, estimate the mixed spectral measure* Γ.

Mixed Stable Distributions on $\mathbb{R}$

We will start with a simpler problem. Suppose that $\mathbf{X} = \sum_{n=1}^{N} \mathbf{X}_N$, where $\mathbf{X}_1, \ldots, \mathbf{X}_N$ are stable random variables with stability exponents $\alpha_1 < \alpha_2 < \ldots < \alpha_N$, respectively, and spectral measures $\Gamma_1, \ldots, \Gamma_N$. We need to determine the number of components, and their relative stability exponents, and then separately estimate each of their spectral measures.

If Φ is the log-characteristic function of $\mathbf{X}$, then

$$\Phi = \sum_{n=1}^{N} \Phi_N$$

where Φ_N is the log-characteristic function of $\mathbf{X}_N$. Thus,

$$\mathbf{re}\left[\Phi(\vec{\xi})\right] \;=\; \sum_{n=1}^{N} \left|\vec{\xi}\right|^{\alpha_n} \cdot \mathbf{re}\left[\Phi_n\left(\theta\right)\right]$$

where $\theta = \frac{\vec{\xi}}{|\vec{\xi}|}$. If we fix $\theta \in \mathbb{S}^{D-1}$, we thus have:

$$\mathbf{re}\left[\Phi(r.\theta)\right] \;=\; -\sum_{n=1}^{N} r^{\alpha_n} \cdot \mathbf{re}\left[\Phi_n\left(\theta\right)\right]$$

When r is large, the summand with the largest exponent dominates the rest. When r is very close to zero, summands with larger exponents get small more quickly; hence, it is the summand with the *smallest* exponent which dominates:

$$\begin{array}{lcl} \dfrac{\mathbf{re}\left[\Phi(r.\theta)\right]}{-r^{\alpha_N}} & \xrightarrow[r\to\infty]{} & \mathbf{re}\left[\Phi_N(\theta)\right] \\ \dfrac{\mathbf{re}\left[\Phi(r.\theta)\right]}{-r^{\alpha_1}} & \xrightarrow[r\to 0]{} & \mathbf{re}\left[\Phi_1(\theta)\right] \end{array}$$

This becomes visible explicitly when we go to a log-log plot. If we define $s = \log(r)$, then $\Phi(r.\theta) = \Phi(e^s.\theta)$, and we have:

$$\begin{array}{lcll} \log\left[-\mathbf{re}\left[\Phi(e^s.\theta)\right]\right] & \asymp & \alpha_N \cdot s & \text{as } s \to +\infty \\ \log\left[-\mathbf{re}\left[\Phi(e^s.\theta)\right]\right] & \asymp & \alpha_1 \cdot s & \text{as } s \to -\infty \end{array}$$

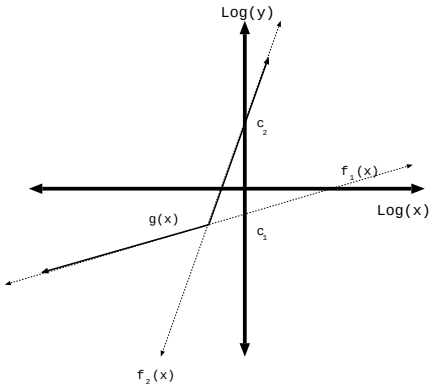

Figure 8.1: $g(x) = e^{c_1}x^{\alpha_1} + e^{c_2}x^{\alpha_2}$, with $\alpha_1 < \alpha_2$, $f_1(x) = e^{c_1}x^{\alpha_1}$, $f_2(x) = e^{c_2}x^{\alpha_2}$.

As $s \to \infty$, the function $\log[-\mathbf{re}[\Phi(e^s.\theta)]]$ will become asymptotic to the line $\alpha_N \cdot s + c_N$, where $c_N = \log[-\mathbf{re}[\Phi_N(\theta)]]$. Similarly, as $s \to \infty$, the $\log[-\mathbf{re}[\Phi(e^s.\theta)]]$ will become asymptotic to the line $\alpha_1 \cdot s + c_1$, where $c_1 = \log[-\mathbf{re}[\Phi_1(\theta)]]$. (see Figure 8.1)

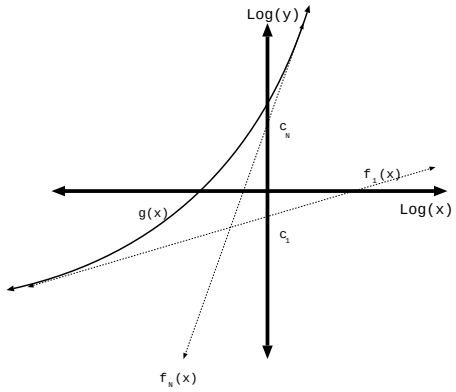

Figure 8.2: $g(x) = e^{c_1}x^{\alpha_1} + e^{c_2}x^{\alpha_2} + \ldots + e^{c_N}x^{\alpha_N}$, with $\alpha_1 < \alpha_2 < \ldots < \alpha_N$, $f_1(x) = e^{c_1}x^{\alpha_1}$, $f_N(x) = e^{c_N}x^{\alpha_N}$

If c_1 and c_N are relatively "large" compared to the values of $c_n = \log[-\mathbf{re}[\Phi_n(\theta)]]$ for $n = 2...N-1$, then the log-log plot takes a sharp "corner" at zero, and converges very rapidly to the two linear asymptotes. If c_1 and c_N are relatively "small" compared to $\{c_2, \ldots, c_{N-1}\}$, however, the plot curves very gradually from one asymptote to the other (see Figure 8.2). In

particular, it will linger at the slope α_2 if c_1 is very small relative to c_2 (see Figure 8.3). Similarly, it will linger at the slope α_{N-1} if c_N is very small relative to c_{N-1}.

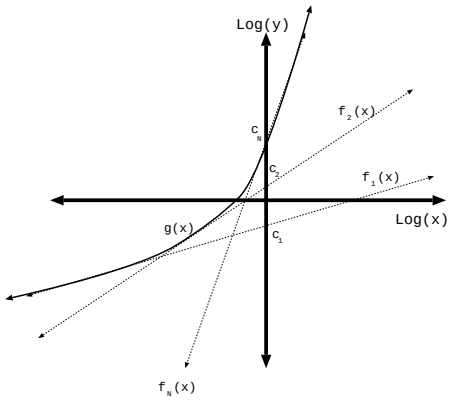

Figure 8.3: $g(x) = e^{c_1} x^{\alpha_1} + e^{c_2} x^{\alpha_2} + \ldots + e^{c_N} x^{\alpha_N}$, with $\alpha_1 < \alpha_2 < \ldots < \alpha_N$, $f_1(x) = e^{c_1} x^{\alpha_1}$, $\quad f_2(x) = e^{c_N} x^{\alpha_N}$, $\quad f_N(x) = e^{c_N} x^{\alpha_N}$

Hence, we can identify the largest and smallest exponents by looking at the linear asymptotics of the log-log plot of $\Psi_1(r) = -\mathbf{re}\,[\Phi(r.\theta)]$. We can then estimate α_1, α_N, c_1, and c_N. Suppose we remove all or most of these summands from the log-characteristic function, and consider the function $\Psi_2(r) = -\mathbf{re}\,[\Phi(r.\theta)] - e^{c_1} r^{\alpha_1} - e^{c_N} r^{\alpha_N}$. If we look at the log-log plot of Ψ_2, we can then identify α_2, α_{N-1}, c_2, and c_{N-1}. We can then remove these summands, and consider Ψ_3, etc.

Proceeding this way we should be able to isolate each of $\alpha_1, \ldots, \alpha_N$ and $c_1, \ldots, c_N$. Of course, identifying the value of c_n and α_n from a log-log plot of Ψ_n will require already having a very accurate estimate of c_{n-1} and α_{n-1}, so the success of each stage of this procedure depends upon the previous stage.

Also, to build good log-log plots, we must have an extremely accurate estimate of $\mathbf{re}\,[\Phi]$. Since $\mathbf{re}\,[\Phi] = \log|\chi|$, and $|\chi| < 1$ and tending to zero, small errors in the estimation of χ can become large errors in the estimation of $\mathbf{re}\,[\Phi]$. Fortunately, estimates of χ converge relatively quickly (see § 4.4 on page 47 for a brief discussion). Nonetheless, this will still be a problem.

Mixed Stable Distributions on $\mathbb{R}^D$

So, suppose that we are able to accurately estimate $\alpha_1, \ldots, \alpha_N$ and $c_1, \ldots, c_N$ at some point $\theta \in \mathbb{S}^{D-1}$. We can repeat this procedure for all elements $\theta \in \Theta$, where $\Theta \subset \mathbb{S}^{D-1}$ is some finite mesh, to get $\alpha_1(\theta), \ldots, \alpha_{N(\theta)}(\theta)$ and $c_1(\theta), \ldots, c_{N(\theta)}(\theta)$ for each $\theta \in \Theta$. Furthermore, for each $\theta \in \Theta$, we can extract the values $\Phi_1(\theta), \ldots, \Phi_{N(\theta)}(\theta)$ by similar means,

Let $\mathfrak{A} = \{\alpha_k(\theta) \; ; \; \theta \in \Theta; \; k \in \mathbb{N}\}$ be the set of all stability exponents detected. For each $\alpha \in \mathfrak{A}$, let Θ_α be the set of all points where α appeared, and let $\mathbb{V}_\alpha$ be the linear span of Θ_α. Thus, an α-stable random vector $\mathbf{X}_\alpha$ "lives in" the subspace $\mathbb{V}$, and we have samples $\{\Phi_k(\theta) \; ; \; \theta \in \Theta_\alpha\}$ of its SLCF. At this point, we can apply the methods of Part II to estimate the spectral measure of $\mathbf{X}_\alpha$.

Chapter 9

Statistical Analysis of Stable Stochastic Processes

In previous chapters, we developed methods for estimating a multivariate stable probability distribution from sample data. However, stably distributed data rarely appears as a sequence of independent random variables drawn from a common distribution. More often, it appears as the output of a *stable stochastic process*, with a potentially complex temporal correlation structure. Even if the datapoints of such a process are identically stably distributed, they are unlikely to be independent. Thus, to understand such a process, we must estimate not only the multivariate distribution of individual data points in isolation, but also their correlation structure over time.

In this chapter, we will show how the "static" methods developed in previous chapters for estimating a single multivariate distribution can be applied to the "dynamic" problem of identifying a stable stochastic process. The key idea is that a finite-dimensional marginal of such a stochastic process can itself be regarded as a (very high-dimensional) multivariate stable distribution; by studying the correlation structure of this distribution, we gain insight into the dynamics of the process.

Stable stochastic processes usually arise as *stochastic integrals*. We assume that there is a unknown source of random, uncorrelated, stably distributed "noise", and that the stochastic process in question is the output of some mechanism driven by this noise. If the process is treated as a random

function $\mathcal{X} : \mathbb{R} \longrightarrow \mathbb{R}^D$, then we can write:

$$\mathcal{X}(T) \;=\; \int_{\infty}^{T} \phi(T,t)\; d\mathcal{N}^{\alpha}$$

where $\mathcal{N}^{\alpha}$ is a "random measure" representing the noise, and ϕ is some integral kernel describing the way the noise drives the process. For example, if $\mathcal{B}$ is classical Brownian motion, then we can write:

$$\forall\, T \geq 0, \quad \mathcal{B}(T) \;=\; \int_{0}^{T} d\mathcal{N}^{2}$$

where $\mathcal{N}^2$ is Gaussian "white noise". In this case, the integral kernel is trivial, indicating that the Brownian particle is memoryless and its dynamics are spacially homogeneous. A more complex process is the Ornstein-Uhlenbeck process $\mathcal{X}$:

$$\forall\, T \in \mathbb{R}, \quad \mathcal{B}(T) \;=\; \int_{\infty}^{T} e^{-\lambda(T-t)}\; d\mathcal{N}^{2}$$

which describes the behaviour of a Brownian particle trapped in an exponential potential well. The Ornstein-Uhlenbeck process is mean-reverting; the particle rarely wanders far from zero.

Brownian motion and the Ornstein-Uhlenbeck process are attractive because they are the solutions of simple stochastic differential equations having natural physical intepretations. The SDE for Brownian motion is

$$d\mathcal{B} \;=\; \mathcal{N}^{2}$$

which basically says that the Brownian particle is randomly kicked around by independent, normally distributed perturbations, and moves for no other reason. The SDE for Ornstein-Uhlenbeck is

$$d\mathcal{X} \;=\; -\lambda\mathcal{X} + \mathcal{N}^{2}$$

which says that the Ornstein-Uhlenbeck process is a simple linear dynamical system with an attractor at zero, being jostled by independent normally distributed perturbations. Other stochastic processes have more complicated stochastic integral formulations, which are harder to explain in terms of simple physical models. Thus, it is desirable to find ways of decomposing more complex processes into simple components like Brownian or Ornstein-Uhlenbeck processes. We will return to this in §9.3.

The α-stable analog of Brownian motion is called *Lévy -stable motion*, and arises by replacing Gaussian white noise with α-stable noise in the

aforementioned stochastic integral. Likewise, there is an α-stable analog to the Ornstein-Uhlenbeck process. A more detailed and lengthy introduction to stable stochastic processes can be found in Appendix D on page 177.

So, suppose $\mathcal{X} : \mathbb{R} \longrightarrow \mathbb{R}^D$ is a stochastic process, determined by stochastically integrating a kernel ϕ against random noise. Given a discrete-time empirical sample $\{\mathbf{X}_n \; ; \; -N \leq n \leq N\}$ of the process, we might wish to determine whether $\mathcal{X}$ is an α-stable process, and if so, *which* process it is. The following are natural questions:

1. What is the shape of the **integration kernel** ϕ? (This is the topic of §9.1).

2. What is the value of the **stability exponent** α (if any)?

3. If $\mathcal{X}$ is a multivariate process, generated by α-stable multivariate noise, then what does the **spectral measure** of the generating α-stable noise look like?

   (These two questions are addressed in §9.2.)

4. Can we find a natural representation of the process as a combination of "elementary" processes, having simple physical explanations? (This is the topic of §9.3.)

9.1 Reconstructing the Kernel of a Moving Average Process

Suppose $\mathcal{X} : \mathbb{R} \longrightarrow \mathbb{R}$ is a univariate stable stochastic process, defined by a **moving average**[1] stochastic integral:

$$\mathcal{X}(T) \;=\; \int_{\mathbb{R}} \Phi(T-t)\; d\mathcal{N}^{\alpha}[t]$$

where $\Phi : \mathbb{R} \longrightarrow \mathbb{R}$ is some integrable "convolution kernel", and $\mathcal{N}^{\alpha}$ is a totally skewed (ie. $\beta = 1$), α-stable random noise on $\mathbb{R}$ of constant intensity. We want to reconstruct Φ from sample data.

First, we will simplify the problem by replacing $\mathcal{X}$ by a discrete-time approximation (see Appendix D.4 on page 196 for details). This is not only much simpler, mathematically speaking, but is also reasonable from a

[1] See § D.3 on page 193.

pragmatic point of view: all real-life computations will be performed using discrete-time data and discrete-time approximations of the process.

So, let $\mathbf{X} : \mathbb{Z} \longrightarrow \mathbb{R}$ be a discrete-time, stationary, stable stochastic process, given by the **moving average**:

$$\forall T \in \mathbb{Z}, \quad \mathbf{X}(T) \;=\; \sum_{n \in \mathbb{Z}} \phi(T-n) \cdot \mathcal{N}_\alpha(n)$$

where $\left[\mathcal{N}_\alpha(n)|_{n\in\mathbb{Z}}\right]$ is an i.i.d. sequence of α-stable random variables with skewness $\beta = 1$ (the discrete time analog of a totally skewed α-stable noise), and where $\phi : \mathbb{Z} \longrightarrow \mathbb{R}$ is some "convolution kernel".

Kernel Reconstruction Problem: Given empirical measurements of $\mathbf{X}$, can we reconstruct the convolution kernel ϕ?

9.1.1 The Spectral Measure of a Process Marginal

Fix $M \in \mathbb{N}$, and for all $t \in \mathbb{Z}$, define

$$\mathbf{X}^{[M]}(t) \;=\; [\mathbf{X}(t),\ \mathbf{X}(t+1),\ \ldots, \mathbf{X}(t+M)] \in \mathbb{R}^{M+1},$$

the $(M+1)$-dimensional marginal of $\mathbf{X}$ at time t. Thus, $\mathbf{X}^{[M]}(t)$ is an $(M+1)$-dimensional α-stable random vector. Since the process is stationary, the distribution of $\mathbf{X}^{[M]}(t)$ is the same for all t, and has some spectral measure $\Gamma^{[M]}$. What does $\Gamma^{[M]}$ look like? Fix $t = 0$, and recall Example 98 on page 188, which says

$$\mathbf{X}^{[M]}(0) \quad = \quad \int_{\mathbb{S}^M} \mathbf{s}\; d\mathcal{N}_{\Gamma^{[M]}}[\mathbf{s}]$$

where $\mathcal{N}_{\Gamma^{[M]}}[\mathbf{s}]$ is the α-stable random noise on $\mathbb{S}^M$ having intensity measure $\Gamma^{[M]}$ and skewness $\beta = 1$. Thus, we can write:

$$\begin{aligned}
\int_{\mathbb{S}^M} \mathbf{s}\; d\mathcal{N}_{\Gamma^{[M]}}[\mathbf{s}] \quad &= \quad \mathbf{X}^{[M]}(0) \\
&= \quad [\mathbf{X}(0),\ \ldots, \mathbf{X}(M)] \\
&= \quad \left[\sum_{n\in\mathbb{Z}} \phi(-n)\cdot\mathcal{N}_\alpha(n),\ \ \ldots,\ \ \sum_{n\in\mathbb{Z}} \phi(M-n)\cdot\mathcal{N}_\alpha(n)\right] \\
&= \quad \sum_{n\in\mathbb{Z}} [\phi(-n),\ \ \ldots,\ \ \phi(M-n)]\cdot\mathcal{N}_\alpha(n)
\end{aligned}$$

$$
\begin{aligned}
&=_{(1)} \quad \sum_{n\in\mathbb{Z}} r(n)\mathbf{s}(n)\cdot\mathcal{N}_\alpha(n) \\
&=_{(2)} \quad \int_{\mathbb{S}^M} \mathbf{s}\; d\mathcal{N}_\Delta[\mathbf{s}]
\end{aligned}
$$

(1) Here, $r(n) = \|\phi(-n),\ \ldots,\ \phi(M-n)\|_2$,
and $\mathbf{s}(n) = \frac{1}{r(n)}\left[\phi(-n),\ \ldots,\ \phi(M-n)\right]$.

(2) Where $\Delta = \sum_{n\in\mathbb{Z}} r(n)\delta_{\mathbf{s}(n)}$ is a sum of point masses.

Hence, we conclude: $\Gamma^{[M]} = \Delta = \sum_{n\in\mathbb{Z}} r(n)\delta_{\mathbf{s}(n)}$ (9.1)

Example 67: *(Ornstein-Uhlenbeck Process)*

Consider the discrete-time analog of an Ornstein-Uhlenbeck process (see § D.3 on page 192), where the kernel is defined

$$
\phi(n) := \begin{cases} \exp(-\lambda\cdot n) & if \quad n \geq 0 \\ 0 & if \quad n < 0 \end{cases}.
$$

Thus, for any $n > M$,

$$
\begin{aligned}
r(n) &= \|\phi(-n),\ \ldots,\ \phi(M-n)\|_2 \\
&= \left(e^{-2\lambda\cdot n} + e^{-2\lambda\cdot(n-1)} + \ldots + e^{-2\lambda\cdot(n-M)}\right)^{1/2} \\
&= e^{-\lambda\cdot n}\left(1 + e^{2\lambda} + \ldots + e^{2\lambda\cdot M}\right)^{1/2},
\end{aligned}
$$

$$
\begin{aligned}
\text{while } \mathbf{s}(n) &= \frac{1}{r(n)}\left[\phi(-n),\ \ldots,\ \phi(M-n)\right] \\
&= \frac{\left[e^{-\lambda\cdot n},\ e^{-\lambda\cdot(n-1)},\ \ldots.\ e^{-\lambda\cdot(n-M)}\right]}{e^{-\lambda\cdot n}\left(1 + e^{-2\lambda} + \ldots + e^{2\lambda\cdot M}\right)^{1/2}} \\
&= \frac{\left[1, e^{2\lambda},\ \ldots,\ e^{2\lambda\cdot M}\right]}{\left(1 + e^{2\lambda} + \ldots + e^{2\lambda\cdot M}\right)^{1/2}}.
\end{aligned}
$$

Thus, $\mathbf{s}(M) = \mathbf{s}(M+1) = \mathbf{s}(M+2) = \ldots$.

Meanwhile, if $n < 0$, then $r(n) = 0$ and $\mathbf{s}(n)$ is not well-defined; effectively, there are no atoms corresponding to these points.

Finally, if $n \in [0..M]$, then

$$\begin{aligned} r(n) &= \|\phi(-n),\ \ldots,\ \phi(M-n)\|_2 \\ &= \left\| \left[e^{-\lambda \cdot n},\ e^{-\lambda \cdot (n-1)},\ \ldots, e^{-2\lambda}, 1, 0, \ldots, 0 \right] \right\|_2 \\ &= \left(e^{-2\lambda \cdot n} + e^{-2\lambda \cdot (n-1)} + \ldots + e^{-2\lambda} + 1 \right)^{1/2}, \end{aligned}$$

$$\begin{aligned} \text{while } \mathbf{s}(n) &= \frac{1}{r(n)} \left[\phi(-n),\ \ldots,\ \phi(M-n) \right] \\ &= \frac{\left[e^{-\lambda \cdot n},\ e^{-\lambda \cdot (n-1)},\ \ldots, e^{-2\lambda}, 1, 0, \ldots, 0 \right]}{\left(e^{-2\lambda \cdot n} + e^{-2\lambda \cdot (n-1)} + \ldots + e^{-2\lambda} + 1 \right)^{1/2}}. \end{aligned}$$

9.1.2 Recovering the kernel from the spectral measure

So, if a large set of samples of $\mathbf{X}^{[M]}$ is analyzed using the methods of **Part II**, and a spectral measure $\Gamma^{[M]}$ is extracted, it should appear to be a countable sum of point masses, as in expression (9.1). The **Kernel Reconstruction Problem** can then be reformulated:

Given $\Gamma^{[M]}$, can we reconstruct ϕ?

To go about doing this, observe the following properties of the points $\mathbf{s}(n)$ $(n \in \mathbb{Z})$ defined in §9.1.1.

Lemma 68: *For all $n \in \mathbb{Z}$, let $\mathbf{s}(n) = (s_0(n),\ s_1(n), \ldots, s_M(n))$. Then:*

1. $\|\mathbf{s}(n)\|_2 = 1$.
2. *For all $m \in [1..M]$,* $\dfrac{s_m(n+1)}{s_1(n+1)} = \dfrac{s_{m-1}(n)}{s_0(n)}$.
3. *Thus, if we define* $\xi_m(n) = \dfrac{s_{m-1}(n)}{s_0(n)}$ *for all $n \in \mathbb{Z}$ and $m \in [1..M]$, then we can write:*

$$\begin{aligned} \mathbf{s}(n+1) = [&s_0(n+1),\ \ s_1(n+1),\ \ s_1(n+1)\xi_2(n),\ \ s_1(n+1)\xi_3(n), \\ &\ldots,\ s_1(n+1)\xi_M(n)] \end{aligned}$$

or, defining $\boldsymbol{\xi}(n) = (\xi_1(n), \xi_2(n), \ldots, \xi_M(n)) = (1, \xi_2(n), \ldots, \xi_M(n))$, *we can write:*

$$\mathbf{s}(n+1) = [s_0(n+1), \quad s_1(n+1) \cdot \boldsymbol{\xi}(n)]$$

4. *Thus, given* $s_0(n+1)$, *we have:*

$$(s_1(n+1))^2 = \frac{1 - (s_0(n+1))^2}{\|\boldsymbol{\xi}(n)\|_2^2}$$

Proof: **Part 1** is because $\mathbf{s} \in \mathbb{S}^M$ by definition.

Proof of Part 2: By definition, $s_m(n) = \frac{1}{r(n)}\phi(m-n)$, hence

$$\begin{aligned}
\frac{s_m(n+1)}{s_1(n+1)} &= \frac{\frac{1}{r(n+1)}\phi(m-n-1)}{\frac{1}{r(n+1)}\phi(1-n-1)} \\
&= \frac{\phi(m-n-1)}{\phi(-n)} \\
&= \frac{\frac{1}{r(n)}\phi(m-1-n)}{\frac{1}{r(n)}\phi(-n)} \\
&= \frac{s_{m-1}(n)}{s_0(n)}
\end{aligned}$$

Part 3 follows immediately from **Part 2**.

Part 4 follows from **Part 3** and **Part 1**.

□

So, suppose that $\mathbb{A} \subset \mathbb{S}^N$ is a countable collection of atoms, $\gamma : \mathbb{A} \longrightarrow \mathbb{R}$ is some function, and

$$\Gamma = \sum_{\mathbf{a} \in \mathbb{A}} \gamma(\mathbf{a}) \delta_{\mathbf{a}}$$

An inductive algorithm to reconstruct ϕ from Γ is as follows:

1. Choose $\mathbf{s}(0) \in \mathbb{A}$ arbitrarily, and define $r(0) = \gamma(\mathbf{s}(0))$.

2. Suppose inductively that you have $\mathbf{s}(n)$ for some $n \in \mathbb{N}$. Define $\xi(n)$ as in the previous lemma. To identify $\mathbf{s}(n+1)$, search for an element

$a_0 \in [-1, 1]$ so that, if we define

$$\begin{aligned} a_1^2 &= \frac{1 - a_0^2}{\|\boldsymbol{\xi}(n)\|_2^2}, \\ \text{and } \mathbf{a} &= [a_0, \quad a_1 \cdot \boldsymbol{\xi}(n)], \end{aligned}$$

then $\mathbf{a} \in \mathbb{A}$.

If such an element exists, define $\mathbf{s}(n+1) = \mathbf{a}$, and then define $r(n+1) = \gamma(\mathbf{s}(n+1)$.

If more than one such element exists, pick one arbitrarily. If no such element exists, then $\mathbf{s}(n)$, as currently defined, cannot be correct; backtrack and choose another value for $\mathbf{s}(n)$.

If only one choice for $\mathbf{s}(n)$ is available at each stage, then this algorithm is simply an inductive construction. If more than one element is available, then this algorithm takes on the character of a depth-first search. This latter possibility is unlikely, however, because the unknown element $\mathbf{a}$ has only one free parameter, a_0, plus a choice of sign for a_1.

One can construct the values for $\mathbf{s}(n)$ and $r(n)$ for $n < 0$ in a similar fashion. When the construction is finished, one should have "used up" all the elements in $\mathbb{A}$; in other words, we should have:

$$\mathbb{A} = \{\mathbf{s}(n) \; ; \; n \in \mathbb{Z}\}.$$

If this is not the case, then a wrong choice of $\mathbf{s}(n)$ was made at some stage, or possibly we are wrong in the hypothesis that Γ is the spectral measure of a moving average process.

Note that, if we are dealing with an Ornstein-Uhlenbeck process, as in Example 67 on page 117, then we will end up with a situation where

1. For $n < 0$, $\mathbf{s}(n)$ does not exist.
2. $\mathbf{s}(0), \; \mathbf{s}(1), \; \ldots, \; \mathbf{s}(M)$ are distinct.
3. $\mathbf{s}(M) = \mathbf{s}(M+1) = \mathbf{s}(M+2) = \ldots$.

In general,

- Condition (1) will occur if and only if $\phi(n) = 0, \; \forall n < 0$ —in other words, whenever the process is "nonanticipating", in the sense that the integration kernel ϕ only "sees" past noise, not future noise.

- Condition (3) will occur if and only if, for all $n > 0$, $\frac{\phi(n)}{\phi(n+1)} = \frac{\phi(n+1)}{\phi(n+2)}$ —in other words, if $\phi(n) = c \cdot \ell^n$ for some constants c and ℓ. Setting $\lambda = -\ln(\ell)$, we see that ϕ is just a scalar multiple of the Ornstein-Uhlenbeck kernel.

Conclusion: The spectral measure $\Gamma^{[N]}$ immediately gives us a method for recognizing **nonanticipating** processes in general, and the **Ornstein-Uhlenbeck** method in particular. For any other arbitrary moving-average process, $\Gamma^{[N]}$ provides us with a dataset which we can use to reconstruct an approximation of the convolution kernel ϕ through an inductive, depth-first search algorithm.

Previous Research: John Nolan ([128], §6) briefly discusses the relationship between the convolution kernel and the spectral measure of a 2-time marginal of a process, concentrating on the example of a **real harmonizable process:**

$$\mathcal{X}(T) \;=\; \mathbf{re}\left[\int_{\mathbb{R}} e^{\mathbf{i}T\cdot t}\, d\mathcal{N}^{\alpha}[t]\right]$$

where $\mathcal{N}^{\alpha}$ is a complex-valued, rotationally invariant spectral measure. He makes reference to further results in [126], but this literature is not yet available.

9.2 Finding the Stability Exponent and Spectral measure:

Proposition 69: *Let $\alpha \in (0,2)$, let Γ be a spectral measure on $\mathbb{S}^{D-1}$, and let $\mathcal{N}_{\Gamma}^{\alpha}$ be the corresponding α-stable noise. Let $\phi : \mathbb{R}\times\mathbb{R} \longrightarrow \mathbb{R}$ be any integration kernel, and let $\mathcal{X}$ be the $\mathbb{R}^D$-valued stochastic process so that, for all T,*

$$\mathcal{X}(T) \;=\; \int_{-\infty}^{\infty} \phi(T,t)\; d\mathcal{N}^{\alpha}[t].$$

Then for all $T \in \mathbb{R}$, $\mathcal{X}(T)$ is an α-stable random vector, whose spectral measure is

$$C_{+}(T)\cdot\Gamma \;+\; C_{-}(T)\cdot\Gamma_{-},$$

where Γ_{-} is the **spherical inversion** *of Γ, defined*

$$\forall\, U \subset \mathbb{S}^{D-1}, \qquad \Gamma_{-}[U] = \Gamma[-U]$$

and where $C_+(T)$ and $C_-(T)$ are the real constants:

$$C_+(T) = \int_{-\infty}^{\infty} \phi(T,t)_+^\alpha \, dt, \qquad C_-(T) = \int_{-\infty}^{\infty} \phi(T,t)_-^\alpha \, dt$$

$$\left(\textit{Where } x_+ = \begin{cases} x & if \quad x > 0 \\ 0 & if \quad x \leq 0 \end{cases}, \text{ and } x_- = \begin{cases} -x & if \quad x < 0 \\ 0 & if \quad x \geq 0 \end{cases} \right)$$

Proof: This follows immediately from Proposition 100 on page 189. □

Corollary 70: *(Symmetric Noise)*
Suppose that Γ is symmetric (ie. $\Gamma_- = \Gamma$). Then for all $T \in \mathbb{R}$, $\mathcal{X}(T)$ is an α-stable random vector with spectral measure $C(T) \cdot \Gamma$, where

$$C(T) = \int_{-\infty}^{\infty} |\phi(T,t)|^\alpha \, dt$$

Corollary 71: *(Spectral measure of moving-average process)*
Suppose $\mathcal{X}$ is a moving average process of the form:

$$\mathcal{X}(T) = \int_{-\infty}^{\infty} \phi(T-t) \, d\mathcal{N}_\Gamma^\alpha[t]$$

where $\phi : \mathbb{R} \longrightarrow [0,\infty)$ is nonnegative. Then for any $T \in \mathbb{R}$, the random vector $\mathcal{X}(T)$ is α-stable, with spectral measure $C \cdot \Gamma$, where

$$C = \int_{-\infty}^{\infty} \phi(t)^\alpha \, dt.$$

Corollary 72: *(Spectral measure and kernel of accumulation process)*
Suppose $\mathcal{X}$ is an accumulation process of the form:

$$\mathcal{X}(T) = \int_{-\infty}^{T} \phi(t) \, d\mathcal{N}_\Gamma^\alpha[t]$$

where $\phi : \mathbb{R} \longrightarrow [0,\infty)$ is nonnegative. Then for any $T \in \mathbb{R}$, the random vector $\mathcal{X}(T)$ is α-stable, with spectral measure $\Gamma(T) = C(T) \cdot \Gamma$, where

$$C(T) = \int_{-\infty}^{T} \phi(t)^\alpha \, dt.$$

Knowing $C(T)$, we can reconstruct ϕ:

$$\phi(T) \;=\; \left(\frac{\partial C}{\partial t}(T)\right)^{1/\alpha}$$

Conclusion: Given sufficient sample data of the random vector $\mathcal{X}(T)$, we can use the methods of Chapter 8 to determine the stability exponent α, if any. We can then use the methods of **Part II** to reconstruct the spectral measure of $\mathcal{X}(T)$. In all three of the cases described above, this information is sufficient to reconstruct the spectral measure of the underlying noise, up to a scalar multiple.

9.3 Harmonic Decomposition of Integral Kernels

Harmonic analysis provides methods for representing functions on the real line as "linear combinations" of simple functions. The Laplace transform allows us to represent a function as a linear combination of functions of the form $\exp\left[(\lambda_1 + \lambda_2\mathbf{i})t\right]$, with $\lambda_1 > 0$ and $\lambda_2 \in \mathbb{R}$, while the Fourier Transform provides a representation in terms of periodic functions $\exp(\mathbf{i}\lambda t)$, where $\lambda \in \mathbb{R}$

Given a "moving average" stochastic process $\mathcal{X}$, determined by convolution with respect to some kernel ϕ, we can use these "harmonic decompositions" of ϕ to represent $\mathcal{X}$ as a linear combination of stochastic processes with exponential kernels.

This is good news if we want to find a "natural explanation" for the process $\mathcal{X}$. Exponential-kernel processes (such as the Ornstein-Uhlenbeck process) have very natural origins as the solutions to linear stochastic differential equations. However, given an arbitrary moving-average process $\mathcal{X}$, determined by some arbitrary convolution kernel ϕ, it is difficult to postulate a natural mechanism to generate it. The decomposition sketched here suggests a mechanism: $\mathcal{X}$ arises as a linear combination of (not necessarily independent) exponential-kernel processes. In other words, the system which generates $\mathcal{X}$ is a combination of subsystems which evolve according to *linear stochastic differential equations* with various time constants.

Inverse Laplace Transforms and Ornstein-Uhlenbeck Processes: Suppose that $\mathcal{X} : \mathbb{R} \longrightarrow \mathbb{R}$ is a *nonanticipating, moving average process.* In

other words, $\mathcal{X}$ is determined by a stochastic integral of the form:

$$\mathcal{X}(T) \;=\; \int_{-\infty}^{T} \phi(T-t)\; d\mathcal{N}[t]$$

where $\mathcal{N}$ is a random noise, and $\phi : [0,\infty) \longrightarrow \mathbb{R}$ is some "convolution kernel" describing how past perturbations influence the present. If ϕ decays sufficiently rapidly, then we can realize it as the Laplace transform of some other function. So, suppose that ϕ is the Laplace Transform of the function $\Phi : [0,\infty) \longrightarrow \mathbb{R}$. In other words, for all $t \in [0,\infty)$,

$$\phi(t) \;=\; \int_{0}^{\infty} \Phi(s) e^{-ts}\; ds$$

Thus, we can write:

$$\begin{aligned}
\mathcal{X}(T) &= \int_{-\infty}^{T} \phi(T-t)\; d\mathcal{N}[t] \\
&= \int_{-\infty}^{T} \int_{0}^{\infty} \Phi(s) e^{-(T-t)s}\; ds\; d\mathcal{N}[t] \\
&= \int_{0}^{\infty} \Phi(s) \int_{-\infty}^{T} e^{-s(T-t)}\; d\mathcal{N}[t]\; ds \\
&= \int_{0}^{\infty} \Phi(s) \mathcal{X}_s(t)\; ds
\end{aligned}$$

where $\mathcal{X}_s$ is an Ornstein-Uhlenbeck process with time constant s:

$$\mathcal{X}_s(T) \;=\; \int_{-\infty}^{T} e^{-s(T-t)}\; d\mathcal{N}[t]\; ds$$

In other words, by representing the kernel ϕ as a Laplace Transform, we can represent the stochastic process $\mathcal{X}$ as a *linear combination* of (non-independent) Ornstein-Uhlenbeck processes with various time constants. The coefficient $\Phi(s)$ measures the "influence" of the subsystem $\mathcal{X}_s$ on the motion of $\mathcal{X}$.

Fourier Transforms and Periodic-Kernel Processes: Now suppose that $\mathcal{X} : \mathbb{R} \longrightarrow \mathbb{R}$ is a *finite history* nonanticipating moving average process:

$$\mathcal{X}(T) \;=\; \int_{0}^{T} \phi(T-t)\; d\mathcal{N}[t]$$

where $\mathcal{N}$ is a random noise, and $\phi : [0, \infty) \longrightarrow \mathbb{R}$ is some convolution kernel.

Extend ϕ to all of $\mathbb{R}$ by symmetry, defining:

$$\phi(-s) \;=\; \phi(s)$$

If $\phi \in \mathbf{L}^1(\mathbb{R})$, then we can take its Fourier Transform, $\widehat{\phi}$. If $\widehat{\phi} \in \mathbf{L}^1(\mathbb{R})$ also, then the Fourier Inversion Formula says

$$\phi \;=\; \int_{-\infty}^{\infty} \widehat{\phi}(\xi)\mathcal{E}_\xi \; d\xi$$

where $\mathcal{E}_\xi(t) = \exp(2\pi \mathbf{i} \xi t)$.

Thus, we can write:

$$\begin{aligned}
\mathcal{X}(T) &= \int_0^T \phi(T-t) \; d\mathcal{N}[t] \\
&= \int_0^T \int_{-\infty}^{\infty} \widehat{\phi}(\xi)\mathcal{E}_\xi(T-t) \; d\xi \; d\mathcal{N}[t] \\
&= \int_{-\infty}^{\infty} \widehat{\phi}(\xi) \int_0^T \mathcal{E}_\xi(T-t) \; d\mathcal{N}[t] \\
&= \int_{-\infty}^{\infty} \widehat{\phi}(\xi)\mathcal{X}_\xi(t) \; d\xi
\end{aligned}$$

where $\mathcal{X}_\xi$ is a nonanticipating moving average process with *periodic* kernel:

$$\mathcal{X}_\xi(T) \;=\; \int_0^T e^{2\pi \mathbf{i} \xi (T-t)} \;\; d\mathcal{N}[t] \; ds$$

In other words, by representing the kernel ϕ as the inverse Fourier transform of $\widehat{\phi}$, we can represent the stochastic process $\mathcal{X}$ as a linear combination of (non-independent) periodic kernel process with various frequencies.

Laplace Transforms and Exponential Kernel processes: Now suppose $\phi : [0, \infty) \longrightarrow \mathbb{R}$ is any function with a well-defined Laplace Transform $\Phi : \mathbb{C}^+ \longrightarrow \mathbb{C}$. Here, $\mathbb{C}^+ = \{\lambda \in \mathbb{C} \,;\, \mathbf{re}\,[\lambda] > 0\}$, and for all such λ,

$$\Phi(\lambda) \;=\; \int_0^\infty \phi(t) \exp(-\lambda t) \; dt$$

If Φ is meromorphic, with all poles of order k or less, for some fixed $k \in \mathbb{N}$, and $r \in \mathbb{R}$ is larger than the real parts of all poles of Φ, then ϕ can be recovered via the Laplace Inversion Integral:

$$\phi(x) \;=\; \int_{-\infty}^{\infty} \Phi(r + s\mathbf{i}) e^{x(r+s\mathbf{i})} \; ds$$

Thus,

$$\begin{aligned}
\mathcal{X}(T) &= \int_{-\infty}^{T} \phi(T-t)\ d\mathcal{N}[t] \\
&= \int_{-\infty}^{T} \int_{-\infty}^{\infty} \Phi(r+s\mathbf{i}) e^{(T-t)(r+s\mathbf{i})}\ ds\ d\mathcal{N}[t] \\
&= \int_{-\infty}^{\infty} \Phi(r+s\mathbf{i}) \int_{-\infty}^{T} e^{(T-t)(r+s\mathbf{i})}\ d\mathcal{N}[t]\ ds \\
&= \int_{-\infty}^{\infty} \Psi(s) \mathcal{X}_{(r+s\mathbf{i})}(t)\ ds
\end{aligned}$$

where $\Psi(s) = \Phi(r+s\mathbf{i})$, and where $\mathcal{X}_{(r+s\mathbf{i})}$ is an stochastic process with a kernel that oscillates in phase with frequency s, while increasing exponentially in magnitude at rate r:

$$\mathcal{X}_{(r+s\mathbf{i})}(T) = \int_{0}^{T} e^{(T-t)(r+s\mathbf{i})}\ d\mathcal{N}[t]$$

These integrals clearly diverge as $T \to \infty$, indicating that this process is very unstable. Over time, the effects of small perturbations are exponentially magnified; as $T \to \infty$, the process $\mathcal{X}(T)$ will "explode off to infinity" in an unpredictable direction.

Part III

Background Material

Appendix A

Univariate Stable Distributions

[For more information on univariate stable probability distributions, the definitive reference is Zolatarev [190]. Other excellent introductions are chapter 1 of Taqqu and Samorodnitsky [154] and a forthcoming book by John Nolan [129]. Other recent references are [51], [143], and [3]; slightly older references are [46] and [1].]

A.1 Introduction

Why is the normal distribution considered the "canonical" distribution in probability theory? Why are Gaussian processes used as the "generic" models for almost any stochastic dynamics? The reason: the normal distribution has two properties:

- **Stability:** The normal distribution is **stable** in the sense that, if $\mathbf{X}$ and $\mathbf{Y}$ are independent random variables, each having a normal distribution, then $\mathbf{X} + \mathbf{Y}$ also has a normal distribution.
- **Renormalization Limit:** The Central Limit Theorem says that the normal distribution is the natural "limiting distribution" of a "suitably renoralized" infinite sum of independent random variables with finite variance.

Hence, the normal is the distribution we expect to see for any quantity which is generated by an infinite sequence of small perturbations; in particular, it is the distribution we expect in a stochastic process driven by *independent increments* of *finite variance*.

These properties uniquely characterize the normal distribution:

Theorem 73: *Characterization of Normal Distributions in* $\mathbb{R}$
Let ρ *be a probability measure on* $\mathbb{R}$. *The following are equivalent:*

1. ρ *has Radon-Nikodym derivative* $\frown[\mu;\sigma]$, *where* $\mu \in \mathbb{R}$ *and* $\sigma > 0$ *are constants, and*

$$\frown[\mu;\sigma](x) \;=\; \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-|x-\mu|^2}{\sigma^2}\right), \quad \forall x \in \mathbb{R}$$

2. ρ *has Fourier Transform:*

$$\chi[\xi] \;=\; \exp\left(\mu\xi\mathbf{i} \;-\; |\sigma \cdot \xi|^2\right), \quad \forall \xi \in \mathbb{R}.$$

3. *Let* $\rho_0 = -\mu^{(+)}\rho$; *hence,* ρ_0 *is the probability distribution of random variable* $(\mathbf{Y} - \mu)$, *where* $\mathbf{Y}$ *has distribution* ρ.

   If $\mathbf{X}_1, \ldots, \mathbf{X}_N$ *are independent random variables, identically distributed according to* ρ_0, *and* $\mathbf{X} = \mathbf{X}_1 + \ldots + \mathbf{X}_N$, *then*

$$\mathbf{D}^{istr}\left[\left(\frac{1}{\sqrt{N}}\right)\mathbf{X}\right] \;=\; \rho_0.$$

4. *Again, let* $\rho_0 = -\mu^{(+)}\rho$. *If* $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \ldots$ *are independent identically distributed random variables, with mean* 0 *and variance* σ^2, *and for all* N, *we define:*

$$\mathbf{Y}_N \;=\; \frac{1}{N^{1/2}} \sum_{n=1}^{N} \mathbf{X}_N$$

   then $\mathbf{D}^{istr}[\mathbf{Y}_N] \xrightarrow[N\to\infty]{} \rho_0$ *in the weak* topology.*

The exponent 1/2 is a recurring theme in this theorem. If we replace this with the exponent $1/\alpha$, where $\alpha \in [0, 2)$, then we can state an analogous theorem, which characterises the so-called α-**stable** distribution.

Theorem 74: *Characterization of Stable Distributions in* $\mathbb{R}$
Let ρ *be a probability measure on* $\mathbb{R}$, *and let* $\alpha \in [0, 2]$ *be a constant. The following are equivalent:*

1. *ρ has Fourier Transform*

$$\chi[\xi] \ = \exp\left(\mu\xi\mathbf{i} \ - \ |\sigma\cdot\xi|^{\alpha} \ - \ \mathcal{B}_{\alpha}\beta\sigma^{\alpha}\xi^{\langle\alpha\rangle}\mathbf{i}\right)$$

where:

- $\mu \in \mathbb{R}$,
- $\sigma \in [0,\infty)$,
- $\beta_0 \in [-1,+1]$,

are constants, and we define:

$$\begin{aligned}\xi^{\langle\alpha\rangle} &= \begin{cases} \mathbf{sign}(\xi)\cdot|\xi|^{\alpha} & if \quad \alpha \neq 1 \\ \xi\cdot\log|\xi| & if \quad \alpha = 1 \end{cases} \\ \text{and} \quad \mathcal{B}_{\alpha} &= \begin{cases} \tan\left(\frac{\pi\alpha}{2}\right) & if \quad \alpha \neq 1 \\ -\frac{2}{\pi} & if \quad \alpha = 1 \end{cases}\end{aligned}$$

2. *ρ has Log-Fourier Transform*

$$\Phi[\xi] = \log\chi[\xi] \ = \ \mu\xi\mathbf{i} \ - \ |\sigma\cdot\xi|^{\alpha} \ - \ \mathcal{B}_{\alpha}\beta\sigma^{\alpha}\xi^{\langle\alpha\rangle}\mathbf{i} \tag{A.1}$$

3. *Let $\rho_0 = -\mu^{(+)}\rho$; hence, ρ_0 is the probability distribution of random variable $(\mathbf{Y}-\mu)$, where $\mathbf{Y}$ has distribution ρ. If $\mathbf{X}_1,\ldots,\mathbf{X}_N$ are independent random variables, identically distributed according to ρ_0, and $\mathbf{X} = \mathbf{X}_1 + \ldots + \mathbf{X}_N$, then*

$$\begin{aligned} &\textit{If } \alpha \neq 1\textit{:} && \mathbf{D}^{istr}\left[\left(\frac{1}{N^{1/\alpha}}\right)\mathbf{X}\right] &&= \ \rho_0. \\ &\textit{If } \alpha = 1\textit{:} && \mathbf{D}^{istr}\left[\left(\frac{1}{N}\right)\mathbf{X} - \mathcal{B}_1\sigma\beta\left(\frac{1}{N}\right)^{\langle 1\rangle}\right] &&= \ \rho_0, \end{aligned}$$

The displacement term $\mathcal{B}_1\sigma\beta\left(\frac{1}{N}\right)^{\langle 1\rangle}$ in the $\alpha = 1$ scenario is a function of the "skewedness", β, of the distribution, it is trivial when $\beta = 0$.

4. *Again, let $\rho_0 = -\mu^{(+)}\rho$; If $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \ldots$ are independent identically distributed random variables with density ω such that*

$$\lim_{|x|\to\infty}\frac{\omega(x)}{|x|^{\alpha+1}} \ = \ C \neq 0, \infty$$

Furthermore...

- *...if* $\alpha > 1$, *then suppose* $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \ldots$ *all have* mean zero.
- *...if* $\alpha \leq 1$, *then assume* $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \ldots$ *are* symmetrically distributed *around zero.*

For all N, *define:*

$$\mathbf{Y}_N \;=\; \frac{c}{N^{1/\alpha}} \sum_{n=1}^{N} \mathbf{X}_N.$$

Then $\mathbf{D}^{istr}\left[\mathbf{Y}_N\right] \xrightarrow[N\to\infty]{} \rho_0$ *in the weak* topology, for some suitably chosen constant* $c > 0$.

A random variable possessing the distribution ρ described by this theorem is called an α-**stable random variable.**

Note: In **Part 2** of the theorem, we are "abusing notation". Strictly speaking, to define the logarithm of the (complex-valued) Fourier transform in **Part 1**, we would need to pick a "branch" of the multivalued log function, along with a "cut" in the complex plane, etc. When treated as a function on $\mathbb{R}$, this "log Fourier Transform" would be discontinuous, arbitrary in definition, and unnatural. Instead, we are simply defining the "log Fourier Transform" of ρ as the argument of the exponential function in **Part1**.

We will employ this convention whenever we speak of "log-characteristic functions" in the discussion that follows.

Proof:

Proof of $1 \iff 2$**:** This is immediate.

Proof of $2 \Longrightarrow 3$**:**

Suppose ρ_0 is an arbitrary probability distribution, with log Fourier transform ϕ. If $\mathbf{X}_1, \ldots, \mathbf{X}_N$ are independently distributed according to ρ, and $\mathbf{X} = \mathbf{X}_1 + \ldots + \mathbf{X}_N$, then $\mathbf{X}$ has log characteristic function $N \cdot \phi$, and thus, $\frac{1}{N^{1/\alpha}}\mathbf{X}$ has log characteristic function

$$\xi \;\mapsto\; N \cdot \phi\left(\frac{1}{N^{1/\alpha}}\xi\right) \qquad (*)$$

Case 1: $\alpha \neq 1$

It follows from $(*)$ that ρ_0 satisfies **Part 3** if and only if:

$$N \cdot \phi\left(\frac{1}{N^{1/\alpha}}\xi\right) \;=\; \phi(\xi). \qquad (A)$$

which clearly happens when ϕ is of the form: $\phi(\xi) = |\sigma\xi|^\alpha + \tau\xi^{\langle\alpha\rangle}\mathbf{i}$

where $\sigma > 0$ and $\tau \in \mathbb{R}$. Now, if we define $\beta = \frac{\tau}{\sigma^\alpha \mathcal{B}_\alpha}$,

then we can rewrite this: $\phi(\xi) = |\sigma\xi|^\alpha + \sigma^\alpha \mathcal{B}_\alpha \beta \mathbf{i} \xi^{\langle\alpha\rangle}$.

Case 2: $\alpha = 1$

Let $c \in \mathbb{R}$. It follows from $(*)$ that $\frac{1}{N}\mathbf{X} - c\left(\frac{1}{N}\right)^{\langle 1\rangle}$ has log-characteristic function

$$\xi \mapsto N \cdot \phi\left(\frac{1}{N}\xi\right) + c\left(\frac{1}{N}\right)^{\langle 1\rangle}\mathbf{i}$$

Thus, ρ_0 satisfies **Part 3** if, and only if:

$$\begin{aligned}\phi(\xi) &= N \cdot \phi\left(\frac{1}{N}\xi\right) + c\left(\frac{1}{N}\right)^{\langle 1\rangle}\mathbf{i} \\ &= N \cdot \phi\left(\frac{1}{N}\xi\right) + c\left(\frac{1}{N}\right)\log\left(\frac{1}{N}\right)\mathbf{i} \qquad (B)\end{aligned}$$

Which clearly happens when ϕ is of the form: $\phi(\xi) = |\sigma\xi| + \tau\xi\log|\xi|\mathbf{i}$,

where $\sigma > 0$ and $\tau \in \mathbb{R}$. Now, if we define $\beta = \frac{\tau}{\sigma\mathcal{B}_1}$,

then we can rewrite this: $\phi(\xi) = |\sigma\xi| + \sigma\mathcal{B}_1\beta\mathbf{i}\xi^{\langle 1\rangle}$

Proof of $3 \Longrightarrow 2$**:** (sketch)

Clearly, distributions satisfying **Part 2** will satisfy **Part 3**. But why must they *necessarily* have this form? For example, why is the parameter β confined to the range $[-1, +1]$? Why is the factor $\mathcal{B}_\alpha$ defined the way it is? Most puzzlingly, why is the case $\alpha = 1$ special, with extra "$\log|\xi|$" term, which makes everything so much more complicated?

These things all stem from the fact that a stable probability distribution must be **infinitely divisible**, and thus, has a log-characteristic function with **Lévy -Khintchine Canonical Form:**[1]

$$\Phi(\xi) = \mu\xi\mathbf{i} - \frac{\sigma^2|\xi|^2}{2} + \int_{\mathbb{R}}\left(\exp(\xi.u.\mathbf{i}) - 1 - \frac{\xi.u.\mathbf{i}}{1+u^2}\right)d\Lambda[u] \quad (A.2)$$

[1] See [20], §30, p. 149, or, alternately, see [46] XVII.2 for more on the the Lévy - Khintchine form.

where Λ is some measure on $\mathbb{R}$, called the **Lévy measure** of the distribution. Λ is absolutely continuous with respect to the Lebesgue measure, and its Radon-Nikodym derivative $\mathbf{L}(x) = \frac{d\Lambda}{d\mathcal{L}^{leb}}$ must be decreasing at least inverse-linearly as $x \to \pm\infty$, so that the function $x\mathbf{L}(x)$ is nonincreasing everywhere.

The proof now proceeds by putting equation (A.2 on the preceding page) into either formula (A) or (B) above, and deducing that the function Φ must have the form of formula (A.1 on page 131). For a more detailed sketch, see [46], XVII.4, p. 540. For a full proof, see [20], §34, p. 164.

Proof of 1 $\iff$ 4**:** See [46], XVII.5, especially the "Concluding Remark" (p. 547); alternately, see [20], §35, page. 171.

□

Remark 75:

- The parameter α is called the **stability exponent**. The normal distribution is the special case when $\alpha = 2$.
- When $\alpha > 1$, the parameter μ is the *mean* of the distribution. When $\alpha \leq 1$, the mean is not well-defined. However, μ can still be thought of as defining the "centroid" of the distribution in some sense, and is called the **shift parameter** of the distribution.
- When $\alpha = 2$, the parameter σ is the *standard deviation* of the distribution. When $\alpha \leq 2$, the variance and standard deviation are not well-defined. However σ can still be thought of as defining "fatness" of the distribution in some sense. σ is called the **variation** or **scale parameter** of the distribution.
- When $\alpha = 2$, the parameter β becomes meaningless, since, by definition, $B_2 = 0$. Hence, β is a feature unique to *non-normal* stable random variables. β is called the **skewness** parameter, and measures measures the asymmetry of the distribution. The smaller α becomes, the more extreme this asymmetry can be.
  - The distribution is symmetric about μ if and only if $\beta = 0$.
  - Although mean of the distribution (when $\alpha > 1$) is not affected by the skewness, the *mode* is. The mode is μ if and only if $\beta = 0$. If $\beta > 0$ ($\beta < 0$), then the mode is greater than (less than) μ.

- Let $\alpha < 1$. If $\beta = +1$, then the support of the distribution is confined to $[\mu, \infty)$. If $\beta = -1$, then the support of the distribution is confined to $(-\infty, \mu]$.

  If $\alpha \geq 1$, then the support of the distribution is all of $\mathbb{R}$, even when $\beta = \pm 1$.

- The distribution is *symmetric* if and only if $\mu = \beta = 0$. In this case, the expression for the characteristic function simplifies to:

$$\chi[\xi] \;=\; \exp\left(\sigma^{\alpha}|\xi|^{\alpha}\right)$$

  which is recognizably analogous to the Gaussian case.

- When $\alpha = 2$, the tails of the distribution decay at the *exponential* rate of $\exp\left(-|x|^2\right)$ as $x \to \infty$. However, when $\alpha < 2$, the tails decay much more slowly, at a *polynomial* rate of $|x|^{-\alpha-1}$ as $x \to \infty$.

  Hence, although the normal distribution has well-defined absolute moments of all degrees, the α-stable distribution (for $\alpha < 2$) only has well-defined absolute moments of degrees less than α. If $\beta \geq \alpha$, and $\mathbf{X}$ is an α-stable random variable, then the βth absolute moment $\mathbf{E}^{xpct}\left[|\mathbf{X}|^{\beta}\right]$ is "infinite".

  In particular, if $\alpha < 2$, then an α-stable random variable has "infinite variance".

We will use the notation $\frown^{\alpha}[\mu, \beta, \sigma]$ to denote this distribution.

Proposition 76: *(Algebra of Stable Random Variables)*
Let $\alpha \in [0, 2)$. Let $\mathbf{X}_1, \mathbf{X}_2 \in \mathbb{R}$ be independent. α-stable random variables, with $\mathbf{D}^{istr}[\mathbf{X}_1] = \frown^{\alpha}[\mu_1, \beta_1, \sigma_1]$ and $\mathbf{D}^{istr}[\mathbf{X}_2] = \frown^{\alpha}[\mu_2, \beta_2, \sigma_2]$. Then

- *$\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2$ is an α-stable random variable, and $\mathbf{D}^{istr}[\mathbf{X}] = \frown^{\alpha}[\mu, \beta, \sigma]$, where*

$$\begin{aligned} \mu &= \mu_1 + \mu_2 \\ \sigma &= \left(\sigma_1^{\alpha} + \sigma_2^{\alpha}\right)^{1/\alpha} \\ \beta &= \frac{\beta_1 \sigma_1^{\alpha} + \beta_2 \sigma_2^{\alpha}}{\sigma^{\alpha}} \end{aligned}$$

- *If* $c \in \mathbb{R}$, *then* $c.\mathbf{X}_1$ *is also an* α*-stable random variable, and* $\mathbf{D}^{istr}\left[c.\mathbf{X}_1\right] = \bigcap^{\alpha}\left[\mu, \beta, \sigma\right]$, *where:*

$$\begin{aligned} \mu &= \begin{cases} c \cdot \mu_1 & if \quad \alpha \neq 1 \\ c \cdot \mu_1 + \mathcal{B}_1 c^{\langle 1 \rangle} \cdot \sigma_1 \beta_1 & if \quad \alpha = 1 \end{cases} \\ \sigma &= |c| \cdot \sigma_1 \\ \beta &= \mathbf{sign}(c) \cdot \beta_1 \end{aligned}$$

- *If* $c_1, c_2 \in \mathbb{R}$, *then* $\mathbf{X} = c_1\mathbf{X}_1 + c_2\mathbf{X}_2$ *is an* α*-stable random variable, and* $\mathbf{D}^{istr}\left[\mathbf{X}\right] = \bigcap^{\alpha}\left[\mu, \beta, \sigma\right]$, *where*

$$\begin{aligned} \mu &= \begin{cases} c_1 \cdot \mu_1 + c_2 \cdot \mu_2 & if \quad \alpha \neq 1 \\ c_1 \cdot \mu_1 + c_2 \cdot \mu_2 + \mathcal{B}_1 c^{\langle 1 \rangle} \cdot \sigma_1 \beta_1 + \mathcal{B}_2 c^{\langle 1 \rangle} \cdot \sigma_2 \beta_2 & if \quad \alpha = 1 \end{cases} \\ \sigma &= \left(|c_1|^{\alpha}\sigma_1^{\alpha} + |c_2|^{\alpha}\sigma_2^{\alpha}\right)^{1/\alpha} \\ \beta &= \frac{\beta_1 \mathbf{sign}(c_1)\sigma_1^{\alpha} + \beta_2 \mathbf{sign}(c_1)\sigma_2^{\alpha}}{\sigma^{\alpha}} \end{aligned}$$

Proof: Let Φ_1, Φ_2 be the log characteristic functions of $\mathbf{X}_1$ and $\mathbf{X}_2$, respectively, as given by **Part 2** of Theorem 74 on page 130. Then the log characteristic function of $\mathbf{X}_1 + \mathbf{X}_2$ is just $\Phi_1 + \Phi_2$. The log characteristic function of $c.\mathbf{X}_1$ is $(\xi \mapsto \Phi_1(c.\xi))$. At this point, the algebra is straightforward.

(for more detail, see [154], Theorems 1.2.1 and 1.2.2). □

A.2 Examples and Explicit Formulae

In general, there is no explicit formula for the *density function* of a stable distribution, except in three special cases

Normal Distribution: When $\alpha = 2$, we have the normal distribution

$$\bigcap^{2}\left[\mu; \sigma\right](x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-|x-\mu|^2}{\sigma^2}\right)$$

Cauchy Distribution: When $\alpha = 1$, and $\beta = 0$, we get the Cauchy Distribution:

$$\bigcap^{1}\left[\mu; 0; \sigma\right](x) = \frac{1}{\pi} \cdot \frac{\sigma}{\left(|x-\mu|^2 + \sigma^2\right)}$$

Lévy Distribution: When $\alpha = \frac{1}{2}$, and $\beta = +1$, we get the Lévy distribution:

$$\smile^{1/2}[\mu; 1; \sigma](x) = \left(\frac{\sigma}{2\pi}\right)^{1/2} \frac{1}{(x-\mu)^{3/2}} \cdot \exp\left(\frac{-\sigma}{2(x-\mu)}\right) \cdot \mathbb{1}_{(\mu,\infty)}(x)$$

A.3 Applications and Examples

Stable probability distributions arise naturally in many contexts; for an excellent overview, see Chapter 1 of [190]. The examples mentioned there include:

- The limiting distributions of certain branching processes.
- The distribution on a flat screen of impact points of radioactive particles from a point source.
- The energy distribution for certain unstable systems in quantum mechanics, where it is called the Lorentz distribution.
- The distribution for the strength, at a point in space, of a random "influence field" of generated by a randomly dispersed collection of particles. For example:
  - The *Holtsmark distribution* [64] for the random gravity field generated by a suitable random distribution of stellar masses.
  - The temperature in a radioactive body.
  - The stresses within a crystal.
  - The magnetic field generated by a random arrangement of magnets
- The character of noise in certain radio engineering applications.
- The statistics of phylogenetic trees in biological taxonomy, related to a power law discovered by Willis [74] and Yule [182].

Stable distributions have also found applications in engineering. For example, Tsakalides and Nikias [132] have suggested using Cauchy distributions to model data from radar sensors. Other researchers have investigated stable noise in the context of signal processing [30],[31] [150], [28], [29], telecommunications [23],[145], and even the power distribution of ocean

waves [135]. In addition, Mandelbrot [105], [108] and Fama [43], [42] have famously advocated the use of stable distributions to model data from financial markets, inspiring a variety of researchers to extend classical mathematical financial methods to the stable universe [155], [44], [187], [142], [10], [11],[38] [117], [152], [52],[53], [99], [100], [68], [67] [82], [21]; for a short review, see Appendix C.1 on page 155; for a more extended reference, see [24].

Finally, since stable distributions are the limiting distributions of sums of independent perturbations following any power law of the form x^β, where $-3 < \beta < 1$, they can be expected to arise in any context where power-law distributions have been observed; for a review of the literature on this subject, see Appendix C on page 153.

I will only discuss the most basic and obvious instances of stable distributions here.

The Renormalization Semigroup: Fix $\alpha \in [0, 2]$. One interpretation **Part 3** of Theorem 74 on page 130 is that the α-stable distributions are the attracting fixed points of the α-**Renormalization Semigroup**[2], acting on the space probability measures on $\mathbb{R}$.

Let $\mathcal{M}_{\mathcal{EAS}}[\mathbb{R}]$ denote the space of Borel probability measures on the real line. $\mathcal{M}_{\mathcal{EAS}}[\mathbb{R}]$ is a semigroup under the action of **convolution**: if ρ and η are two probability measures, we define $\rho * \eta$ to be the probability measure so that, for any measurable $U \subset \mathbb{R}$,

$$\rho * \eta(U) \ = \ \int_{\mathbb{R}} \rho[U - x] \ d\eta[x]$$

If $\mathbf{X}$ and $\mathbf{Y}$ are independent random variables with distributions ρ and η respectively, then the random variable $\mathbf{X} + \mathbf{Y}$ has distribution $\rho * \eta$.

Now, fix $\alpha \in [0, 2]$, and, for all $N \in \mathbb{N}$, define $T_\alpha^N : \mathcal{M}_{\mathcal{EAS}}[\mathbb{R}] \longrightarrow \mathcal{M}_{\mathcal{EAS}}[\mathbb{R}]$ to be the map:

$$T_\alpha^N(\rho) \ = \ \left(\frac{1}{N^{1/\alpha}}\right)^\times \left(\underbrace{\rho * \rho * \ldots * \rho}_{N}\right)$$

In other words, $T_\alpha^N(\rho)$ is the distribution of the random variable

$$\left(\frac{1}{N^{1/\alpha}}\right)(\mathbf{X}_1 + \mathbf{X}_2 + \ldots \mathbf{X}_N)$$

[2]In physics literature, this is often called the Renormalization *Group*. Since it is not actually a group, this terminology is somewhat misleading.

where $\mathbf{X}_1$, $\mathbf{X}_2, \ldots,$ $\mathbf{X}_N$ are independent random variables with distribution ρ.

The family $\left\{T_\alpha^N \; ; \; N \in \mathbb{N}\right\}$ forms a semigroup, isomorphic to the semigroup of natural numbers under multiplication, because for any $N, M \in \mathbb{N}$,

$$T_\alpha^N \circ T_\alpha^M \;=\; T_\alpha^{N \cdot M}$$

Suppose that ρ is symmetrically distributed about 0. What is the weak* limit of $T_\alpha^N \rho$, as $N \to \infty$? According to **Part 3** of Theorem 74 on page 130, if ρ has a density with asymptotic decay rate of order $|x|^{-\alpha-1}$ as $|x| \to \infty$, then

$$\lim_{N \to \infty} T_\alpha^N \rho \;=\; \frown^{\alpha} [0; \beta; \sigma]$$

in the weak* topology, for some suitable choice of parameters β and σ.

Of course, we can generalize this to the case where ρ is symmetrically distributed about $\mu \neq 0$, by redefining the action of T_α^N to shift the distribution by $N\mu$. The weak* limit would then be $\frown^{\alpha} [\mu; \beta; \sigma]$.

First Contact Times for Brownian Motion: Suppose that $\mathcal{B} : [0, \infty) \longrightarrow \mathbb{R}$ is a one-dimensional Brownian motion process, and let $x > 0$. Define the stopping time

$$\tau_x \;=\; \min \{t > 0 \; ; \; \mathcal{B}(t) = a\}$$

Recall the following properties of Brownian Motion:

- (A) **Self-affinity:** For any $\lambda > 0$, the process $t \mapsto \mathcal{B}(\lambda.t)$ is identical to the process $t \mapsto \sqrt{\lambda}.\mathcal{B}(t)$.
- (B) **Strong Markov Property:** Brownian motion "restarted" at any stopping time is identical to Brownian motion. Thus, the process $\mathcal{B}'(t) \;=\; \mathcal{B}(t + \tau_r) - \mathcal{B}(\tau_r)$ is identical to $\mathcal{B}$.
- (C) The path $\mathcal{B}$ is **almost-surely continuous**. Thus, by the Intermediate Value Theorem, if $y > x > 0$, then $\tau_y > \tau_x$.

From property (A), it follows that, if $\kappa = \sqrt{\lambda}$, then process $\mathcal{B}(\kappa^2 t)$ is identical to the process $\kappa \mathcal{B}(t)$, so

$$\kappa^2 \tau_x \quad \stackrel{\cong}{_{distr}} \quad \tau_{\kappa x}$$

Suppose $0 < r < s$. From property (C), we know that τ_s must be larger than τ_r. It follows from (B) that $\tau_s - \tau_r$ is identically distributed with $\tau_{(s-r)}$.

In other words, $\tau_s \overset{\cong}{_{distr}} \tau_r + \tau_{(r-s)}$, treated as a sum of two independent random variables. Hence, if $\rho_r = \mathbf{D}^{istr}[\tau_r]$, then we have

$$\rho_s = \rho_r * \rho_{(s-r)}$$

thus,

$$\begin{aligned} \underbrace{\rho * \ldots * \rho}_{N} &= \mathbf{D}^{istr}\left[\tau_{(Nx)}\right] \\ &= N^2 \mathbf{D}^{istr}[\tau_x] \\ &= N^{1/\alpha}\rho \end{aligned}$$

where $\alpha = 1/2$. Hence, we conclude that, for any x, τ_x is a $(1/2)$-stable random variable. Since τ_x must be positive, this variable is furthermore totally skewed to the right –hence, the skewness parameter β must equal 1. A straightforward computation using properties of Brownian motion reveals

$$\mathbf{D}^{istr}[\tau_x] = \mathcal{S}^{1/2}[x; 1; 0]$$

(ie. $\sigma = x$ and $\mu = 0$). (see [46], P. 171)

A.4 Statistical Methods

The problem of parameter estimation for stable random variables in one dimension has been well-studied. The earliest methods were developed by Fama and Roll [39], who developed a method based on the quantiles of the empirical distribution. Simple heuristic formulae involving these sample quantiles provided crude but rapid estimates ofparameters; more accurate estimates could be made using precalculated quantile tables. Unfortunately, Fama and Roll only addressed the case of symmetric distributions with $\alpha > 1$,

Next, Press [141] developed a method for estimating the parameters of a stable distribution, using empirical estimates of the characteristic functions. Press uses the fact that the log characteristic function follows a power law; hence, a log-log plot of the log characteristic function will be linear, and it's slope will be the stability exponent α. Other parameters are estimated using similar "linear" methods. Press also extends his method to deal with a limited class of symmetric multivariate distributions.

Koutrouvelis [92] also uses the linear properties of the log-log plot of the log characteristic function. He develops a method for estimating the

parameters based on linear regression. He further refines this method in [93]. Feuerverger and McDunnogh [9], Arad [10], and Paulson, Holcomb and Leitch [41] also develop methods using the sample characteristic function.

McCulloch [116] developes a more sophisticated quantile-based method, which is applicable for asymmetric distributions and $\alpha \geq 0.6$. For example McCulloch uses

$$\frac{x_{.95} - x_{.05}}{x_{.75} - x_{.25}}$$

to estimate α, and

$$\frac{x_{.95} - 2x_{.50} + x_{.05}}{x_{.95} - x_{.05}}$$

to estimate β. Estimation of parameters is again performed by reference to precomputed tables of order statistics.

Badahdah and Siddiqui [153] also use the the quantiles of the empirical sample, along with derived statistics (eg. "Winsorized" means, trimmed means), to develop robust estimators of the μ parameter, and to compute tables of the Fisher information of stable distributions. DuMouchel [36] also studies the Fisher information of these distributions.

Nolan [127] developes maximum-likelihood estimation techniques for the parameters of stable distributions, and together with Fofack [7], performs a detailed analysis of the modes and Paretian tails of the distributions.

Other estimation methods have been developed by Paulson and Delahanty [12], and Zolatarev [189].

Appendix B

Multivariate Stable Distributions

[For more information on multiivariate stable probability distributions, the best references are chapter 2 of Taqqu and Samorodnitsky [154], and the review provided by [128].]

B.1 Characteristic Functions and Spectral Measures

Theorem 77: *Characterization of Stable Distributions in* $\mathbb{R}^D$

Let ρ *be a probability measure on* $\mathbb{R}^D$*, and let* $\alpha \in [0, 2)$ *be a constant. The following are equivalent:*

1. ρ *has Log-Fourier Transform*[1]

$$\Phi[\vec{\xi}] \quad = \quad \left\langle \vec{\mu}, \vec{\xi} \right\rangle \mathbf{i} \quad - \quad \int_{\mathbb{S}^{D-1}} \eta_{\vec{\xi}}^{(\alpha)}(\mathbf{s}) \; d\Gamma[\mathbf{s}]. \qquad \text{(B.1)}$$

where $\eta_{\vec{\xi}}^{(\alpha)}(\mathbf{s}) \; = \; \left| \left\langle \vec{\xi}, \mathbf{s} \right\rangle \right|^{\alpha} \; + \; \mathcal{B}_\alpha \left\langle \vec{\xi}, \mathbf{s} \right\rangle^{\langle \alpha \rangle} \mathbf{i}$, *with*

$$\xi^{\langle \alpha \rangle} \; = \; \begin{cases} \mathbf{sign}(\xi) \cdot |\xi|^{\alpha} & if \quad \alpha \neq 1 \\ \xi \cdot \log |\xi| & if \quad \alpha = 1 \end{cases} \quad \text{and} \quad \mathcal{B}_\alpha \; = \; \begin{cases} \tan\left(\frac{\pi\alpha}{2}\right) & if \quad \alpha \neq 1 \\ -\frac{2}{\pi} & if \quad \alpha = 1 \end{cases} .$$

Here, Γ *is some nonegative Borel measure on the unit sphere* $\mathbb{S}^{D-1} = \left\{ \vec{x} \in \mathbb{R}^D \; ; \; \|\vec{x}\| = 1 \right\}$.

[1] See note after Theorem 74 on page 132, .

2. *Let $\rho_0 = -\vec{\mu}^{(+)}\rho$; hence, ρ_0 is the probability distribution of random variable $\mathbf{Y} - \vec{\mu}$, where $\mathbf{Y}$ has distribution ρ.*

   If $\mathbf{X}_1, \ldots, \mathbf{X}_N$ are independent random vectors in $\mathbb{R}^D$, identically distributed according to ρ_0, and $\mathbf{X} = \mathbf{X}_1 + \ldots + \mathbf{X}_N$, then

   $$\text{If } \alpha \neq 1\text{:} \qquad \mathbf{D}^{istr}\left[\left(\frac{1}{N^{1/\alpha}}\right)\mathbf{X}\right] = \rho_0.$$

   $$\text{If } \alpha = 1\text{:} \qquad \mathbf{D}^{istr}\left[\left(\frac{1}{N}\mathbf{X}\right) - \left(\frac{1}{N}\right)^{\langle 1\rangle}\vec{\beta}\right] = \rho_0,$$

   The displacement term $\vec{\beta} \in \mathbb{R}^D$ is a measure of the "asymmetry" of the spectral measure. If Γ is symmetric (ie. $\Gamma[-U] = \Gamma[U]$, for all $U \subset \mathbb{S}^{D-1}$), then $\vec{\beta} = 0$.

3. *Again, let $\rho_0 = -\vec{\mu}^{(+)}\rho$. If $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N$ are independent identically distributed random vectors in $\mathbb{R}^D$ with density ω such that*

   $$\lim_{\|x\|\to\infty} \frac{\omega(x)}{\|x\|^{\alpha+1}} = C \neq 0, \infty.$$

   Furthermore, for all N, define $\mathbf{Y}_N = \frac{1}{N^{c/\alpha}} \sum_{n=1}^{N} \mathbf{X}_N$. If either

   - *$\alpha > 1$, and $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \ldots$ all have* mean zero.
     or
   - *$\alpha \leq 1$, and $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \ldots$ are* symmetrically distributed *around zero,*

   then $\mathbf{D}^{istr}[\mathbf{Y}_N] \xrightarrow[N\to\infty]{} \rho_0$ in the weak topology, for some suitably chosen scalar $c > 0$*

Proof: See [154], §2.3, p.65, or [95]. □

Γ is called the **spectral measure** of the distribution[2]. What is the

[2]This terminology has become standard, but is somewhat unfortunate. The "spectral measure" of a multivariate stable distribution is *entirely unrelated* to any of the other spectral measures or other "spectral" objects in mathematics and probability theory. The "spectrum" of an electromagnetic or acoustic signal, "power spectral density" of a stationary stochastic process, the "spectral measure" of a normal operator on a Hilbert space, the "spectra" of C* algebra, Banach algebras, and commutative rings, and even the energy spectra of atoms are all loosely related; the spectral measure of a stable distribution is related to none of them. Perhaps it would be more appropriate to call it a **Feldheim measure**, since Feldheim [45] was the first to define it.

meaning of Γ? For all $s \in \mathbb{S}^{D-1}$, suppose $\mathbf{X}_s$ are independent, $\alpha-$stable random variables in $\mathbb{R}$, totally "skewed" to the right —ie. having log-characteristic functions:

$$\Phi_{\mathbf{X_s}}(\xi) \quad = \quad -|\xi|^{\alpha} + \mathcal{B}_{\alpha}\xi^{\langle\alpha\rangle}\mathbf{i}$$

Recall that, if $\mathbf{X}$ and $\mathbf{Y}$ are independent random variables, with log-characteristic functions $\Phi_{\mathbf{X}}$ and $\Phi_{\mathbf{Y}}$, then

$$\Phi_{\mathbf{X+Y}} \quad = \quad \Phi_{\mathbf{X}} + \Phi_{\mathbf{Y}}$$

So, if $\mathbf{X}$ is α-stable, then:

$$\begin{aligned}\Phi_X(\vec{\xi}) &= \left\langle \vec{\mu}, \vec{\xi} \right\rangle \mathbf{i} \;-\; \int_{\mathbb{S}^{D-1}} \left(\left| \left\langle \vec{\xi}, \mathbf{s} \right\rangle \right|^{\alpha} \cdot \; + \mathcal{B}_{\alpha} \left\langle \vec{\xi}, \mathbf{s} \right\rangle^{\langle\alpha\rangle} \mathbf{i} \right) \; d\Gamma[\mathbf{s}] \\ &= \mathbf{i} \left\langle \vec{\mu}, \vec{\xi} \right\rangle \;+\; \int_{\mathbb{S}^{D-1}} \Phi_{(\mathbf{X}_s \cdot \mathbf{s})} \; d\Gamma[\mathbf{s}],\end{aligned}$$

Thus, intuitively $\mathbf{X}$ can be thought of as an "integral" of a collection of independent $\alpha-$stable random variables, indexed by $\mathbb{S}^{D-1}$:

$$\mathbf{X} \;=\; \mu \;+\; \int_{\mathbb{S}^{D-1}} \mathbf{X}_s \cdot \mathbf{s} \; d\Gamma[\mathbf{s}]$$

This heuristic can be made precise by representing $\mathbf{X}$ as the *stochastic integral* of a certain multivariate "α-stable noise" on $\mathbb{S}^{D-1}$, with respect to the measure Γ. This is called the **stochastic integral representation** of $\mathbf{X}$ (see Example 98 on page 188, or [154], Theorem 3.5.6, p. 131).

Thus, if much of mass of the spectral measure is concentrated in some small region of the sphere, this means that the stable random vector is a sum of many independent stable perturbations, *most* of which lie along rays transecting this small region. Hence, the tails of the multivariate distribution should decay most slowly within the cone spanned by this region.

More generally, we would expect the radial decay rates of the multivariate distribution in different directions to be somehow proportional to the concentration of spectral mass in those directions. This indeed is exactly the content of the following

Theorem 78: *(Araujo and Giné, 1980) Let ρ be a stable probability distribution in $\mathbb{R}^D$, with spectral measure Γ; assume that ρ is centered at the origin. Let $\mathcal{U} \subset \mathbb{S}^{D-1}$ be some measurable subset, and define*

$$Cone[\mathcal{U};\; R] \quad = \quad \{r.\mathbf{u} \;;\; \mathbf{u} \in \mathcal{U},\; r > 0\}$$

For any $R > 0$, let $\mathbf{A}_R = \{\mathbf{x} \in \mathbb{R}^D \; ; \; |\mathbf{x}| > R\}$.
Then:

$$\lim_{R\to\infty} \frac{\rho\left[Cone[\mathcal{U};\; R] \cap \mathbf{A}_R\right]}{\rho\left[\mathbf{A}_R\right]} \quad = \quad \frac{\Gamma[\mathcal{U}]}{\Gamma\left[\mathbb{S}^{D-1}\right]}$$

Proof: See Chapter 3, Corollary 6.20, part (b) of [1], on page 152. □

Multivariate Stable Densities: Unfortunately, just as in the univariate case, there is no explicit expression for the density function ρ of a multivariate stable distribution. Nolan and Abdul-Hamid [80] offer a formula where $\rho(\mathbf{x})$ is expressed as an integral over $\mathbb{S}^{D-1}$ of a certain function depending on $\mathbf{x}$ and on the skewness and variance parameters characterizing the one-dimensional marginals of ρ –this is an improvement of an earlier, similar formula by Abdul-Hamid [4].

B.2 Examples

Example 79: *One dimensional Distributions:*

If $D = 1$, then $\mathbb{S}^{D-1} = \{\pm 1\}$, so the expression can be rewritten as:

$$\begin{aligned} \log \Phi_X(\vec{\xi}) &= \mathbf{i}\mu\xi \;-\; \sum_{s=\pm 1} \left(\left| \left\langle \vec{\xi}, \vec{s} \right\rangle \right|^{\alpha} \;+\; B_\alpha \left\langle \vec{\xi}, \vec{s} \right\rangle^{\langle\alpha\rangle} \mathbf{i} \right) \Gamma[\{s\}] \\ &= \mathbf{i}\mu\xi \;-\; |\sigma\xi|^{\alpha} \;-\; \sigma B_\alpha \beta \xi^{\langle\alpha\rangle} \mathbf{i} \end{aligned}$$

where $\sigma = (\Gamma\{+1\} + \Gamma\{-1\})^{1/\alpha}$ is the "variation", and $\beta = \frac{\Gamma\{+1\} - \Gamma\{-1\}}{\sigma^\alpha}$ is the one-dimensional "skewness" parameter. (To see this, use Proposition 76 on page 135.)

We recover the "symmetric" case by setting $\beta = 0$ —that is, $\Gamma\{+1\} = \Gamma\{-1\}$.

Example 80: *Symmetric Distributions:*

Suppose that Γ is a **symmetric** measure —that is, $\Gamma[U] = \Gamma[-U]$ for any measurable subset $U \subset \mathbb{S}^{D-1}$. Then the "skewness" terms all cancel off, and the log-characteristic function can be rewritten as:

$$\log \Phi_X(\vec{\xi}) \quad = \quad \mathbf{i} \left\langle \vec{\mu}, \vec{\xi} \right\rangle \quad - \quad \int_{\mathbb{S}^{D-1}} \left| \left\langle \vec{\xi}, \vec{s} \right\rangle \right|^{\alpha} d\Gamma[\vec{s}]$$

This is called a **symmetric α-stable** random vector. For every such vector, there is a unique symmetric measure Γ making the previous formula true.

Example 81: *"Standard" Subgaussian Distribution:*

Suppose that $\mathbf{G}$ is random vector in $\mathbb{R}^D$, with multivariate standard normal distribution, and $\mathbf{Y}$ is an (independent) $\frac{\alpha}{2}$-stable random scalar, totally skewed to the right, and let

$$\mathbf{X} \;=\; \mathbf{Y}^{1/2} \cdot \mathbf{G}.$$

Then $\mathbf{X}$ is α-stable random variable. The spectral measure Γ is a *uniform* measure on $\mathbb{S}^{D-1}$ —ie. $\Gamma = C \cdot \mathcal{L}^{bsg}$, where C is a constant, and $\mathcal{L}^{bsg}$ is the Lebesgue measure on the sphere.

If $\mathbf{G}$ has any multivariate normal distribution, then the product $\mathbf{Y}^{1/2} \cdot \mathbf{G}$ will be α-stable. The formula for the spectral measure is more complicated when $\mathbf{G}$ has nontrivial covariance structure, however.

Example 82: *First Contact Point of Multivariate Brownian Motion:*

Let
$\mathcal{B} : [0, \infty) \longrightarrow \mathbb{R}^{D+1}$ be the "standard" multivariate Brownian motion — ie. $\mathcal{B} = (\mathcal{B}_0, \mathcal{B}_1, \ldots, \mathcal{B}_D)$, where $\mathcal{B}_0, \mathcal{B}_1, \ldots, \mathcal{B}_D$ are independent Brownian processes on $\mathbb{R}$. For any $r > 0$, let $\mathbb{X}_r = \{r\} \times \mathbb{R}^D \subset \mathbb{R}^{D+1}$, and let $\mathcal{X}_r$ be the *first contact point* of $\mathcal{B}$ with the hyperplane $\mathbb{X}_r$ —in other words, if $\tau_r = \min\{t > 0 \,;\, \mathcal{B}_0(t) = r\}$, then $\mathcal{X}_r \;=\; (\mathcal{B}_1, \ldots, \mathcal{B}_D)(\tau_r)$.

We claim that, as a random vector in $\mathbb{R}^D$, $\mathcal{X}_r$ is a subgaussian stable random variable with stability exponent $\alpha = 1$. This follows from four properties of Brownian Motion:

- (A) **Self-affinity:** For any $\lambda > 0$, the process $t \mapsto \mathcal{B}(\lambda.t)$ is identically distributed with the process $t \mapsto \sqrt{\lambda}.\mathcal{B}(t)$.
- (B) **Strong Markov Property:** Brownian motion "restarted" at any stopping time is identical to Brownian motion. Thus, the process $\mathcal{B}'(t) \;=\; \mathcal{B}(t + \tau_r) - \mathcal{B}(\tau_r)$ is identically distributed with $\mathcal{B}$.
- (C) The path $\mathcal{B}$ is **almost-surely continuous**. Thus, by the Intermediate Value Theorem, if $y > x > 0$, then $\tau_y > \tau_x$.

- (D) **Spherical Symmetry:** If $F : \mathbb{R}^{D+1} \longrightarrow \mathbb{R}^{D+1}$ is an orthogonal linear transformation, then the process $F \circ \mathcal{B}$ is identically distributed with $\mathcal{B}$.

Suppose $0 < r < s$. By property (C), the first contact of $\mathcal{B}$ with $\mathbb{X}_s$ must happen *after* the first contact with $\mathbb{X}_r$. It follows from (B) that $\mathcal{X}_s - \mathcal{X}_r$ is identically distributed with $\mathcal{X}_{(s-r)}$. In other words, $\mathcal{X}_s \overset{\cong}{_{distr}} \mathcal{X}_r + \mathcal{X}_{(r-s)}$, treated as a sum of two independent random variables. Hence, if $\rho_r = \mathbf{D}^{istr}[\mathcal{X}_r]$, then we have

$$\rho_s = \rho_r * \rho_{(s-r)}$$

By the self-affinity of property (A), we can rescale space to conclude that, for any $\lambda > 0$,

$$\mathcal{X}_{\lambda.r} \overset{\cong}{_{distr}} \lambda.\mathcal{X}_r$$

Thus, $\mathcal{X}_1$ must be stable with stability exponent $\alpha = 1$. By property (D), the distribution of $\mathcal{X}_1$ must be spherically symmetrical in $\mathbb{R}^D$. The only spherically symmetric stable distributions are those which are subgaussian, with an underlying standard normal distribution. ———

Example 83: *A Sum of Principal Components:*

Let $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N \in \mathbb{R}^D$ be vectors of unit length, and let $\mathbf{X}_1, \ldots, \mathbf{X}_N$ be stable random variables in $\mathbb{R}$, independent, with $\mathbf{D}^{istr}[\mathbf{X}_n] = \int\!\!\!\!\!\!\sim^{\alpha} [\sigma_n; 1, 0]$.

Suppose $\mathbf{X} = \mathbf{X}_1\mathbf{y}_1 + \mathbf{X}_2\mathbf{y}_2 + \ldots + \mathbf{X}_N\mathbf{y}_N$.

Then the skewness measure Γ for $\mathbf{X}$ is defined:

$$\Gamma = \sum_{n=1}^{N} \sigma_n \delta_{\mathbf{y}_n}$$

$\delta_{\mathbf{y}_n}$ is the point mass situated at $\mathbf{y}_n$.

It may seem peculiar that the sub-Gaussian distribution and the sum of independent "principal components" have different (indeed, in a sense "opposite") spectral measures. After all, shouldn't the coordinates of a sub-Gaussian random vector be independent?

Apparently not. One way of seeing this is to compare the probability density functions of $\mathbf{X}^{(1)} = \mathbf{X}_1 + \mathbf{X}_2 + \ldots + \mathbf{X}_D$ (a sum of independent α-stable coordinates), versus that of $\mathbf{X}^{(2)} = \mathbf{Y}^{1/2} \cdot \mathbf{G}$ (a sub-Gaussian). The contours of equal probability density for $\mathbf{X}^{(2)}$ are spheres around

the origin. The contours of equal probability for $\mathbf{X}^{(1)}$ start out roughly spherical, but, as you move away from the origin, they start to "bulge" out along the axis of the space, so that, for example, when $D = 2$, they begin to look like "fat plus signs" (see [105], p. 387 for a picture). ——

B.3 Algebraic Properties of Spectral Measures

Proposition 84: *(Scalar multiplication of a Stable random vector)* *Let $\alpha \in [0, 2)$, $\alpha \neq 1$. Let $\mathbf{X} \in \mathbb{R}^D$ be an α-stable random vector, with spectral measure Γ. If $c \in \mathbb{R}$, then $c \cdot \mathbf{X}$ is also α-stable.*

1. *If $c > 0$, then $c \cdot \mathbf{X}$ has spectral measure $c^\alpha \cdot \Gamma$.*
2. *If $c < 0$, then $c \cdot \mathbf{X}$ has spectral measure $-c^\alpha.\widetilde{\Gamma}$, where $\widetilde{\Gamma}$ is the* **spherical inversion** *of Γ; in other words, for all subsets $U \subset \mathbb{S}^{D-1}$, $\widetilde{\Gamma}[U] = \Gamma[-U]$.*

Proof: If Φ is the log-characteristic function of $\mathbf{X}$, then the log characteristic function of $c.\mathbf{X}$ is $(\xi \mapsto \Phi_1(c.\xi))$. The algebra is now straightforward. □

Proposition 85: *(Sums of Independent Stable Random Vectors[3])* *Let $\alpha \in [0, 2)$. Let $\mathbf{X}_1, \mathbf{X}_2 \in \mathbb{R}^D$ be independent α-stable random vectors, with spectral measures Γ_1, Γ_2, respectively. Let $\mathbf{X}_0 = \mathbf{X}_1 + \mathbf{X}_2$ have spectral measure Γ_0. Then $\Gamma_0 = \Gamma_1 + \Gamma_2$.*

Proof: Suppose, for $k = 0, 1, 2$, that $\mathbf{X}_k$ has distribution ρ_k, characteristic function χ_k, and log-characteristic function Φ_k. Then

$$\Big(\mathbf{X}_0 = \mathbf{X}_1 + \mathbf{X}_2\Big) \Longrightarrow \Big(\rho_0 = \rho_1 * \rho_2\Big) \Longrightarrow \Big(\chi_0 = \chi_1 \cdot \chi_2\Big) \Longrightarrow \Big(\Phi_0 = \Phi_1 + \Phi_2\Big).$$

But recall from Theorem 77 on page 143 that, for any $\boldsymbol{\xi} \in \mathbb{R}^D$, $\Phi_k(\boldsymbol{\xi}) = \int_{\mathbb{S}^{D-1}} \eta_\xi^{(\alpha)}(\mathbf{s})\, d\Gamma_k[\mathbf{s}]$, where $\eta_\xi^{(\alpha)}(\mathbf{s}) = |\langle \boldsymbol{\xi}, \mathbf{s}\rangle|^\alpha + \mathcal{B}_\alpha \langle \boldsymbol{\xi}, \mathbf{s}\rangle^{\langle\alpha\rangle}\, \mathbf{i}$. Thus, we have:

$$\int_{\mathbb{S}^{D-1}} \eta_\xi^{(\alpha)}(\mathbf{s})\, d\Gamma_0[\mathbf{s}] = \Phi_0(\boldsymbol{\xi}) = \Phi_1(\boldsymbol{\xi}) + \Phi_2(\boldsymbol{\xi}) = \int_{\mathbb{S}^{D-1}} \eta_\xi^{(\alpha)}(\mathbf{s})\, d(\Gamma_1 + \Gamma_2)[\mathbf{s}]$$

This is true for every $\boldsymbol{\xi}$; this forces $\Gamma_0 = \Gamma_1 + \Gamma_2$. ——— □

[3] I am grateful to Jeremy Quastel for pointing out an error in the original statement of this result.

(Another way to see this is to use the "stochastic integral" interpretation of spectral measures given in Example 98 on page 188).

B.4 Marginal Distributions

Theorem 86: *Suppose that* $\mathbf{X}$ *has a* D*-dimensional* α*-stable random variable with spectral measure* Γ *and mean* $\vec{\mu}$*. If* $\mathbf{X}_d$ *be the* d*th coordinate of* $\mathbf{X}$*, then* $\mathbf{X}_d$ *is alsp* α*-stable, with distribution* $\int\!\!\!\!\!\frown^{\alpha} [\sigma_d, \nu_d, \beta_d]$*, where*

$$\begin{aligned} \nu_d &= \begin{cases} \mu_d & if \quad \alpha \neq 1 \\ \mu_d - \frac{1}{\pi} \int_{\mathbb{S}^{D-1}_{\pm d}} s_d \log\left(s_d^2\right) \, d\Gamma[\vec{s}] & if \quad \alpha = 1 \end{cases}, \\ \sigma_d &= \left(\int_{\mathbb{S}^{D-1}} |s_d|^{\alpha} \, d\Gamma[s_d] \right)^{1/\alpha} \\ and \;\; \beta_d &= \frac{1}{\sigma^{\alpha}} \int_{\mathbb{S}^{D-1}} s_d^{\alpha} \, d\Gamma[s_d] \end{aligned}$$

Proof: The general expression for the spectral measure of an N-dimensional marginal of $\mathbf{X}$, with $1 \leq N < D$, is provided on [154] p. 72. Applying this in the case when $N = 1$, we conclude that the distribution of $\mathbf{X}_d$ is given

$$\Phi_{\mathbf{X}_d}(\xi) = \mathbf{i}\nu_d\xi - \sum_{s=\pm 1} \left(|\xi|^{\alpha} \cdot + B_{\alpha} \left\langle \vec{\xi}, \vec{s} \right\rangle^{\langle \alpha \rangle} \mathbf{i} \right) \cdot \gamma_d(s)$$

where

$$\begin{aligned} \gamma_d(\pm 1) &:= \int_{\mathbb{S}^{D-1}_{\pm d}} s_d^{\alpha} \, d\Gamma[\vec{s}] \\ \mathbb{S}^{D-1}_{+d} &= \left\{ \vec{s} \in \mathbb{S}^{D-1} \, ; \; s_d > 0 \right\} \\ \mathbb{S}^{D-1}_{-d} &= \left\{ \vec{s} \in \mathbb{S}^{D-1} \, ; \; s_d < 0 \right\} \end{aligned}$$

and

$$\nu_d := \begin{cases} \mu_d & if \quad \alpha \neq 1 \\ \mu_d - \frac{1}{\pi} \int_{\mathbb{S}^{D-1}_{\pm d}} s_d \log\left(s_d^2\right) \, d\Gamma[\vec{s}] & if \quad \alpha = 1 \end{cases}$$

The parameters σ_d and β_d are derived from this "spectral measure" by:

$$\begin{aligned} \sigma_d &= \left(\gamma_d\{+1\} + \gamma_d\{-1\} \right)^{1/\alpha} \\ \beta_d &= \frac{\gamma_d\{+1\} - \gamma_d\{-1\}}{\sigma^{\alpha}} \end{aligned}$$

□

The problem of reconstructing the original distribution from its one-dimensional marginals is thus the problem of reconstructing the measure Γ on $\mathbb{S}^{D-1}$ from the $2D$ points $\{\gamma_d(s)\ ;\ d \in [1..D],\ s = \pm 1\}$. In the Gaussian case, this is relatively straightforward; all we need to complete the picture is the "correlation structure", as represented by the components of the **correlation matrix**. For the general α-stable case, however, the problem is far more complicated.

B.5 Simulating $\alpha-$stable Random Variables

According to [154], *p.*42, the simplest way to simulate a one-dimensional $\alpha-$stable random variable, with $\sigma = 1, \beta = \mu = 0$, is as follows:

- Let $\mathbf{U}$ be a uniform random variable on $\left(\frac{-\pi}{2}, \frac{\pi}{2}\right)$.
- Let $\mathbf{E}$ be an exponentially distributed random variable, with mean 1, independent of $\mathbf{U}$.

Then

$$\mathbf{X} := \frac{\sin(\alpha \cdot \mathbf{U})}{(\cos(\mathbf{U}))^{1/\alpha}} \cdot \left(\frac{\cos\left[(1-\alpha)\cdot \mathbf{U}\right]}{\mathbf{E}}\right)^{\frac{1-\alpha}{\alpha}}$$

is an $\alpha-$stable random variable with $\sigma = 1$ and $\beta = \mu = 0$.

Samorodnitzky and Taqqu [154] claim that a "similar" formula exists for the skewed case ($\beta \neq 0$), but they do not state it explicitly, although they refer to [36] and [77] for further information. Instead, they provide source code for a `FORTRAN` program called `rstab`[4], which will simulate the skewed random variable (see [154] Section 1.7, p. 46)

Simulating $\alpha-$stable Random Vectors

From this, we can simulate an $\alpha-$stable random vector with skewness measure Γ, simply by using a program like `rstab` to simulate a $\mathbb{S}^{D-1}-$indexed collection of skewed random variables, and then integrating them together with the skewness measure Γ.

If we wish to simulate a *symmetric* random vector, our task is even simpler; in this case, we can use the explicit formula I've written above in terms of random variables $\mathbf{E}$ and $\mathbf{U}$.

[4]Originally by John M. Chambers [77], adapted by John Nolan, 1992, modified by Vadim Teverovsky, 1993

I have written **C** code to simulate multidimensional stable random vectors. The vectors can potentially have arbitrary spectral measures, although the code is optimized to work for the most commonplace scenarios: purely atomic spectral measures, subgaussian stable vectors, and linear combinations of these two extremes.

A similar method is described in [79], whose authors have written a `FORTRAN` program for generating pseudorandom stable vectors. They confine themselves to the case when the spectral measure is a sum of atoms, which is sufficient because any spectral measure can be approximated by such a sum. The authors use the results of [174] to bound the error of this approximation. A review of these simulation techniques is provided in section 4 of [128].

Appendix C

The Prevalence of Power Laws

Most of the phenomena we encounter in nature exhibit probability densities with rapidly decaying tails. For example, the tails of the ubiquitous *normal distribution* decay like $\exp(-|x|^2)$ as $x \to \infty$. Densities with rapidly decaying tails have finite first and second moments, and hence, have a well-defined *mean* and *variance*. Intuitively, the *mean* of the distribution represents the "typical scale" of the phenomenon; the *variation*, represents the "typical degree of deviation" from this typical scale.

The ubiquity of the normal distribution, with its convenient mean and variation, has engendered a complacent assumption —or perhaps a hope —within the statistical sciences that, in the "real world", *all* distributions had such rapid tail decay. Hence, the normal distribution is the default used to model any empirically encountered random variable of unknown distribution, and *Gaussian processes* are the default model for any empirically encountered stochastic process.

Increasing evidence exists, however, that many phenomena in the real world do *not* exhibit distributions with rapid tail decay. Instead, these phenomena have tails which decay according to **power laws** of the form $x^{-\alpha}$ as $x \to \infty$, where α is a "scaling exponent", characteristic of the system. If $\alpha < 3$, then the distribution does not have a well-defined variance, and many of the $\mathbf{L}^2$-norm related tools of classical probability theory are rendered inapplicable. If $\alpha \leq 2$, then the distribution does not even have a well-defined mean.

As a consequence, these phenomena have no "typical scale"; we cannot identify events of a certain size as being "typical events". Indeed, if a distri-

bution really closely follows an $x^{-\alpha}$ power law, then it possesses the property of **scale invariance**; the distribution actually looks exactly the same when we perform a change of scale. Suppose, for example, that we were looking at the distribution of "lengths" of some object. Scale invariance means that, were the "tick marks" erased from the X axis, it would be impossible, from just looking at the graph of the distribution, to discern whether the X axis was on a scale of millimetres, metres, or kilometres.

Another consequence of power-law decay rates is the surprising frequency of *large events*. In a normal distribution, "large" events more than a few standard deviations away from the mean are so improbable that, for practical purposes, we can exclude them as impossibilities. In power-law distributions, however, arbitrarily large events occur with only slowly decaying frequency. When the phenomena in question have real human impact (such as, for example, earthquakes or stock market crashes), the frequency of large events must be modelled accurately enough to formulate sound policies of risk management.

C.1 Examples

The Gutenberg-Richter Law: The Gutenberg-Richter law is empirically observed power law in the distribution of earthquake sizes. Let E denote the total energy released during an earthquake event; this reflects the "size" of the earthquake. Let $P(E)$ be the the distribution of earthquake sizes along some particular fault. Gutenberg and Richter found that $P(E) \sim E^{-\alpha}$, where α is an exponent which ranges from 1.8 to 2.2, depending upon the fault being examined [76, 157, 158, 163, 148].

Although the Gutenberg-Richter law is accepted within the geological community, there is some controversy over its exact range of applicability. Earthquakes are propagate within the *schizosphere*, the part of the Earth's crust that is rigid enough to be ruptured by earthquake events. A fault can be imagined as a plane vertically bisecting the schizosphere. Smaller earthquakes propagate in two dimensions along the fault plane. Larger earthquakes, however, penetrate all the way to the bottom of the schizosphere; at this point, further earthquake propagation can only occur in one dimension, along the line of the fault. Thus, although both small and large earthquakes follow power laws, the exponent is different for small earthquakes than for large ones, reflecting the dimension of the fault geometry within which the earthquake occurs.

The Pareto Income Exponent: Vilfredo Pareto ([133], reprinted in [134]) was the first to observe that the distribution of personal incomes seems to follow a power law distribution, with the exponent α between 2 and 3. Various other authors have refined this estimate; for example, Mandelbrot [106] suggests that different exponents may apply to different "sectors" of the employment market (eg. salaried professionals vs. skilled wage earners vs. unskilled wage earners), and attempts to offer a theoretical explanation for why this is not only possible, but plausible. Others have rejected Pareto's claims; for example, Macaulay [103] strongly criticized Pareto, while Gibrat [54] asserted, contrary to Pareto, that the distribution of personal incomes is better modeled by a *log-normal* distribution.

Finance: In 1963, Benoit Mandelbrot published a seminal paper [108], in which he repudiated the classical "Brownian motion" model of speculative price movement first developed by Bachelier [14], and proposed instead that, at least in *certain* markets, the temporal variation of prices was better modelled using α-stable distributions, which he called "Pareto-Lévy " or "L-stable" law. This, for Mandelbrot, was the explanation for the surprising frequency of very large market events —a frequency which was inexplicable within Bachelier's Gaussian framework. In [108], Mandelbrot explores the theoretical consequences of his conjecture. In a sister paper, [107], he musters an impressive body of empirical evidence, demonstrating how empirical data from markets as diverse as cotton, wheat, and railroad stocks exhibit stable laws.

Although Mandelbrot's ideas were promulgated by himself and his student Fama [43, 42], and generated considerable interest at the time (earning a mention in Feller's encyclopaedic introduction to probability theory [46]), they soon fell out of favour, largely because the Bachelier's model was far more computationally tractable. The finite variance of normal distributions allows us to identify "market risk" with distribution variance: this forms the basis of the Sharpe-Markovitz theory of risk [161, 114]. The existence of Itô integrals for the Brownian process made possible the development of the Black-Scholes stochastic differential equations [18]. The well-defined *mean* of a normal distribution allows sensible discussion of the "expected future price" of a stock, and therefore allows us to meaningfully hypothesise that price processes are *martingales*. The martingale assumption is called the *Rational Expectations* hypothesis in the mathematical finance community, and is almost an axiom of financial mathematics.

Stable distributions, in contrast, do not generally have finite variance;

thus, we must invent a new conception of *risk*. Although stochastic integration with respect to stable processes is well-defined, it does not have all the nice properties of the Itô integral. For example: unlike Brownian motion, stable processes do *not* always have almost-surely continuous sample paths —a highly undesirable, "nonphysical" property. Finally, if the tail exponent is less than 1, a stable random variable does not even have a *mean*; a stable process generated by an accumulation of such random variables thus has no "expected future value" calling into question the very meaning of "Rational Expectations."

Recently, researchers have again started seriously exploring the idea that financial markets exhibit non-Gaussian price variations. Mandelbrot has published a book [105], summarising his work in the field over the last forty years. Inspired by the theory of *Self-Organized Criticality* (see § C.2 on page 162) originally suggested by Bak, Tang, and Wiesenfeld [136], many researchers have begun searching for "critical phenomena" in financial markets [8, 115], or within the economy in general [94, 137, 156].

New empirical work has confirmed Mandelbrot's early claims; for example, Rosario and Mantegna [112] have observed that the movements of the Standard & Poor's 500 index follow a "truncated" α-stable distribution with $\alpha \approx 1.4$.

Others have tried generalizing classical portfolio optimization techniques to this new, non-Gaussian framework. Generalizations of Sharpe-Markowitz portfolio theory under the hypothesis of *stable* asset distributions have been carried out by a variety of authors. As a natural measure of portfolio risk, all of these authors consider the *variation* of a stable distribution as the obvious surrogate for the variance used in the Gaussian regime. All of them work only with *symmetric* distributions, because there is no clear interpretation of "skewness" (β) of an asymetric distribution in terms of portfolio risk or return.

Samuelson [155] examined the case when the assets are assumed to be symmetric and independent (this corresponds to pure-atomic spectral measure with orthogonal atoms). Fama [44] worked with the Sharpe-Markowitz diagonal model. Press [142] generalized this to an arbitrary "pseudo-Gaussian" distribution (pure atomic spectral measure whose atoms are orthogonal relative to some inner product). These two cases where further studied by Ziemba [187]. Arad [10, 11] showed that the efficient portfolio set is convex. This allowed Belkacem, Lévy-Vé'hel, and Walter [99, 100, 68, 67] to extend the Capital Asset Pricing Model (CAPM) to a general multivariate symmetric distribution, making use of the concept of **covariation** (originally developed by [120], and discussed extensively in [154]) to develop an

equation describing the relationship between risk and return for an optimal portfolio. These results are extended by Gamba [52, 53], who develops an interpretation of portfolio minimization as a "norm minimization" problem. Other studies have been done by Bawa *et al.* [38] and McCulloch [117]. The subject has now become extensive enough that Embrechts, Klüppelberg, and Mikosch recently wrote a whole book on it [24].

Bouchaud, Sornette, Walter, and Aguilar [82] propose natural extensions of portfolio-optimization methodologies to a world where price variations are not only allowed to be non-Gaussian, but are further only constrained to satisfy the very weak hypothesis of power-law decay in the tails (see also [21] §3.3.2). The obervations of Bouchaud *et al.* therefore not only apply to stable distributions, but to any other power-law type distribution.

Zipf's Law: One famous early discussion of power law distributions was in the book *Human behavior and the principal of least effort : an introduction to human ecology* by George Kingsley Zipf [188]. Zipf ranked the major urban centres of the United States in decreasing order of population, and empirically observed what has come to be known as **Zipf's Law**: as $N \to \infty$, the Nth city on the list has a population proportional to $1/N$. Hence, for example, the "30th largest" city in the U.S. has a population approximately one thirtieth that of the largest city.

This observation has been confirmed many times, using urban population data from many countries (see [94], I.3). For small N, the data fits the curve rather poorly (the "second largest" city is *not* half as large as the largest). However, as N becomes large, the fit is almost exact.

Zipf noticed a similar pattern in many other distributions. For example, one gets an identical curve if one looks at the frequency of appearance of words in the English language: if one orders the words of the English language in decreasing order of frequency, then the Nth word in the list has a frequency proportionate to $1/N$.

To see how Zipf's Law corresponds to a power law, note that the "ranking" of a city corresponds to the value of the cumulative distribution of its inverse-population. Let the total number of cities in our sample be N, and let c be a constant so that, as n gets large, the nth largest city has population approximately c/n, and thus, has *inverse* population n/c. Zipf's law can then be restated: "For any n, the proportion of cities whose inverse-population is smaller than c/n is approximately n/N." In other words, if X is the inverse-population of a randomly chosen city, then the cumulative distribution of the random variable X is approximately $(c/N) \cdot X^{-1}$. Thus,

the probability density function is approximately $(-c/2N) \cdot X^{-2}$

At this point, the astute reader will realize that Zipf's law cannot *literally* be true for $n \to \infty$. In real life, there are a finite number of cities, and thus, there is a maximum n for which Zipf's law is meaningful. Stated another way: the minimum population of an "urban centre" is one person, so the urban centres of Zipf's law cannot get smaller than a certain size. Indeed, if we were to interpret Zipf's law as *literally* true for all $n \in \mathbb{N}$, we would end up concluding that the country had an infinite population, because we would be summing a harmonic series.

This illustrates a difficulty with power law distributions: there are often physical and mathematical reasons why the power law *must* become inapplicable at some very large scale. Physically, we realize that, although "large events" are possible, one cannot have events that are bigger than the system in question. One cannot have earthquakes bigger than the Earth, or Nile floods bigger than the entire Nile watershed, or market crashes bigger than the market, or an inverse-population bigger than 1 (and thus, a *population* smaller than one person). Mathematically, if we take a power law too literally, we may end up computing infinite values for quantities which must obviously be finite, such as total seismic energy dissipation, total volume of water displaced, total volume traded, or total national population.

The power-law distribution must have a "cut-off" at some point. The exact location of this "cut-off", and the mathematical consequences it entails, are subtleties which are not yet properly understood.

Internet Traffic patterns: Design of communications networks involves complex resource-allocation problems concerning the placement of routers, the capacities of packet-buffers, and the bandwidth of channels. In addition to these "static" allocation problems, one must confront the "dynamic" allocation problem of efficiently routing packets to maximize transmission speed while minimizing congestion and router overload.

Early mathematical models of Internet traffic were based on a simple extrapolation of the mathematical *teletraffic theory* which had been used by engineers to design telephone networks. Teletraphic theory is just a generalisation of queuing theory, and is based upon two assumptions:

- (A) Clients connect or disconnect to nodes in the network according to *Poisson processes* with constant rates.
- (B) Each client, while connected, uses a constant and predictable amount of network resources.

	Telephone	Internet
Clients	Humans	Computers
Purpose	Human conversation	Digital data transfer
Connection Behaviour	Only one connection at a time	Many connections simultaneously
Load per Client	Constant, predictable	Wildly Varying
Corrolation between clients	Almost none, except during emergencies	Large and unpredictable
"Typical" timescale of behaviour	One timescale	Many timescales

Figure C.1: A Comparison of Voice Telephony with Digital Communications

Implicit in (A) is that the behaviour of different clients is *independent*; hence, the individual fluctuations in network load are all independent events. Assumption (B) says that the average total network load will be linearly proportional to number of clients, and thus, related to the rates of the Poisson processes of assumption (A) in a simple way.

While this model had served the telephone system well for years, engineers soon found it to be totally inadequate for managing Internet traffic. Inspection revealed that the statistics of Internet traffic were totally incompatible with the Poisson model, while theoretical considerations suggested that there were fundamental differences between a digital communications network and a telephone network; these differences are summarized in table C.1

The "typical" timescale of human-to-human connections is relatively well-defined and invariant, assuming that most people have roughly similar telephone conversation styles. This time-scale takes the form of a relatively constant "rate", λ, for the Poisson processes defining (dis)connection events; λ is proportional to the *average call-length*. (To be sure, we can refine this assumption, asserting that there may be one time-scale for "personal" calls (usually placed during the evening) and another for "business" calls (placed during business hours). Nonetheless, these two time-scales differ by less than one order of magnitude.)

In contrast, computer-to-computer connections take place on many different time-scales, depending upon the *purpose of the connection* and the *nature of the data being transmitted*. For example, `HTTP` connections take the form of very brief requests, sent from a client to webserver, followed by

several relatively short but intense bursts of data from webserver to client, followed usually by a long period of "silence" on both ends. In contrast, an FTP file-transfer usually takes the form of a much longer, sustained period of constant, very high-intensity dataflow. On the other hand, text-based human-to-computer interactions (for example, via "telnet") involve an intermittent, very low-intensity stream of packets in one direction (user keyboard input) alternating with brief bursts in the other direction (server response). Human-to-human interactions (via ICQ , "talk", and other "online chat" mediums) involve intermittent, very low-intensity stream of packets in both directions.

For similar reasons, the load-per-client on a digital communications network can vary over many orders of magnitude. Consider the range of scales in the World Wide Web alone. The smallest, simplest web pages are text only, and occupy perhaps between 1 and 10 kilobytes of space. Larger web pages make use of inlined graphics, and require perhaps 100 to 1000 kilobytes, assuming the graphics is stored in some compressed format. The fanciest "multimedia" webpages make extensive use of sound and animation; even in highly compressed formats, a ten second soundfile occupies hundreds of kilobytes, and an interesting animation easily exceeds 10 megabytes. Because of this, a single mouse-click by a web-browsing individual can trigger file transfers ranging in scale over 5 orders of magnitude.

In contrast, the load-per-client on a telephone network is constant: each human-to-human connection occupies one circuit of the circuit-switched network, and requires (and gets) exactly the same amount of bandwidth as every other connection.

Also, human-to-human telephone connections are usually uncorrelated with one another, except during events of widespread public importance (ie. emergencies, breaking news items, etc.). In contrast, the design of digitial networks often results in "cascading" events. For example, when a client's web browser reads an HTML file from a web-server, that HTML file often contains embedded "links" to many other files, on other servers (containing, for example, graphics, "banner advertising", etc.). Hence, a single HTTP request can trigger a "cascade" of secondary HTTP requests; in this way, the activity on one webserver actually becomes correlated to that on many others.

Also, TCP/IP routing algorithms are designed to reroute packets to avoid congestion. Because of this, congestion in one "region" of a network can actually affect the routing of packets in neighbouring regions, creating further correlations. Finally, many forms of networked data are widely "mirrored". For example, USENET news data is mirrored on thousands of "news servers"

all over the network. This means that every USENET message, large or small, is duplicated thousands of times, resulting in a "multiplication" of any fluctuations in traffic patterns.

Empirical evidence [122, 121] indicates that Internet traffic statistics exhibit tails with power law decays. As in many other contexts where power laws manifest, Internet traffic also exhibits long duration correlations in time and apparent self-similarity on different scales. For a review, see [181].

Other Examples: Mandelbrot [109] catalogs many other cases where power law distributions have been found in empirical data. These include:

- The population distribution of mutant strains in a bacterial culture [160].
- The energy distribution of incoming cosmic rays [47].
- The relative frequencies of words in a human language, when words are "ranked" in decreasing order of frequency. This $1/f$ distribution was originally called the *Zipf-Mandelbrot Law*, and was the subject of much discussion in the early 1950's [188, 104, 110].
- The property damage caused by fires [97].
- Other financial risk distributions [183, 184, 185, 186].
- The distribution of scientific papers produced by a given scientist follows a power law with exponent $\alpha = 2$ ([101], see also p.77 of [105]).

Wentian Li of Rockefeller University maintains an excellent and very thourough online resource of information on power laws and $1/f$ systems at

`http://linkage.rockefeller.edu/wli/1fnoise/`.

In particular, he provides an extensive list of examples of power law behaviour in:

Electronic Devices	Ecological Systems	Heartbeats
Biology	Network Traffic	Traffic Flow
Granular Flow	Music and Speech	Neurological Systems
Human Coordination	Astronomy	Magnetic Systems
DNA Sequences	Chemical Systems	Geophysical Records
Number Systems	Radioactive Decay	Optical Systems
Leaking Faucets	Meteorology	Written Language
Work-Related Tardiness	Images	Financial Data

C.2 Origins and Causes

The ubiquity of the normal distribution in real phenomena is explained by the Central Limit Theorem, which basically says that a normal distribution is the inevitable outcome of any accumulation of a large number of small, independent random perturbations with finite variance. Hence, in the world of finite-variance processes, the normal distribution is "generic". We need a similar "generic" mathematical mechanism to explain the increasingly frequent manifestation of power law distributions in empirical data. Few mathematical or physical mechanisms have yet been proposed; we will review some of them here.

Self-Organized Criticality

In 1988, Per Bak, Chao Tang, and Kurt Wiesenfeld published a now famous paper [136] in which they suggested a novel paradigm they called *Self-Organized Criticality* ("SOC"). This term has been much used and abused since its introduction, and some semantic clarification is in order. To understand the meaning of "self-organized criticality", we must first understand, seperately, the terms "self-organized" and "criticality".

Critical phenomena are familiar to physicists studying condensed matter, and to mathematicians studying percolation theory and interacting particle systems. Unfortunately, the term means different things in different contexts, so there is some ambiguity attached to its usage. Critical phenomena usually arise in the context of *phase transitions*: sudden, discontinuous qualitative changes in a mathematical or physical system which arise when a control parameter crosses a certain threshold or *critical point*.

For example, in percolation theory [56, 89, 37], one randomly assigns the value of "0" or "1" to each site[1] in a lattice; one then studies the distribution of sizes of connected blocks of 1's in the lattice. Let $\lambda \in [0, 1]$ describe the proportion of sites which get the value 1. As λ increases from zero to one, the expected size of a connected block increases nonlinearly, and as λ exceeds a critical value λ_c, the expected size suddenly goes to infinity. At this critical point, a "phase transition" occurs; the system goes from a collection of seperate, isolated blocks of finite size, to a single, infinite connected component (with a few disconnected finite "satellites").

In solid-state magnetic physics, the canonical mathematical model is the

[1] Strictly speaking, this is so-called *site* percolation theory; in *bond* percolation theory, one assigns "0" and "1" to the *edges* connecting sites.

so-called *Ising Ferromagnet* [98]. Again, one begins with a lattice of sites; this time, each represents a micromagnet (say, an atom), which can assume one of two polarities or "spins" ("up" or "down"). Neighbouring magnets exert a force on one another which causes them to try to equalize their polarities: a state where neighbouring magnets have equal polarities has less potential energy than one where their polarities are opposite.

Left undisturbed, the system as a whole "wants" to assume one of the two global spin configurations of minimal energy: either all "up", or all "down". Confounding this uniformizing trend, however, is randomness introduced through thermal noise. We assume that the system is subjected to a certain amount of "heat", which causes the micromagnets to occassionally flip their spins in defiance of their neighbours.

In this model, the parameter λ represents the "temperature" of the system. When $\lambda = 0$, there is no thermal noise; the system crystalizes into either an all-up, or an all-down configuration. As $\lambda \to \infty$, the interaction between neighbouring sites becomes weaker and weaker, compared to the degree of thermal perturbation, until, when $\lambda = \infty$, the spins at each site are effectively independent random variables. At this point, the spin-configuration of the lattice is what, in ergodic theory, is called a *Bernoulli process*. In particular, as a stochastic process on $\mathbb{Z}^D$, the spin state is *ergodic*. Intuitively, this means that, starting from any spin-configuration, we can get arbitrarily close to any other spin-configuration, simply by *translating* the configuration far enough on the lattice.

In contrast, when the $\lambda = 0$, the system is *not* ergodic, because there are two invariant configurations: *all-up*, and *all-down*. If you are in the *all-up* configuration, you cannot attain the *all-down* configuration simply by translating the configuration across the lattice by some large amount.

So, if we "quench" an Ising ferromagnet by continuously decreasing the temperature parameter λ from ∞ towards 0, at some point, λ must cross a **critical value**, λ_c, when the system experiences a "breakdown of ergodicity". For large λ, there is really only one "state" the system as a whole can be in, from a statistical point of view[2], whereas for smaller λ, there are many. This breakdown of ergodicity is called a "phase transition" (for more information, see [98], chapt. IV).

Systems at a phase transition exhibit many interesting properties, which we collectively refer to as "critical phenomena". For one thing, as the system approaches the phase transition, we find spatial correlations occuring

[2]Formally speaking, what I mean is that the translation-invariant measure has no nontrivial invariant subsets.

over increasingly long distances. Consider the Ising ferromagnet. When $\lambda = \infty$, there is *no* correlation between even immediately adjacent sites. When $\lambda = 0$, there is *total* correlation between even very distant sites, since all sites must have identical spins. As λ goes from ∞ down to 0, the degree of correlation between neighbouring sites must increase. In particular, when λ is near the critical point, we find that the correlation between sites (as measured by the autocorrelation function of the configuration) decays over distance according to a *power law*. This is to be contrasted with an *exponential* decay rate when λ is large compared to the critical point. (For more information, see [76], sect. 2.4)

The problem, however, is that the critical phenomena are not "generic". The parameter λ has to be "tuned" to be close to the critical value, so that the critical state is inherently unstable. The claim of Bak *et al.* was that, in certain physical phenomena, the system *self-organizes* to a critical state, and, once it arrives there, it hovers on the edge of criticality forever. Thus, Bak *et al.* argue, far from being exotic, the critical state is in fact the *generic* state for certain systems.

The term *self-organization* has received a lot of use lately, due to the burgeoning study of "Complexity", which originated at the Santa Fe Institute ([63, 164, 75, 124, 125] etc.). Theorists studying emergent phenomena in complex systems have identified self-organization as the mystery ingredient needed to explain such enigmas as the origin of life [86, 85, 87, 149] the evolution of complex organisms [55] or the nature of cognition and consciousness [88]. In the popular science press, the term has assumed almost mystical connotations, which has hurt its credibility within the academic scientific community.

"Self-organization" at least the way it is used by Bak *et al*, however, simply refers to the existence of an *attractor* within some dynamical system. *Self-organized criticality* is a property of a (stochastic) dynamical system: it means that the *attractor subset* consists mainly of critical or near-critical states.

Bak *et al.* illustrate this with the famous "Sandpile Model". The idea is this: sand falls randomly onto a flat surface, and accumulates into a random landscape of hills and valleys. As long as the slope of the pile at some point is below a critical value, the force of friction is sufficient to keep the sand immobile. If the slope exceeds a certain critical value, however, then gravity will cause an avalanche to occur until enough sand has flowed from the hills into the valleys to bring the slope below the critical slope everywhere in the landscape.

Intuitively, the system is in a "near-critical" state if the slope is at or near the critical value at a large number of locations. In such a state, the addition of a single grain can cause the slope to exceed the critical value at some locale. A small avalanche will begin, transfering sand from this locale to neighbouring locales. If the neighbouring locales are also at the critical point, then they, in turn, will themselves avalanche. In this way, an avalanche can propagate through the landscape in a chain reaction, eventually involving a very large amount of sand.

If the landscape is *super-critical* (ie. the slope exceeds the critical value at any sites), then it will immediately collapse into widespread avalanche activity, until the slope is below the critical value everywhere. If the landscape is *sub-critical* (ie. the slope is far below the critical value everywhere), then the addition of new sand-grains will not trigger any avalanches, but will instead tend to increase the slopes, bringing the system closer to criticality. When the system is near-critical, avalanches are frequent events, and occur with a wide range of sizes. However, none is large enough to take the system very far away from criticality.

In other words, as sand is continually and randomly added to the system, it approaches the near-critical region, and then stays there. The system *self-organizes* to criticality.

Bak *et al.* never intended their model to be a description of *real* sandpiles[3]; the "sandpile" metaphor was only intended to stimulate intuition. Instead, Bak *et al.* imagined self-organized criticality as being a "universal" phenomena[4], arising, in one form or another, in a wide variety of physical processes. Since the paradigm was introduced, researchers have attempted to apply it, with varying degrees of success, to earthquakes [131, 61], economics [137, 156, 118, 119], finance [115], evolutionary biology [16], and superconductivity [151], as well a wide variety of cellular automata, such as Conway's famous Game of Life [15].

What does any of this have to do with power law distributions? In

[3]Hence, the flurry of experiments attempting to "empirically confirm" the "predictions" of the Sandpile model in real sandpiles [26, 62, 102], and rice-piles [167, 49, 139, 140] are sort of missing the point. In particular, those researchers who claim [26, 62] that Bak *et al.* have been "refuted" because real sand-piles do *not* exhibit exactly the critical behaviour described in [136] are definitely missing the point. For a summary of this experimental activity, see [76], §3.2, 3.3.

[4]Bak's own attempt, in his book *How Nature Works* [15], to promulgate self organizing criticality as some kind of "universal theory of complexity" are perhaps a little bit *too* ambitious in this regard, and have done nothing to assuage the skepticism of the scientific community about the status of his claims.

their original papers, Bak *et al.* describe self-organized criticality as "an explanation for $1/f$ noise" —indeed, this is the subtitle of their paper [138] in *Physical Review Letters*. "$1/f$ noise" is one of the older terms used to describe power law distributions[5]. One of the central theses of Bak *et al.* was that self-organized critical systems spontaneously exhibit power-law distributions in many observables. Their justification is entirely empirical in nature: they ran extensive computer simulations of the sand-pile model, along with an analogous "3-dimensional" model, and found that, if one empirically measured the distribution of avalanche sizes according to any of a number of criteria (total amount of sand displaced, total physical area covered by the avalanche, avalanche lifetime, etc.), one recovered power laws with various exponents.

Bak *et al.* thus propose SOC as an explanation for the power laws observed in processes such as earthquakes, stock markets, and Nile floods. Unfortunately, this explanation sometimes seems, to borrow Dennet's [35] words, nothing more than a *virtus dormitiva*[6] : it merely displaces the problem, rather than solving it. Now, instead of asking, "What is the origin of all these power-law distributions?" we ask, "What exactly *is* self-organized criticality? What are necessary and sufficient conditions for its manifestation?" No precise mathematical characterization of self-organized criticality exists, and thus, although many mathematical models "appear to exhibit" SOC in computer simulations, it is not clear how to prove *rigorously* that *any* of them do, or even if the question can be rigorously formulated. Indeed, it is clear from their writings that some researchers seem to consider "self-organized criticality" as *synonymous* with "power law distribution", hence making any explanation of the latter in terms of the former entirely circular.

Catastrophic Dissipation of a Constant Energy Influx

Common to many of the mathematical models exhibiting self-organized criticality is the following feature: "stress" or "energy" enters the system at a constant rate, and builds up within. This stress is periodically dissipated in

[5] In particular, the term describes the power spectra of certain stochastic process; the term "$1/f$" refers to the fact that power is inversely proportional to frequency. For some reason, the exponent is suppressed; the proper term should be "$1/f^{\alpha}$ noise", but this just wasn't as catchy.

[6] Dennet derives this term from a comedy by Moliére, called *La Malade Imaginaire* (1673), where, during a satirical medical-school exam, a character "explains" opium's ability to cause sleep by positing that opium has a *virtus dormitiva* (Latin: "sleep-causing power").

"catastrophic" events, and it is in the statistics associated with these catastrophic events that we find the power laws. For example, in the Sandpile Model, stress enters the system as sand piles up in increasingly precarious slopes; this stress is dissipated by intermittent avalanches. In SOC models of earthquakes, stress enters the system as two plates move relative to one another, and mechanical stress builds up along the "sticky" boundary between them. This stress is dissipated by intermittent slippage events at the boundary —-that is, earthquakes.

Perhaps we can duplicate the power law statistics of SOC by constructing a simplified model of this phenomena. We need four ingredients:

- (1) "Energy" enters the system at a constant rate, ρ. In other words, if $E(t)$ denotes the energy level at time t, then

$$\frac{dE}{dt} = \rho$$

  except during "catastrophes".

- (2) "Catastrophe" events occur according to some Poisson process, with instantaneous rate $\mathcal{J}(t)$, which possibly varies over time.

- (3) When a catastrophe occurs, it dissipates *all* the energy which has built up since the last event.

- (4) The "size" of the catastrophe is simply the amount of energy it dissipates.

We want the size distribution of catastrophes to follow a power law. What sort of instantaneous rate $\mathcal{J}(t)$ will we need to ensure this? By conditions (3) and (4), the size of a catastrophe is equal to the total energy in the system, at the moment the catastrophe occurs. By condition (1), this energy is linearly proportional to the length of time since the previous catastrophe. Hence, a power-law distribution of catastrophe *sizes* is equivalent to a power-law distribution of *inter-catastrophe intervals*.

It turns out that a necessary and sufficient condition for this is:

- (5) The instantaneous catastrophe rate is inversely proportional to the energy load in the system, up to some maximum rate M:

$$\mathcal{J}(t) = \min\left\{\frac{\alpha}{E(t)},\ M\right\}$$

Let C be the (random) size of a catastrophe. Thus, $C = \rho \cdot \tau$, where τ is the (random) length of time that has elapsed since the previous catastrophe.

Thus, assuming that $x > \frac{\alpha}{M}$, we have:

$$\begin{aligned}
\mathbf{Prob}\left[C = x\right] &= \mathbf{Prob}\left[\tau = x/\rho\right] \\
&= \exp\left(-\int_0^{\frac{x}{\rho}} \mathcal{J}(t)\ dt\right) \\
&= \exp\left(-\int_0^{\frac{x}{\rho}} \min\left\{\frac{\alpha}{E(t)},\ M\right\}\ dt\right) \\
&= \exp\left(-\int_0^{\frac{x}{\rho}} \min\left\{\frac{\alpha}{\rho \cdot t},\ M\right\}\ dt\right) \\
&= \exp\left(-\int_0^{\frac{\alpha}{\rho M}} M\ dt \quad - \quad \int_{\frac{\alpha}{\rho M}}^{\frac{x}{\rho}} \frac{\alpha}{\rho \cdot t}\ dt\right) \\
&= \exp\left(\frac{-\alpha}{\rho} \ - \ \frac{\rho}{\alpha} \cdot \int_M^{\frac{x}{\alpha}} \frac{1}{s}\ ds\right) \\
&= \exp\left(\frac{-\alpha}{\rho} \ - \frac{\rho}{\alpha} \cdot \left[\ln\left(\frac{x}{\alpha}\right) - \ln(M)\right]\right) \\
&= \left(e^{-\alpha/\rho}(\alpha M)^{-\rho/\alpha}\right) \cdot x^{-\rho/\alpha} \\
&= K \cdot x^{-\beta}
\end{aligned}$$

where $K = \left(e^{-\alpha/\rho}(\alpha M)^{-\rho/\alpha}\right)$, and $\beta = \frac{\rho}{\alpha}$. Notice that we must assume that M is finite; if we let M go to infinity, then the integral of $\mathcal{J}(t)$ is infinite, which forces catastrophe intervals to have size zero: the system is in a "permanent state of catastrophe", and all catastrophe events have size zero. By making M finite, however, we impose a minimum time-scale on which the power law applies: for extremely short time scales (ie. extremely "small" catastrophe events), the distribution of catastrophe sizes is exponentially decaying, rather than a power law.

Thus, an extremely simple mathematical model can reproduce the power law phenomena. What physical interpretation can we give to condition (5)? It seems that, paradoxically, the more energy we load onto the system, the more stable the system becomes.

Random Exponential Growth I: Bursting Bubbles

There is a another simple "catastrophe" model which exhibits a power law distribution. We will describe it in terms of stochastically bursting "bubbles". These bubbles satisfy the following two axioms:

- Bubbles are born, grow, and die in succession. Each bubble is born at the moment the previous bubble bursts.
- When a bubble is born, it has initial size 1.
- After a bubble is born, it grows exponentially over time. In other words, if $E(t)$ is the size of the bubble at time t, then E satisfies the differential equation:
$$\frac{dE}{dt}(t) = \beta \cdot E(t)$$
so that, if $E(0) = 1$, then, for any $t > 0$, $E(t) = e^{\beta t}$.
- Bubbles burst according to a Poisson process with constant instantaneous rate α. In other words, if τ is the (random) lifetime of a bubble, then for any $t > 0$,
$$\frac{d}{dt}\mathbf{P}_{\text{bub}}\left[\tau \leq t\right] = \alpha \cdot e^{-\alpha t}$$
- When a bubble "bursts", it releases energy in direct proportion to its size. These intermittent releases of energy are the "catastrophes" of the system.

Let's look at the distribution of "catastrophe sizes". Let C be the (random) size of the catastrophe associated with a particular bubble. Thus, $C = e^{\beta\tau}$, where τ is the (random) lifetime of the bubble. So, for any $x > 1$,

$$\begin{aligned}
\frac{d}{dx}\mathbf{P}_{\text{bub}}\left[C \leq x\right] &= \frac{d}{dx}\mathbf{P}_{\text{bub}}\left[e^{\beta\tau} \leq x\right] \\
&= \frac{d}{dx}\mathbf{P}_{\text{bub}}\left[\tau \leq \frac{\ln(x)}{\beta}\right] \\
&= \frac{1}{\alpha}e^{-\alpha\frac{\ln(x)}{\beta}}\frac{1}{\beta}\left(\frac{d}{dx}\ln\right)(x) \\
&= \frac{1}{\alpha\beta}x^{-\frac{\alpha}{\beta}}\frac{1}{x} \\
&= K \cdot x^{-\gamma}
\end{aligned}$$

where $K = \frac{1}{\alpha\beta}$, and $\gamma = \frac{\alpha}{\beta} + 1$. Note that, again, we must impose a minimal "cut-off" in order to get sensible results. This time, we do this by assuming that all bubbles begin with a minimal size of 1.

Random Exponential Growth II: Simon's Model of Urban Growth[7]

Another "random exponential growth" model that exhibits power laws is based on Herbert Simon's [162], proposed explanation for Zipf's Law [188] of urban sizes (see Section C.1 on page 157).

Simon makes the following assumptions:

- New "cities" are created at a constant rate α, and begin with a population of 1.
- New "citizens" arrive in the country one at a time, and each chooses a city to inhabit. The probability that a given citizen will pick a certain city as their home is directly proportional to its size[8].

The growth rate of each individual city is therefore random and roughly exponential. In Simon uses this to deduce a kind of "discrete Fokker-Planck" equation for the distribution of city sizes, treated as a probability distribution on $\mathbb{N}$. He concludes that the only stationary distribution of this Fokker-Planck equation is one satisfying a power law. (Krugman provides more detailed discussion of Simon's model in [94].)

Here I will present a generalization of Simon's model. For simplicity, we will make the following assumption, which, although ridiculous if interpreted too literally, are mathematically convenient.

- Instead of parameterizing the "size" of each individual city *discretely* (on $\mathbb{N}$), we will parameterize it *continuously*, (on $[0, \infty)$). This will allow us to make explicit use of the Fokker-Planck formalism.
- Instead of assuming new cities are created at a constant rate, with an initial "size" of 1, we will assume that there is an infinite "reservoir" of cities of size zero; all cities begin with zero size and grow over time.

[7] I am grateful to Don McLeish for pointing out some errors in an earlier version of this material.

[8] We can explain this as follows: if new citizens are "born", then, assuming the birthrate is the same in all cities, the population growth rate of each city is directly proportional to its size. If new citizens are "immigrants", then they are likely to settle in a city where they already know someone; the probability that a given immigrant has a contact in a given city is again directly proportional to its size.

- There are an infinite population of cities; the distribution of their *logarithmic* sizes at time t is given by a measure, absolutely continuous with respect to Lebesgue measure, with Radon-Nikodym derivative $\Phi_t : [0, \infty) \longrightarrow [0, \infty)$. In other words, for any $L_1 < L_2$, the number of cities whose *log-population* lies between L_1 and L_2 is given by:

$$\int_{L_1}^{L_2} \Phi_t(\ell)\ d\ell$$

We will also assume:

- Each city grows exponentially over time, but each is also subjected to an white-noise perturbation, in direct proportion to its size. In other words, if X_t represents the size of a city at time t, then X satisfies the stochastic differential equation:

$$\partial_t X_t \ = \ \alpha \cdot X_t \ + \ \gamma X_t \, d\mathcal{B}_t,$$

where $\mathcal{B}_t$ is a version of Brownian Motion. Thus, if $L_t = \log(X_t)$, then L_t satisfies:

$$\partial_t L_t \ = \ \alpha \ + \ \gamma \, d\mathcal{B}_t$$

- The white noise processes perturbing each city are independent.

Thus, as a function of time, Φ must obey the Fokker-Planck PDE:

$$\partial_t \Phi_t \ = \ \alpha \cdot \partial_L \Phi_t + \gamma \triangle \Phi_t \tag{C.1}$$

Suppose that Φ is a fixed point of this PDE. Then Φ must satisfy the ODE:

$$0 \ = \ \partial_t \Phi \ = \ \alpha \cdot \partial_L \Phi + \gamma \triangle \Phi$$

The general solution to this ODE is:

$$\Phi(L) \ = \ c_1 + c_2 \exp\left(-\frac{\alpha}{\gamma} L\right) \tag{C.2}$$

where c_1 and c_2 are arbitrary constants.

Suppose that $\Psi : [1, \infty) \longrightarrow [0, \infty)$ is the Radon-Nikodym derivative of the distribution of (nonlogarithmic) urban populations —in other words, for any $X_1 < X_2$, the number of cities whose population lies between X_1 and X_2 is given by:

$$\int_{X_1}^{X_2} \Psi(x)\ dx$$

Thus, by a simple change of variables $\ell = \log(x)$, we must have:

$$\begin{aligned}\Psi(x) &= \Phi(\log(x)) \left(\frac{\partial}{\partial x} \log\right)(x) \\ &= \frac{\Phi(\log(x))}{x} \\ &= \frac{c_1 + c_2 \exp\left(-\frac{\alpha}{\gamma}\log(x)\right)}{x} \\ &= \frac{c_1}{x} + \frac{c_2}{x^{1+(\alpha/\gamma)}},\end{aligned}$$

a sum of two power-law functions.

Random Partitions

Power law distributions can also arise when studying combinatorial structures which have a canonical decomposition into "irreducible" components. Consider, for example:

1. The decomposition of a natural number n into prime factors.
2. The decomposition of a polynomial $f(x)$ in over the field $\mathbb{Z}_{/p}$ into a product of irreducible polynomials.
3. The decomposition of a directed graph $\mathbf{G}$ into connected components.
4. If $f : \mathbf{X} \longrightarrow \mathbf{X}$ is a map on a finite space, then we can decompose $\mathbf{X}$ into f-invariant subsets. This is really a special case of #2, where we think of $\mathbf{X}$ as having a digraph structure with edges of the form $x \mapsto f(x)$.
5. The decomposition of a permutation σ of $\{1, 2, \ldots, N\}$ into cyclic permutations (this is really a special case of #3).

In all of these examples, the following ingredients are present:

- A natural measure of the "size" of the object in question. In the examples above, this size is given, respectively, by:
  1. $\ln(n)$.
  2. **degree**$[f]$.
  3. $\mathcal{C}_{ard}\,[\mathbf{G}]$.

4. $\mathcal{C}_{and}[\mathbf{X}]$.
5. N.

- A finite set of objects in a given size range: (eg. all graphs on N vertices, all permutations of N elements, etc.)

- A natural measure of the size of the components of the resulting decomposition, so that we have the formula:

$$|\mathbf{X}| \;=\; |\mathbf{X}_1| + |\mathbf{X}_2| + \ldots + |\mathbf{X}_N|$$

where $|\mathbf{X}|$ represents the size of the original object, and $\mathbf{X}_1, \ldots, \mathbf{X}_N$ are its component parts.

Suppose $\mathcal{S}_N$ is the (finite) set of all objects of size N, and we pick $\mathbf{X}$ randomly from S_N (with a uniform probability distribution). For all $n \in [1..N]$, let $\#_n(\mathbf{X})$ be the number of components of $\mathbf{X}$ of size exactly n. Then $\#_1(\mathbf{X})$, $\#_2(\mathbf{X}), \ldots, \#_N(\mathbf{X})$ are random variables, satisfying the relation:

$$\#_1(\mathbf{X}) \;+\; 2 \cdot \#_2(\mathbf{X}) \;+\; 3 \cdot \#_3(\mathbf{X}) \;+\; \ldots + \; N \cdot \#_N(\mathbf{X}) \;\;=\;\; N$$

Define $\mathbf{C}_n^{[N]} \;=\; \#_n(\mathbf{X})$, where $\mathbf{X}$ is drawn uniformely from $\mathcal{S}_N$.

We can regard $\mathbf{C}_1^{[N]}$, $\mathbf{C}_2^{[N]}, \ldots, \mathbf{C}_N^{[N]}$ the first N elements of an infinite sequence of random variables, where we define $\mathbf{C}_n^{[N]} = 0$ for all $n > N$. We might then ask what the limiting distribution of this sequence of random variables is, as $N \to \infty$.

In many cases, it turns out that the random sequence $\left[\mathbf{C}_n^{[N]}|_{n=1}^{\infty}\right]$ converges, in distribution, to a random sequence of the form $[\mathbf{C}_n|_{n=1}^{\infty}]$. If we then define a random noise

$$\mathcal{M}_N \;=\; \sum_{n=1}^{\infty} \mathbf{C}_n \cdot \delta_{(n/N)}$$

and let $\mathcal{M} = \lim_{N\to\infty} \mathcal{M}$, then, in many cases, $\mathcal{M}$ is a **scale-invariant Poisson measure**, with intensity $\rho(x) = x^{-1}$ (ie. the expected number of points in interval (a, b) is $\ln(b/a)$.) Hence, once again we see the emergence of a "$1/f$" distribution out of a natural and fairly universal class of random processes.

The discussion here is loosely based on the survey article [5], which provides an extensive bibliography of related literature.

Other Explanations:

Mandelbrot [109] catalogs several other explanations for power laws, none of which he finds particularly satisfactory, perhaps because none of them assigns any special role to *fractal* or *self-affine* phenomena, which Mandelbrot himself favours. These include:

- "Trivial" explanations, whereby a power law distribution is an immediate consequence of the mathematical relationships between certain random distributions. For example:

  - The ratio of two independent standard normal random variables is has a Cauchy stable distribution (hence, the tails decay with order $1/f$).
  - The ratio of two independent exponential random variables has a power law distribution.

  Hence, if we can plausible represent the observable in question as a ratio of two other random quantities with the aforementioned distributions, then its power law behaviour follows immediately. Mandelbrot apparently considers such explanations too tautological to be interesting.

- What Mandelbrot calls "elimination of an intrinsic variable between two intrinisically meaningful quantities". The "bursting bubbles" model we provided above (page 169) is an example of this.

- What Mandelbrot calls "random proportionate effect" models, which involve *logarithmic Brownian motion*.

  Brownian motion is normally explained by imagining the observable $\mathbf{X}$ as a sum of a very large number of microscopic "perturbations". The justification for a *logarithmic Brownian motion* is similar, except that now, we imagine that the scale of these "perturbations" is growing or shrinking in size, in proportion to the current scale of $\mathbf{X}$.

  For example, we expect the fluctuations in the luminosity of a large star to occur on a larger scale than those of a small star, or the fluctuations in the market-capitalization of a large firm to occur on a larger scale than those of a smaller firm. If a firm has a market capitalization of \$10 billion, then a fluctuation of \$50 million is not all that shocking; on the other hand, if the firm has a capitalization of \$80 million, then

a fluctuation of \$50 million represents either phenomenal growth or a catastrophic loss of shareholder confidence.

In this context, it seems more sensible to measure fluctuations in terms of *percentage*, rather than in terms of absolute numerical figures: for both the "large" firm and the "small" one, a fluctuation of 2% in market value has roughly the same significance. Once we switch to percentages, we are essentially dealing in logarithmic units; hence if we want to model evolution using a Wiener process, we are essentially dealing with log-Brownian motion

Notwithstanding this argument, Mandelbrot rejects log Brownian motion. This seems to be part of a more general rejection ([105], p. 112) of *any* distribution obtained obtained by "transforming" the variable $\mathbf{X}$ into a new variable $\mathbf{Y} = \phi(\mathbf{X})$, which Mandelbrot makes on three grounds:

- Suppose we are simultaneously examining several empirical processes $\mathbf{X}_1, \ldots, \mathbf{X}_N$. Aside from the seperate univariate distributions of $\mathbf{X}_1, \ldots, \mathbf{X}_N$, the other question of interest is the nature of the *dependency*, if any, between these observables. If we apply some transform ϕ, to get a new collection of random variables $\mathbf{Y}_1 = \phi(\mathbf{X}_1), \ \ldots, \mathbf{Y}_N = \phi(\mathbf{X}_N)$, then we may end up with more convenient univariate distributions, but we may confound and distort the *dependency structure* of the data in the process.
- As previously mentioned, the process $\mathbf{X}$ may plausibly be generated by *additive* perturbations, if it represents a quantity where addition is actually *physically meaningful* (for example: mass, energy, income, momentum, etc.). Hence, it is sensible to expect $\mathbf{X}$ to evolve according to some Gaussian process, because of the Central Limit Theorem. However, the quantity $\mathbf{Y} = \pi(\mathbf{X})$ is probably *not* additive (what is the physical meaning of adding *log-incomes*, or *log-momentums*?). Hence, there is no plausible reason to expect $\mathbf{Y}$ to have a Gaussian distribution; indeed, hypothesising that $\mathbf{Y}$ is Brownian actually raises more questions than it answers.
- In real science, we never know the distribution of $\mathbf{X}$ exactly, but only "approximately"; we use statistical tests to identify a "neighbourhood" of plausible distributions for the data. If we apply our statistical methods instead to some transformed data

$\mathbf{Y} = \phi(\mathbf{X})$, then we had best be very aware of how these "distribution neighbourhoods" are transformed by ϕ. Such care is rarely taken.

- What Mandelbrot calls "random proportionate growth" models. Herbert Simon's model of urban growth (page 170) is an example of this.

C.3 Power Laws and Stable Random Variables

Stable random variables arise naturally as "suitable renormalized limits" of summations of independent random variables exhibiting power law distributions. If $\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3, \ldots$ are random "perturbations" obeying a power law with exponent $-1-\alpha$ (for $\alpha \in [0,2]$) , then the random variables

$$\mathbf{X}_N \quad := \quad \frac{1}{N^{1/\alpha}} \sum_{n=1}^{N} \mathbf{Y}_n$$

converge, in distribution, to a stable random variable with "stability exponent" α. Hence, the ubiquity of power laws in physical phenomena suggests an important role for stable distributions.

Appendix D

Stable Stochastic Processes

D.1 Preliminaries

Suppose that $\mathbf{X}(t)$ ($t \in \mathbb{T}_{ime}$) is some real-valued stochastic process, and imagine that $\mathbf{X}(t)$ evolves by an accumulation of independent perturbations; that is, for any $T > 0$, and any N, we can write:

$$\mathbf{X}(T) \;=\; \sum_{n=1}^{N} \Delta\mathbf{X}_n$$

where $\Delta\mathbf{X}_1, \ldots, \Delta\mathbf{X}_N$ are independent random variables. If we assume that these random perturbations have finite variance, then, as N gets large, the variable $\mathbf{X}(t)$ is likely to become Gaussian in distribution.

Because of this, **Gaussian processes** are the most widely used models of evolving stochastic phenomena. Two, in particular, are well-known

- The **Ornstein-Uhlenbeck process** is the unique Gaussian process $\mathbf{X}(t)$ ($t \in \mathbb{R}$) which is *stationary*, *Markovian*, and has *continuous sample paths*.

- The **Wiener process** (also called **Brownian motion**) is the unique Gaussian process $\mathbf{X}(t)$ ($t \geq 0$) having *continuous sample paths* and *stationary, independent increments*. The *increments* of a process are the random variables of the form $\mathbf{X}(t \to T) = \mathbf{X}(T) - \mathbf{X}(t)$, where $t < T$. To say that Brownian motion has stationary, independent increments is to say that, for every $\delta > 0$, the random variables

  $$\ldots, \mathbf{X}(-2\delta \to -\delta), \quad \mathbf{X}(-\delta \to 0), \quad \mathbf{X}(0 \to \delta), \quad \mathbf{X}(\delta \to 2\delta), \ldots$$

  are independent and identically distributed.

Gaussian processes have several convenient properties:

- For all $T \in \mathbb{T}_{ime}$, the variable $\mathbf{X}(T)$ has well-defined, finite mean and variance.
- The increments of $\mathbf{X}$ are also Gaussian, and have finite variance.
- The increment $\mathbf{X}(t \longrightarrow T)$ tends to be **Fickian** as a function of the time difference $(T - t)$, in the sense that the variance of $\mathbf{X}(T) - \mathbf{X}(t)$ grows proportionate to $\sqrt{T - t}$. Stated another way, Gaussian processes exhibit "very rapid" motion on a short time scale but quite "slow" motion on longer time scales.
- In particular, very "large" movements are extremely rare.
- When larger than average movements occur, they are usually quickly "washed out" by the noise that follows. Short-term behaviour tends to make little difference in the long term. This is related to the fact that, in a summation of independent normal variables, no single term is likely to dominate the whole sum. Because of this property, Mandelbrot [111] characterises the randomness of Gaussian processes as "mild".
- For any times $t_1, t_2 \in \mathbb{T}_{ime}$, the **covariance** $\mathbf{E}^{xpct}[\mathbf{X}(t_1) \cdot \mathbf{X}(t_2)]$ is well-defined. The **autocorrelation** function $R(t_1, t_2) = \mathbf{E}^{xpct}[\mathbf{X}(t_1) \cdot \mathbf{X}(t_2)]$ is easy to compute, and completely characterises the Gaussian process.
- The autocorrelation function provides a complete description of the *dependence structure* of the process, allowing us to easily quantify the extent to which the process exhibits *long memory*.

So familiar are these properties that analysts have come to take them for granted. It comes as a real surprise, therefore, when empirically observed processes fail to satisfy them. The increasing evidence for such "anomalous" behaviour has motivated interest in *non-Gaussian* processes. The conception of a natural phenomenon as an "accumulation of perturbations", however, still suggests that $\mathbf{X}(t)$ should be a *stable* random variable. Hence, we are led to look at **stable stochastic processes**.

Definition 87: *α-stable stochastic process*

Let $\alpha \in [0,2]$. Let $\mathcal{X} : \mathbb{T}_{time} \longrightarrow \mathbb{R}^D$ be a stochastic process, where $\mathbb{T}_{time}$ is some space parameterizing "time". $\mathcal{X}$ is an **α-stable process** *if, for all $t_1, t_2, \ldots, t_N \in \mathbb{T}_{time}$, the random vector $(\mathcal{X}(t_1), \mathcal{X}(t_2), \ldots, \mathcal{X}(t_N)) \in \mathbb{R}^{D \times N}$ has an $\alpha-$stable distribution.*

For our purposes, $\mathbb{T}_{time} = \mathbb{R}$ or $\mathbb{T}_{time} = [0, \infty)$.

Stable stochastic processes behave, in many ways, quite differently than Gaussian processes:

- The random variable $\mathbf{X}(T)$ may not have a well-defined variance (if $\alpha < 2$) or even mean (if $\alpha < 1$).
- Increments also may not have well-defined mean or variance.
- Given $t_1, t_2 \in \mathbb{T}_{time}$, the covariance $\mathbf{E}^{xpct}[\mathbf{X}(t_1) \cdot \mathbf{X}(t_2)]$ is no longer well-defined, and thus, there is no longer a well-defined autocorrelation function. Formal analogs of the autocorrelation function (for example $R(t_1, t_2) = \mathbf{E}^{xpct}\left[\mathbf{X}(t_1)^{1/\alpha} \cdot \mathbf{X}(t_2)^{1/\alpha}\right]$) are not necessarily computationally tractable, and do not completely describe the process anyways.
- There is no simple description of the *dependence structure* of the process. Stable processes can exhibits *long memory* in ways that are subtle and difficult to quantify.
- The increment $\mathbf{X}(T) - \mathbf{X}(t)$ is no longer "Fickian" as a function of the time difference $(T - t)$. Informally speaking, the "typical size" of $\mathbf{X}(T) - \mathbf{X}(t)$ grows proportionate to $|T - t|^{1/\alpha}$. As α decreases from 2 towards 0, an $\alpha-$stable process increasingly tends to exhibit "slower" movements in the short term (relative to the Gaussian), but "larger" movements in the long term. This seems to require some sort of "coherence" or "bias" amongst the seperate increments, but it doesn't. The seperate increments can be quite independent of one another; large long-term movements are usually caused by one or a few increments which are so large (relative to the rest) that they completely dominate the sum.
- In other words, occassionally, short-term behaviour can be "extreme" enough to have a long-lasting legacy. This is because in a sum of independent stable random variables, a single summand can often dominate the whole sum. This property defines what Mandelbrot [111] calls "wild" randomness.

- As a result, very "large" movements are not uncommon.

Like Gaussian processes, stable processes often have **stationary increments**, or and are sometimes **stochastically self-affine**, and these can be useful properties.

Definition 88: *Stationary Increments*

Let $\mathcal{X} : \mathbb{T}_{ime} \longrightarrow \mathbb{R}^D$ *be a stochastic process, where* $\mathbb{T}_{ime} \subseteq \mathbb{R}$. *Assume for simplicity that* $0 \in \mathbb{T}_{ime}$.

$\mathcal{X}$ *has* **stationary increments** *if, for any* $t, T \in \mathbb{T}_{ime}$, *and any* s *such that* $T - t$ *is also in* $\mathbb{T}_{ime}$, *the increment:*

$$\mathcal{X}(t \to T) \;=\; \mathcal{X}(T) - \mathcal{X}(t).$$

is identically distributed with the increment

$$\mathcal{X}\left(0 \to (T-t)\right) \;=\; \mathcal{X}(T-t) - \mathcal{X}(0).$$

If $\mathcal{X}$ is α-stable, with stationary increments, we will say that $\mathcal{X}$ is "α-SI".

Definition 89: *Stochastically Identical*

Let X *be some statespace,* $\mathbb{T}_{ime}$ *some timeline, and let* $\mathcal{X}_1 : \mathbb{T}_{ime} \longrightarrow X$, *and* $\mathcal{X}_2 : \mathbb{T}_{ime} \longrightarrow X$, *be two stochastic processes.* $\mathcal{X}_1$ *and* $\mathcal{X}_2$ *are* **stochastically identical** *if, for any finite subset* $\{t_1, \ldots, t_N\} \subset \mathbb{T}_{ime}$, *the finite dimensional distributions*

$$\mathbf{D}^{istr}\left[(\mathcal{X}_1(t_1), \ldots, \mathcal{X}_1(t_N)) \in X^N\right]$$

and

$$\mathbf{D}^{istr}\left[(\mathcal{X}_2(t_1), \ldots, \mathcal{X}_2(t_N)) \in X^N\right]$$

are identical as probability measures on X^N.

Definition 90: *Stochastic Self-Affinity*

Let $H \in (0, \infty)$. *Let* $\mathcal{X} : \mathbb{T}_{ime} \longrightarrow \mathbb{R}^D$ *be a stochastic process, where* $\mathbb{T}_{ime}$ *is some space parameterizing "time".* $\mathcal{X}$ *is* **stochastically self affine** *with* **self-affinity exponent** H *(for short, "*H**-self-affine**") *if the process* $\lambda^{-H}\mathcal{X}(\lambda t)$ *is a* version *of* $\mathcal{X}$.

If $\mathcal{X}$ is H-self affine, we will sometimes say that $\mathcal{X}$ is "H-SA". An H-self affine, α-stable process with stationary increments will be called H-SA-α-SI.

For example, Brownian motion is $\frac{1}{2}$-SA, and 2-SI.

D.2 Stochastic Integration and Stochastic Noise

Let $(X, \mathcal{X}, \nu)$ be a measure space, and let $\mathbf{L}(X)$ be some vector space of $\mathcal{X}$-measurable functions from X to $\mathbb{R}$. A **random noise** on X is a "random measure" on the space X, which we can use to "integrate" elements of $\mathbf{L}(X)$. In other words, a noise is a random linear functional $\mathcal{N} : \mathbf{L}(X) \longrightarrow \mathbb{R}$.

A linear functional is an element of the space $\mathbb{R}^{\mathbf{L}(X)}$; hence, the probability distribution of $\mathcal{N}$ is some measure on $\mathbb{R}^{\mathbf{L}(X)}$. We will use the simplest sigma-algebra available for $\mathbb{R}^{\mathbf{L}(X)}$: the sigma-algebra generated by all evaluation maps

$$\begin{array}{rcl} \mathbf{pr}_f : \mathbb{R}^{\mathbf{L}(X)} & \longrightarrow & \mathbb{R} \\ \mathcal{N} & \mapsto & \mathcal{N}(f) \end{array}$$

According to the Kolmogorov Consistency Theorem, to specify a probability measure on $\mathbb{R}^{\mathbf{L}(X)}$, it suffices to "consistently" specify its finite-dimensional marginals. Thus, for any finite subset $\{f_1, f_2, \ldots, f_N\} \subset \mathbf{L}(X)$, we want to define the probability measure $\mathbf{P}_{\{f_1, f_2, \ldots, f_N\}} \in \mathcal{MEAS}\left[\mathbb{R}^N\right]$ where, for all $U_1, U_2, \ldots, U_N \subset \mathbb{R}$,

$$\mathbf{P}_{\{f_1, f_2, \ldots, f_N\}}(U_1, U_2, \ldots, U_N) \;=\; \mathbf{P}_{\mathcal{N}\mathcal{O}}\left[\mathcal{N}(f_1) \in U_1,\; \mathcal{N}(f_2) \in U_2,\; \ldots \mathcal{N}(f_N) \in U_N\right]$$

The Kolmogorov Consistency condition we must satisfy is this: any $f_0 \in \mathbf{L}(X)$,

$$\mathbf{P}_{\{f_0, f_1, f_2, \ldots, f_N\}}(\mathbb{R}, U_1, U_2, \ldots, U_N) \;=\; \mathbf{P}_{\{f_1, f_2, \ldots, f_N\}}(U_1, U_2, \ldots, U_N)$$

Definition 91: *α-stable Noise*

Let $(X, \mathcal{X}, \nu)$ be a measure space. If $\alpha \neq 1$, then let

$$\mathbf{L}^{\langle\alpha\rangle}(X) \;=\; \left\{ f : X \longrightarrow \mathbb{R}\;;\; f \text{ is measurable, and } \int_X |f|^\alpha \, d\nu < \infty \right\}$$

whereas, if $\alpha = 1$, then let

$$\mathbf{L}^{\langle 1\rangle}(X) \;=\; \left\{ f : X \longrightarrow \mathbb{R};\;\; f \text{ is measurable, } \int_X |f| \, d\nu < \infty, \right.$$
$$\left. \text{and } \int_X |f| \log |f| \, d\nu < \infty \right\}$$

Let $\beta \in [-1, +1]$. An **α-stable random noise** *on X with* **skewness**[1] *β is a random linear function $\mathcal{N} : \mathbf{L}^{\langle\alpha\rangle}(X) \longrightarrow \mathbb{R}$ whose finite dimensional marginals are multivariate α-stable random variables with the following characteristic functions.*

First, let $\eta_{\alpha;\beta} : \mathbb{R}^N \times \mathbb{R}^N \longrightarrow \mathbb{C}$ be defined:

$$\eta_{\alpha;\beta}(\vec{\xi}, \mathbf{s}) \;=\; \left|\left\langle \vec{\xi}, \mathbf{s} \right\rangle\right|^{\alpha} + \mathcal{B}_{\alpha}\beta \left\langle \vec{\xi}, \mathbf{s} \right\rangle^{\langle\alpha\rangle} \mathbf{i}.$$

If $\{f_1, f_2, \ldots, f_N\} \subset \mathbf{L}^{\langle\alpha\rangle}(X)$, let $F : X \longrightarrow \mathbb{R}^N$ be defined:

$$F(x) \;=\; [f_1(x), \ldots, f_N(x)]\,,$$

If $\Phi_{\{f_1,f_2,\ldots,f_N\}}$ is the log Fourier transform of $\mathbf{P}_{\{f_1,f_2,\ldots,f_N\}} = \mathbf{D}^{istr}[F]$, then, for all $\vec{\xi} \in \mathbb{R}^N$,

$$\Phi_{\{f_1,f_2,\ldots,f_N\}} \;=\; -\int_X \eta_{\alpha;\beta}\left(\vec{\xi}, F(x)\right)\, d\nu[x] \qquad (*)$$

Remark 92: *ν as Intensity Measure*

Note that the definition of the noise $\mathcal{N}$ depends upon the measure ν on the space X. Different choices of ν will produce different noises. In some sense, ν defines the "loudness" of the noise at each point in X; if ν is concentrated in a certain region of X, then a function $f \in \mathbf{L}(X)$ supported in that region will tend to "pick up more noise" than a function g supported elsewhere, and as a result, $\mathcal{N}(f)$ will be have a larger *variation* than $\mathcal{N}(g)$.

If (X, ν) is a space parameterising "time" (for example, $(\mathbb{R}, \mathcal{L}^{lsg})$) then the measure ν can be thought of as measuring the rate at which time is passing. The choice of Lebesgue measure $\mathcal{L}^{lsg}$ corresponds to a "constant flow of time"; if we chose another measure ν with $d\nu = \phi d\mathcal{L}^{lsg}$, then, intuitively, the time periods where ϕ is large are the time-periods when ν-relative time is most "concentrated" compared to $\mathcal{L}^{lsg}$-relative time. If $\phi(t) = 60$, then, heuristically speaking, one "second" of $\mathcal{L}^{lsg}$-relative time is equivalent to one "minute" of ν-relative time (and hence, one full minute of ν-relative stochastic noise).

Because of this interpretation ν is sometimes called the **intensity measure** of the noise $\mathcal{N}$.

[1] In theory, the *skewness* of this noise can vary as a function of time. The **skewness function** $\beta(t)$ must satisfy certain conditions. This introduces a lot of technical complications, however, without adding any clear benefit, so we will assume that β is constant.

Proposition 93: *Let $\Phi_{\{f_1,f_2,\ldots,f_N\}}$ be as defined previously.*

1. *$\Phi_{\{f_1,f_2,\ldots,f_N\}}$ is the characteristic function of a stable probability measure on $\mathbb{R}^N$.*

2. *The family of measures $\left\{\mathbf{P}_{\{f_1,f_2,\ldots,f_N\}} \,;\, \{f_1, f_2, \ldots, f_N\} \subset \mathbf{L}^{\langle\alpha\rangle}(X)\right\}$ satisfy the Kolmogorov consistency condition, and therefore define a probability measure on $\mathbb{R}^{\mathbf{L}^{\langle\alpha\rangle}(X)}$.*

3. *If $\mathcal{N} \in \mathbb{R}^{\mathbf{L}^{\langle\alpha\rangle}(X)}$ is a random object with this distribution, then for all $f_1, f_2 \in \mathbf{L}^{\langle\alpha\rangle}(X)$, and $c_1, c_2 \in \mathbb{R}$,*

$$\mathcal{N}(c_1 f_1 + c_2 f_2) \;=\; c_1\mathcal{N}(f_1) + c_2\mathcal{N}(f_2)$$

almost surely; in other words, $\mathcal{N}$ is a linear functional.

Proof: (for more explanation, see [154], pp.115-118.)

Proof of Part 1: Fix $\{f_1, f_2, \ldots, f_N\} \subset \mathbf{L}^{\langle\alpha\rangle}(X)$, and let $F : X \longrightarrow \mathbb{R}^N$ be defined as above. We want to make formula $(*)$ look like the characteristic function of a multivariate stable distribution; to do this, we need to make the right hand side look like

$$\int_{\mathbb{S}^{D-1}} \eta(\vec{\xi}, \mathbf{s})\; d\Gamma[\mathbf{s}]$$

The right hand side of $(*)$ has almost the right integrand; unfortunately, we are integrating over the wrong space (X), with the wrong measure (ν). So, to accomplish our goal, we will need to perform a change of variables.

Define the functions $\Theta : X \longrightarrow \mathbb{S}^{D-1}$ and $R : X \longrightarrow \mathbb{R}$ by

$$\Theta(x) \;=\; \frac{F(x)}{|F(x)|}, \qquad\qquad R(x) \;=\; |F(x)|^\alpha,$$

and let $\Gamma_0 = \Theta^*\nu$ be the projected probability measure on $\mathbb{S}^{D-1}$. Thus, ν has a "disintegration" into "fibre measures":

$$\nu \;=\; \int_{\mathbb{S}^{D-1}} \nu_{\mathbf{s}}\; d\Gamma_0[\mathbf{s}] \qquad\qquad (\dagger)$$

where $\nu_{\mathbf{s}}$ is a measure on the fibre $\Theta^{-1}\{\mathbf{s}\}$.

When $\alpha \neq 1$, $\quad -\Phi_{\{f_1,f_2,\ldots,f_N\}} \;=\; \displaystyle\int_X \eta_{\alpha;\beta}\left(\vec{\xi},\, F(x)\right)\; d\nu[x]$

$$
\begin{aligned}
=_{(1)} \quad & \int_X R(x) \cdot \eta_{\alpha;\beta}\left(\vec{\xi},\, \Theta(x)\right)\; d\nu_{[}x] \\
=_{(2)} \quad & \int_{\mathbb{S}^{D-1}} \int_{\Theta^{-1}\{\mathbf{s}\}} R(x) \cdot \eta_{\alpha;\beta}\left(\vec{\xi},\, \Theta(x)\right)\; d\nu_{\mathbf{s}}[x]\; d\Gamma_0[\mathbf{s}] \\
= \quad & \int_{\mathbb{S}^{D-1}} \left(\int_{\Theta^{-1}\{\mathbf{s}\}} R(x)\; d\nu_{\mathbf{s}}[x] \right) \eta_{\alpha;\beta}\left(\vec{\xi}, \mathbf{s}\right)\; d\Gamma_0[\mathbf{s}] \\
=_{(3)} \quad & \int_{\mathbb{S}^{D-1}} \eta_{\alpha;\beta}\left(\vec{\xi}, \mathbf{s}\right)\; d\Gamma_1[\mathbf{s}] \\
=_{(4)} \quad & \int_{\mathbb{S}^{D-1}} \left[\left(\frac{1+\beta}{2}\right) \eta_{\alpha;(+1)}\left(\vec{\xi}, \mathbf{s}\right) \;+\; \left(\frac{1-\beta}{2}\right) \eta_{\alpha;(-1)}\left(\vec{\xi}, \mathbf{s}\right) \right]\; d\Gamma_1[\mathbf{s}] \\
=_{(5)} \quad & \int_{\mathbb{S}^{D-1}} \left[\left(\frac{1+\beta}{2}\right) \eta_{\alpha;1}\left(\vec{\xi}, \mathbf{s}\right) \;+\; \left(\frac{1-\beta}{2}\right) \eta_{\alpha;1}\left(\vec{\xi}, -\mathbf{s}\right) \right]\; d\Gamma_1[\mathbf{s}] \\
=_{(6)} \quad & \int_{\mathbb{S}^{D-1}} \eta_{\alpha;1}\left(\vec{\xi}, \mathbf{s}\right)\; d\Gamma[\mathbf{s}].
\end{aligned}
$$

- (1) makes use of **Claim 1** (see below)
- (2) employs the "disintegration" of ν (equation (†)).
- (3) Γ_2 is the measure on $\mathbb{S}^{D-1}$ so that, for all $\mathbf{s} \in \mathbb{S}^{D-1}$,

$$\frac{d\Gamma}{d\gamma}[\mathbf{s}] \;=\; \int_{\Theta^{-1}\{\mathbf{s}\}} R(x)\; d\nu_{\mathbf{s}}[x].$$

- (4) It is not difficult to verify that $\eta_{\alpha;\beta} \;=\; \left(\frac{1+\beta}{2}\right) \eta_{\alpha;(+1)} + \left(\frac{1-\beta}{2}\right) \eta_{\alpha;(-1)}$.
- (5) $\eta_{\alpha;(-1)}(\vec{\xi}, \mathbf{s}) = \eta_{\alpha;(+1)}(\vec{\xi}, -\mathbf{s})$.
- (6) Here, Γ is the measure on $\mathbb{S}^{D-1}$ defined: for any $U \subset \mathbb{S}^{D-1}$

$$\Gamma[U] \;=\; \left(\frac{1+\beta}{2}\right) \Gamma_1[U] + \left(\frac{1-\beta}{2}\right) \Gamma_1[-U].$$

The last expression is clearly the formula for the log-characteristic function of a stable random variable, with Γ playing the role of spectral measure. It remains to justify step (1). This involves the following:

Claim 1: *Let $\vec{\xi}, \mathbf{s} \in \mathbb{R}^D$, and let $c \in \mathbb{R}$. Then*

$$\eta_{\alpha;\beta}(\vec{\xi},\, c{\cdot}\mathbf{s}) \;=\; \begin{cases} |c|^\alpha \cdot \eta_{\alpha;\beta}(\vec{\xi},\; \mathbf{sign}(c) \cdot \mathbf{s}) & if \quad \alpha \neq 1 \\ |c| \cdot \eta_{1;\beta}(\vec{\xi},\; \mathbf{sign}(c) \cdot \mathbf{s}) \;+\; \mathcal{B}_1 \beta \cdot c^{\langle 1 \rangle} \left\langle \vec{\xi}, \mathbf{s} \right\rangle \mathbf{i} & if \quad \alpha = 1 \end{cases}.$$

Proof: If $\alpha \neq 1$, then

$$\begin{aligned}
\eta_{\alpha;\beta}(\vec{\xi}, c.\mathbf{s}) &= \left|\left\langle \vec{\xi}, c.\mathbf{s}\right\rangle\right|^{\alpha} + \mathcal{B}_\alpha\beta \left\langle \vec{\xi}, c.\mathbf{s}\right\rangle^{\langle\alpha\rangle} \mathbf{i} \\
&= |c|^{\alpha}\left|\left\langle \vec{\xi}, \mathbf{s}\right\rangle\right|^{\alpha} + |c|^{\alpha}\mathcal{B}_\alpha\beta \left\langle \vec{\xi}, \mathbf{sign}(c).\mathbf{s}\right\rangle^{\langle\alpha\rangle} \mathbf{i} \\
&= |c|^{\alpha}\eta_{\alpha;\beta}(\vec{\xi}, \mathbf{sign}(c)\cdot\mathbf{s}).
\end{aligned}$$

If $\alpha = 1$, then

$$\begin{aligned}
\eta_{1;\beta}(\vec{\xi}, c.\mathbf{s}) &= \left|\left\langle \vec{\xi}, c.\mathbf{s}\right\rangle\right| + \mathcal{B}_1\beta \left\langle \vec{\xi}, c.\mathbf{s}\right\rangle \log\left|\left\langle \vec{\xi}, c.\mathbf{s}\right\rangle\right| \mathbf{i} \\
&= |c|\left|\left\langle \vec{\xi}, \mathbf{s}\right\rangle\right| \; + \; |c|\cdot\mathcal{B}_1\beta \left\langle \vec{\xi}, \mathbf{sign}(c)\cdot\mathbf{s}\right\rangle \left(\log\left|\left\langle \vec{\xi}, \mathbf{s}\right\rangle\right| + \log|c|\right) \mathbf{i} \\
&= |c|\cdot\eta_{1;\beta}(\vec{\xi}, \mathbf{sign}(c)\cdot\mathbf{s}) \; + \; \mathcal{B}_1\beta\cdot c\log|c| \left\langle \vec{\xi}, \mathbf{s}\right\rangle \mathbf{i} \\
&= |c|\cdot\eta_{1;\beta}(\vec{\xi}, \mathbf{sign}(c)\cdot\mathbf{s}) \; + \; \mathcal{B}_1\beta c^{\langle 1\rangle} \left\langle \vec{\xi}, \mathbf{s}\right\rangle \mathbf{i}.
\end{aligned}$$

..... □ [Claim 1]

The argument in the case when $\alpha = 1$ is similar, only now we end up with a dangling constant.

Proof of Part 2: The Kolmogorov Consistency Criterion,

$$\text{“}\mathbf{P}_{\{f_0,f_1,\ldots,f_N\}}(\mathbb{R}, U_1, \ldots, U_N) \;=\; \mathbf{P}_{\{f_1,\ldots,f_N\}}(U_1, \ldots, U_N)$$

for all $f_0, f_1, \ldots, f_N \in \mathbf{L}^{\langle\alpha\rangle}(X)$ and $U_1, \ldots, U_N \subset \mathbb{R}$,” is equivalent to the following condition on the characteristic functions: “For all $f_0, f_1, \ldots, f_N \in \mathbf{L}^{\langle\alpha\rangle}(X)$, and all $\xi_1, \ldots, \xi_N \in \mathbb{R}$,

$$\chi_{\{f_0,f_1,\ldots,f_N\}}(0, \xi_1, \ldots, \xi_N) \;=\; \chi_{\{f_1,\ldots,f_N\}}(\xi_1, \ldots, \xi_N),\text{”}$$

which is equivalent to the condition:

$$\Phi_{\{f_0,f_1,\ldots,f_N\}}(0, \xi_1, \ldots, \xi_N) \;=\; \Phi_{\{f_1,\ldots,f_N\}}(\xi_1, \ldots, \xi_N)$$

This condition is clearly satisfied by the functions $\Phi_{f_1,f_2,\ldots,f_N}$ as we have defined them.

Proof of Part 3: Let $f_0 = c_1 f_1 + c_2 f_2$, and consider the marginal log-Fourier transform $\Phi_{\{f_0,f_1,f_2\}}$. The statement

$$\text{“}\mathcal{N}(c_1 f_1 + c_2 f_2) \;=\; c_1\mathcal{N}(f_1) + c_2\mathcal{N}(f_2)\text{”}$$

is equivalent to: "For any $\vec{\xi} \in \mathbb{R}^3$, if $\vec{\xi}$ is normal to the subspace $V = \left\{\vec{\xi} \in \mathbb{R}^3 \ ; \ \xi_0 = c_1\xi_1 + x_2\xi_2\right\}$, then $\chi_{\{f_0,f_1,f_2\}}(\vec{\xi}) = 1$."

The orthogonal complement of V is one-dimensional, and generated by the vector $\vec{\xi} = (-1, c_1, c_2)$, so it suffices to show that $\chi_{\{f_0,f_1,f_2\}}(\vec{\xi}) = 1$, or equivalently, that $\Phi_{\{f_0,f_1,f_2\}}(\vec{\xi}) = 0$. But

$$\Phi_{\{f_0,f_1,f_2\}}(\vec{\xi}) = -\int_X \left[\left|\left\langle \vec{\xi}, F(x)\right\rangle\right|^\alpha + \mathcal{B}_\alpha\beta\left\langle \vec{\xi}, F(x)\right\rangle^{\langle\alpha\rangle}\mathbf{i}\right] d\nu[x],$$

where $F(x) = (f_0(x), f_1(x), f_2(x))$. By construction of F and $\vec{\xi}$, $\left\langle \vec{\xi}, F(x)\right\rangle = 0$ for any $x \in X$. □

To emphasis the heuristic that $\mathcal{N}$ is a "random measure", we will normally use the notation:

$$\int_X f \ d\mathcal{N}$$

to denote the value $\mathcal{N}(f)$.

Example 94: *The Lévy process*

Let $\mathcal{L}^{lbsg}$ be the Lebesgue measure on $[0, \infty)$, and let $\mathcal{N}$ be an α-stable random noise on the measure space $([0, \infty), \mathcal{L}^{lbsg})$, with some skewness $\beta \in [-1, +1]$.

The **α-stable Lévy process** with **skewness** β is the stochastic process $\mathcal{X} : [0, \infty) \longrightarrow \mathbb{R}$ defined:

$$\mathcal{X}(T) = \int_{\mathbb{R}} \mathbb{1}_{0,T} \ d\mathcal{N}$$

When $\alpha = 2$, the parameter β is irrelevant, and we have the Weiner process, also called Brownian Motion.

Proposition 95: *Properties of the Lévy Process*

- $\mathcal{X}(0) = 0$ *almost surely.*
- *For any* $T > 0$, $\mathcal{X}(T)$ *is an* α*-stable random variable with distribution* $\mathcal{S}^\alpha[\sigma_T; \beta_T; \mu_T]$ *where:*

$$\begin{aligned} \sigma_T &= T^{1/\alpha}, \\ \beta_T &= \beta, \\ \text{and } \mu_T &= \begin{cases} 0 & if \quad \alpha \neq 1 \\ \mathcal{B}_1\beta \cdot T & if \quad \alpha = 1 \end{cases}. \end{aligned}$$

- $\mathcal{X}$ *has* **stationary increments**.

Proof: **Part 1** and **Part 3** are clear from the definition. **Part 2** follows from Corollary 102 on page 191 (see below). □

Example 96: *Lévy process with variable-speed time*

Let ν be a nonnegative measure on $[0, \infty)$, and let $\mathcal{N}$ be an α-stable random noise on the measure space $([0, \infty), \nu)$, with some skewness $\beta \in [-1, +1]$.

The **α-stable Lévy process** with **skewness** β and **time measure** ν is the stochastic process $\mathcal{X} : [0, \infty) \longrightarrow \mathbb{R}$ defined:

$$\mathcal{X}(T) \;=\; \int_{\mathbb{R}} \mathbb{1}_{0,T} \; d\mathcal{N}$$

Proposition 97: *Properties of the variable-time Lévy Process*

- $\mathcal{X}(0) = 0$ *almost surely.*
- *For any* $T > 0$, $\mathcal{X}(T)$ *is an* α*-stable random variable with distribution* $\bigcap^{\alpha} [\sigma_T; \beta_T; \mu_T]$ *where:*

$$\begin{aligned} \sigma_T &= \nu[0,T]^{1/\alpha}, \\ \beta_T &= \beta, \\ \text{and } \mu_T &= \begin{cases} 0 & if \quad \alpha \neq 1 \\ \mathcal{B}_1\beta \cdot \nu[0,T] & if \quad \alpha \neq 1 \end{cases}. \end{aligned}$$

- $\mathcal{X}$ *does* not *necessarily have stationary increments.*

Proof: **Part 1** and **Part 3** are clear from the definition. **Part 2** follows from Corollary 102 on page 191 (see below). □

Example 98: *Spectral Measures*

Let $\Gamma \in \mathcal{M}_{\mathcal{EAS}}\left[\mathbb{S}^{D-1}\right]$, and let $\mathcal{N}$ be a α-stable random noise on the measure space $(\mathbb{S}^{D-1}, \Gamma)$ with skewness $\beta = 1$. For all $d \in [1..D]$, let $\mathbf{pr}_d : \mathbb{R}^D \longrightarrow \mathbb{R}$ be projection onto the dth coordinate. Define the random vector

$$\mathbf{X} \;=\; \left(\int_{\mathbb{S}^{D-1}} \mathbf{pr}_1 \; d\mathcal{N},\; \int_{\mathbb{S}^{D-1}} \mathbf{pr}_2 \; d\mathcal{N},\; \ldots,\; \int_{\mathbb{S}^{D-1}} \mathbf{pr}_D \; d\mathcal{N} \right)$$

Then $\mathbf{X}$ is α-stable, and has spectral measure Γ.

Proof: See [154], Theorem 3.5.6, p. 131 □

Stable noise allows us to easily define one-dimensional stochastic processes. But suppose we want to define a stochastic process that ranges over a D-dimensional statespace. It seems we will need "D-dimensional noise" —in other words, a "random $\mathbb{R}^D$-valued measure."

Definition 99: *D-dimensional α-stable noise*

Let $\alpha \in [0, 2]$, and let $\Gamma \in \mathcal{M}_{\mathcal{EAS}}\left[\mathbb{S}^{D-1}\right]$. Let $(X, \mathcal{X}, \nu)$ be an arbitrary measure space. The **D-dimensional α-stable random noise** *on $(X, \mathcal{X}, \nu)$ with* **spectral measure** *Γ is defined as follows:*

- *Let $\mathcal{M}$ be an α-stable random noise on the measure space $\left(X \times \mathbb{S}^{D-1},\, \nu \otimes \Gamma\right)$, with skewness $\beta = 1$.*
- *For any $f \in \mathbf{L}^{\langle\alpha\rangle}(X, \nu)$, define $f_1, f_2, \ldots, f_D \in \mathbf{L}^{\langle\alpha\rangle}\left(X \times \mathbb{S}^{D-1},\, \nu \otimes \Gamma\right)$ by:*

$$f_d(x, \mathbf{s}) \;=\; f(x) \cdot \mathbf{pr}_d(\mathbf{s}) \;=\; f(x) \cdot s_d$$

Now, define the random measure $\mathcal{N}$ on X as follows: for any $f \in \mathbf{L}^{\langle\alpha\rangle}(X, \nu)$ define:

$$\mathcal{N}(f) \;=\; (\mathcal{M}(f_1),\; \mathcal{M}(f_2), \ldots,\; \mathcal{M}(f_D))$$

Proposition 100: *For any $f \in \mathbf{L}^{\langle\alpha\rangle}(X,\nu)$, the random vector $\mathcal{N}(f) \in \mathbb{R}^D$ is α-stable, with spectral measure*

$$C_+ \cdot \Gamma \; + \; C_- \cdot \Gamma_-,$$

where Γ_- is the "spherical inversion" of Γ, defined

$$\forall\, U \subset \mathbb{S}^{D-1}, \quad \Gamma_-[U] = \Gamma[-U]$$

and where C_+ and C_- are the real constants:

$$C_+ \;=\; \int_X f(x)_+^\alpha \; d\nu[x], \qquad C_- \;=\; \int_X f(x)_-^\alpha \; d\nu[x],$$

where, for any $r \in \mathbb{R}$, $r_+ = \begin{cases} r & if \quad r > 0 \\ 0 & if \quad r \leq 0 \end{cases}$, and $r_- = \begin{cases} -r & if \quad r < 0 \\ 0 & if \quad r \geq 0 \end{cases}$.

- *If $\alpha \neq 1$, the shift parameter of $\mathcal{N}(f)$ is equal to zero.*
- *If $\alpha = 1$, the shift parameter of $\mathcal{N}(f)$ is $\vec{\mu}_f \; = \mathcal{B}_\alpha \langle f \rangle^{\langle 1 \rangle} \vec{\gamma}$,*

$$\text{where } \vec{\gamma} \;=\; \int_{\mathbb{S}^{D-1}} \mathbf{s} \; d\Gamma[\mathbf{s}] \;\; \text{is the "centroid" of } \Gamma,$$

$$\text{and where } \; \langle f \rangle_{\langle 1 \rangle} \;=\; \int_X f(x) \log |f(x)| \; d\nu[x].$$

Proof: Let $\mathbf{X} = \mathcal{N}(f) = (\mathcal{M}(f_1),\; \mathcal{M}(f_2), \ldots,\; \mathcal{M}(f_D))$, where $\mathcal{M}$ and $f_1, \ldots, f_D$ are defined as previously. Then the log-characteristic function of $\mathbf{X}$ is just the joint log-characteristic function of $(\mathcal{M}(f_1),\; \mathcal{M}(f_2), \ldots,\; \mathcal{M}(f_D))$, that is, the log-Fourier transform of the marginal probability distribution $\mathbf{P}_{\{f_1,\ldots,f_D\}}$, which, according to the definition of the noise $\mathcal{M}$ is the function

$$\Phi_{\{f_1,f_2,\ldots,f_D\}}(\vec{\xi}) \;=\; -\int_{X \times \mathbb{S}^{D-1}} \eta_{\alpha;1}\left(\vec{\xi}, F(x,\mathbf{s})\right) \; d(\nu \otimes \Gamma)[x,\mathbf{s}]$$

where $F : X \times \mathbb{S}^{D-1} \longrightarrow \mathbb{R}^D$ is defined: $F(x,\mathbf{s}) \;=\; f(x) \cdot \mathbf{s}$, and where $\eta_{\alpha;1}(\vec{\xi},\mathbf{s}) \;=\; \left|\left\langle \vec{\xi}, \mathbf{s} \right\rangle\right|^\alpha + \mathcal{B}_\alpha \left\langle \vec{\xi}, \mathbf{s} \right\rangle^{\langle\alpha\rangle} \mathbf{i}$. Thus, if $\alpha \neq 1$, then

$$-\Phi_{\{f_1,f_2,\ldots,f_D\}}(\vec{\xi})$$
$$= \quad \int_{X \times \mathbb{S}^{D-1}} \eta_{\alpha;1}\left(\vec{\xi},\; f(x) \cdot \mathbf{s}\right) \; d(\nu \otimes \Gamma)[x,\mathbf{s}]$$

$$
\begin{aligned}
=_{(1)} \quad & \int_{\mathbb{S}^{D-1}} \int_X |f(x)|^\alpha \eta_{\alpha;1}\left(\vec{\xi}, \mathbf{sign} f(x) \cdot \mathbf{s}\right)\ d\nu[x]\ d\Gamma[\mathbf{s}] \\
=_{(2)} \quad & \int_{\mathbb{S}^{D-1}} \int_{X_+} |f(x)|^\alpha \eta_{\alpha;1}\left(\vec{\xi}, \mathbf{s}\right)\ d\nu[x]\ d\Gamma[\mathbf{s}] \\
& \qquad + \int_{\mathbb{S}^{D-1}} \int_{X_-} |f(x)|^\alpha \eta_{\alpha;1}\left(\vec{\xi}, -\mathbf{s}\right)\ d\nu[x]\ d\Gamma[\mathbf{s}] \\
= \quad & \int_{\mathbb{S}^{D-1}} \left(\int_{X_+} |f(x)|^\alpha\ d\nu[x]\right) \eta_{\alpha;1}\left(\vec{\xi}, \mathbf{s}\right)\ d\Gamma[\mathbf{s}] \\
& \qquad + \int_{\mathbb{S}^{D-1}} \left(\int_{X_-} |f(x)|^\alpha\ d\nu[x]\right) \eta_{\alpha;1}\left(\vec{\xi}, \mathbf{s}\right)\ d\Gamma_-[\mathbf{s}] \\
= \quad & \int_{\mathbb{S}^{D-1}} C_+ \cdot \eta_{\alpha;1}\left(\vec{\xi}, \mathbf{s}\right)\ d\Gamma[\mathbf{s}] \quad + \quad \int_{\mathbb{S}^{D-1}} C_- \cdot \eta_{\alpha;1}\left(\vec{\xi}, \mathbf{s}\right)\ d\Gamma_-[\mathbf{s}] \\
= \quad & \int_{\mathbb{S}^{D-1}} \eta_{\alpha;1}\left(\vec{\xi}, \mathbf{s}\right)\ \ d(C_+\Gamma + C_-\Gamma_-)[\mathbf{s}]
\end{aligned}
$$

- (1) follows from **Claim 1** of Proposition 93 on page 183.
- (2) Here, $X_+ = \{x \in X\ ;\ f(x) > 0\}$ and $X_- = \{x \in X\ ;\ f(x) < 0\}$.

Similarly, when $\alpha = 1$, apply **Claim 1** of Proposition 93 on page 183 to conclude:

$$
\begin{aligned}
-\Phi_{\{f_1, f_2, \ldots, f_D\}}(\vec{\xi}) \quad &= \quad \int_{X \times \mathbb{S}^{D-1}} \eta_{1;1}\left(\vec{\xi},\ f(x) \cdot \mathbf{s}\right)\ d(\nu \otimes \Gamma)[x, \mathbf{s}] \\
&= \quad \int_{\mathbb{S}^{D-1}} \eta_{1;1}\left(\vec{\xi}, \mathbf{s}\right)\ d(C_+\Gamma + C_-\Gamma_-)[\mathbf{s}] \\
&\qquad + \quad \int_{X \times \mathbb{S}^{D-1}} \mathcal{B}_1 f(x)^{\langle 1 \rangle} \left\langle \vec{\xi}, \mathbf{s} \right\rangle \mathbf{i}\ d(\nu \otimes \Gamma)[x, \mathbf{s}].
\end{aligned}
$$

$$
\begin{aligned}
\text{But} \quad & \int_{X \times \mathbb{S}^{D-1}} \mathcal{B}_1 f(x)^{\langle 1 \rangle} \left\langle \vec{\xi}, \mathbf{s} \right\rangle \mathbf{i}\ d(\nu \otimes \Gamma)[x, \mathbf{s}] \\
&= \quad \int_{\mathbb{S}^{D-1}} \mathcal{B}_1 \left(\int_X f(x)^{\langle 1 \rangle}\ d\nu[x]\right) \left\langle \vec{\xi}, \mathbf{s} \right\rangle \mathbf{i}\ d\Gamma[\mathbf{s}] \\
&= \quad \int_{\mathbb{S}^{D-1}} \mathcal{B}_1 \left\langle f \right\rangle_{\langle 1 \rangle} \mathbf{i} \cdot \left\langle \vec{\xi}, \mathbf{s} \right\rangle\ d\Gamma[\mathbf{s}] \\
&= \quad \mathcal{B}_1 \left\langle f \right\rangle_{\langle 1 \rangle} \mathbf{i} \cdot \left\langle \vec{\xi},\ \int_{\mathbb{S}^{D-1}} \mathbf{s}\ d\Gamma[\mathbf{s}] \right\rangle \\
&= \quad \mathcal{B}_1 \left\langle f \right\rangle_{\langle 1 \rangle} \mathbf{i} \cdot \left\langle \vec{\xi},\ \vec{\gamma} \right\rangle \\
&= \quad \left\langle \vec{\xi},\ \mathcal{B}_1 \left\langle f \right\rangle_{\langle 1 \rangle} \vec{\gamma} \right\rangle \mathbf{i}
\end{aligned}
$$

$$= \left\langle \vec{\xi},\ \vec{\mu} \right\rangle \mathbf{i}$$

Hence, in both cases, the spectral measure is $C_+\Gamma + C_-\Gamma_-$. When $\alpha \neq 1$, the shift parameter is equal to zero. But when $\alpha = 1$, the shift parameter is $\vec{\mu}$.

□

Corollary 101: *Let Γ_f be the spectral measure of the α-stable random vector $\int_X f\ d\mathcal{N}$ of Proposition 100 on page 189. Then*

$$\Gamma_f\left[\mathbb{S}^{D-1}\right] = \|f\|_\alpha^\alpha\, \Gamma\left[\mathbb{S}^{D-1}\right]$$

where the "norm" $\|\bullet\|_\alpha$ is defined[2] $\|f\|_\alpha = \left(\int_X |f(x)|^\alpha\ d\nu[x]\right)^{1/\alpha}$.

Proof: Let C_+ and C_- be defined as in Proposition 100 on page 189. Then

$$\begin{aligned}\Gamma_f\left[\mathbb{S}^{D-1}\right] &= C_+ \cdot \Gamma\left[\mathbb{S}^{D-1}\right] + C_- \cdot \Gamma_-\left[\mathbb{S}^{D-1}\right] \\ &= (C_+ + C_-) \cdot \Gamma\left[\mathbb{S}^{D-1}\right],\end{aligned}$$

and $C_+ + C_- = \int_X f(x)_+^\alpha + f(x)_-^\alpha\ d\nu[x] = \int_X |f(x)|^\alpha\ d\nu[x]$. □

Corollary 102: *Let $(X, \mathcal{X}, \nu)$ be a measure space, and let $f \in \mathbf{L}^{\langle\alpha\rangle}(X)$. Let $\mathcal{N}$ be a one-dimensonal, α-stable noise on (X, ν), with skewness β. Then $\int_X f\ d\mathcal{N}$ is an α-stable random variable having distribution $\mathcal{S}^\alpha[\mu_f; \beta_f; \sigma_f]$, where*

$$\begin{aligned}\text{if } \alpha \neq 1, \qquad \mu_f &= 0, \\ \sigma_f &= \|f\|_\alpha, \\ \text{and } \beta_f &= \frac{\langle f\rangle^{\langle\alpha\rangle}}{\sigma_f^\alpha}\beta,\end{aligned}$$

$$\begin{aligned}\text{while if } \alpha = 1, \qquad \mu_f &= \mathcal{B}_1 \beta \langle f\rangle_{\langle 1\rangle} \\ \sigma_f &= \|f\|_\alpha, \\ \text{and } \beta_f &= \frac{\beta}{\sigma_f}\int_X f\ d\nu\end{aligned}$$

[2] If $\alpha \geq 1$, this is just the $\mathbf{L}^\alpha$ norm. If $\alpha < 1$, then it is not actually a norm.

where, as before, $\|f\|_\alpha = \left(\int_X |f(x)|^\alpha \, d\nu[x]\right)^{1/\alpha}$, *and* $\langle f \rangle^{\langle\alpha\rangle} = \int_X f(x)^{\langle\alpha\rangle} \, d\nu[x]$.

Proof: Recall that the "spectral measure" for the one-dimensional distribution $\mathcal{S}\alpha[\mu;\beta;\sigma]$, is the measure $\Gamma = \gamma_{+1}\delta_{+1} + \gamma_{-1}\delta_{-1}$, where γ_{+1} and γ_{-1} are the unique positive values so that

$$\begin{aligned} \sigma &= (\gamma_{+1} + \gamma_{-1})^{1/\alpha} \\ \text{and } \beta &= \frac{\gamma_{+1} - \gamma_{-1}}{\sigma^\alpha} \end{aligned}$$

Now apply Proposition 100 on page 189. □

D.3 Examples

Most α-stable stochastic processes can be represented as a **stochastic integral**:

$$\mathcal{X}(T) = \int_{-\infty}^{\infty} \phi(T,t) \, d\mathcal{W}_\alpha(t)$$

where $\mathcal{W}_\alpha(t)$ is an **α-stable noise** and $\phi(T,t)$ is some kernel. (For an excellent introduction to stochastic integration and stochastic noise, see [154], chapter 7.)

Brownian Motion:

$$\mathcal{X}(T) = \int_0^T d\mathcal{W}(t)$$

where $\mathcal{W}$ is "Gaussian white noise". Hence, $\phi(T,t) = \mathbb{1}_{[0,T]}(t)$.

Ornstein-Uhlenbeck Process:

$$\mathcal{X}(T) = \int_{-\infty}^{T} \exp(-\lambda(T-t)) \, d\mathcal{W}(t)$$

Fractional Brownian Motion: For any $H \in [0,1], H \neq 1/2$, define

$$\mathcal{X}(T) \;=\; \int_{-\infty}^{\infty} \left(|T-t|^{H-1/2} - |t|^{H-1/2} \right) \, d\mathcal{W}(t)$$

This gives us ***H*-self-affine fractional Brownian motion**.

Now, let $\alpha \in [0,2]$. In contrast to the Gaussian case, there are many different α-stable noises, determined by different degrees of *skewness*. For a *one-dimensional* process, we must choose a value for the **skewness parameter**, β. For a *D-dimensional* process, we must choose a **spectral measure**[3], Γ. For a given choice of skewness β (or spectral measure Γ) we will write the corresponding noise as "$\mathcal{W}_\alpha^\beta$" (or "$\mathcal{W}_\alpha^\Gamma$"). The following examples are defined in [154] as one-dimensional processes, but clearly, the same dimension will work in higher dimensions: fix a spectral measure Γ, and integrate the kernels with respect to the noise $\mathcal{W}_\alpha^\Gamma$. In the following discussion, we will suppress the "β" (or "Γ").

Lévy -stable Motion:

$$\mathcal{X}(T) \;=\; \int_0^T \, d\mathcal{W}_\alpha[t]$$

where $\mathcal{W}_\alpha$ is an α-stable white noise.

Moving Averages: Let $\phi : \mathbb{R} \longrightarrow \mathbb{R}$. The **moving average** process generated by ϕ is defined:

$$\mathcal{X}(T) \;=\; \int_{-\infty}^{\infty} \phi(T-t) \, d\mathcal{W}_\alpha[t]$$

For example, the Ornstein-Uhlenbeck process is an example of such a process.

"Balanced" Fractional Stable Motion: For any $H \in [0,1]$, with $H \neq \frac{1}{\alpha}$, define

$$\mathcal{X}(T) \;=\; \int_{-\infty}^{\infty} \left(|T-t|^{H-1/\alpha} - |t|^{H-1/\alpha} \right) \, d\mathcal{W}[t]$$

This gives us ***H*-self-affine balanced fractional stable motion**.

[3]**Note:** Do not confuse the spectral measure of the stable noise $\mathcal{W}_\alpha^\Gamma$ with the *entirely different* "spectral measure" associated with the autocorrelation of a Gaussian process. This is but one of many reasons why the terminology "spectral measure" is unfortunate in the realm of stable random variables.

"Unbalanced" Fractional Stable Motion: More generally, again let $H \in [0,1]$, with $H \neq \frac{1}{\alpha}$, but now, for any $a, b \in \mathbb{R}$, with at least one of a or b nonzero. Let

$$\phi(T,t) \;=\; a\left((T-t)_+^{H-1/\alpha} - (t)_+^{H-1/\alpha}\right) \;+\; b\left((T-t)_-^{H-1/\alpha} - (t)_-^{H-1/\alpha}\right)$$

where

$$u_+ \;=\; \begin{cases} u & if \quad u \geq 0 \\ 0 & if \quad u < 0 \end{cases} \qquad \text{and} \qquad u_- \;=\; \begin{cases} -u & if \quad u \leq 0 \\ 0 & if \quad u > 0 \end{cases}$$

Then the process

$$\mathcal{X}(T) \;=\; \int_{-\infty}^{\infty} \phi(T,t)\; d\mathcal{W}(t)$$

is **H-self-affine linear fractional α-stable motion** with "coefficients" (a,b). (see [154], 7.4, p. 343). The coefficients of the motion completely characterises it, but different coefficients a and b can produce identical processes.

Proposition 103: *Let $\alpha \in [0,2)$, and suppose that $a_1, a_2, b_1, b_2 \in \mathbb{R}$, and that at least one element of each pair (a_1, b_1) and (a_2, b_2) is nonzero.*

For $k = 1, 2$, let $\mathcal{X}_k(t)$ be an H-self-affine linear fractional α-stable motions with coefficients (a_k, b_k). Then $\mathcal{X}_1(t)$ and $\mathcal{X}_2(t)$ are stochastically identical up to a multiplicative constant if and only if one of the following conditions obtains:

$$a_1 = 0 = a_2,$$

$$a_1 = 0 = a_2,$$

$$or \quad \frac{a_1}{b_1} \;=\; \frac{a_2}{b_2}$$

In this case, $\mathcal{X}_2(t)$ is a version of $C.\mathcal{X}_1(t)$, where $C = \frac{a_1}{a_2}$ or $C = \frac{b_1}{b_2}$ (or both), depending upon which is well-defined.

Proof: If $D = 1$, see [154], Theorem 7.4.5, p. 347, for a proof of the one-dimensional version of this theorem. If $D > 1$, then any one-dimensional projection of the process will be a one-dimensional fractional stable motion; apply the one-dimensional case to extract the unique type r for this one-dimensional process. □

According to this theorem, the important property of the coefficients a and b is the *ratio* between them, where we generalize the definition of ratio by defining:

$$\frac{a}{b} \;=\; \begin{cases} 0 & if \quad a = 0 \neq b \\ a/b & if \quad a \neq 0 \neq b \\ \infty & if \quad a \neq 0 = b \end{cases}$$

(and at least one of a or b must be nonzero).

Because of this, if the motion has coefficients a and b, we will say that it is of **type** $r = \frac{a}{b}$, using this generalized understanding of ratio. We will use the symbol $\mathcal{LFM}_r^{H;\alpha}$ to represent the **H-self-affine linear fractional α-stable motion of type** r, where we assume that the coefficients a and b are normalized so that $|a| + |b| = 1$.

Remark 104:

- If $r = \infty$ –ie. $a > 0$, $b = 0$, then we have **nonanticipating** fractional stable motion; the value of $\mathcal{X}(T)$ depends only on the noise "before" time T.
- If $r = 1$ –ie. $a = b$, then we have **well-balanced** fractional stable motion; the value of $\mathcal{X}(T)$ depends only on the noise "before" time T.
- If $\alpha = 2$, then, in contrast to the previous theorem, *all* choices of coefficients a and b produce the same process; there is only one "fractional linear Brownian motion".
- A fractional linear motion can be thought of, heuristically, as a sum:

$$\mathcal{X}(T) \;=\; \mathcal{X}_1(T) \;-\; \mathbf{X}$$

where $\mathcal{X}_1$ is the "moving average" process:

$$\mathcal{X}_1(T) \;=\; \int_{-\infty}^{\infty} \left(a\,(T-t)_+^{H-1/\alpha} \;+\; b\,(T-t)_-^{H-1/\alpha} \right)\, d\mathcal{W}_\alpha,$$

while $\mathbf{X}$ is the random vector:

$$\mathbf{X} \;=\; \int_{-\infty}^{\infty} \left(a\,(t)_+^{H-1/\alpha} + b\,(t)_-^{H-1/\alpha} \right) d\mathcal{W}_\alpha,$$

However, these integrals do not converge, so this decomposition makes no sense, formally.

Logarithmic Fractional α-stable motion: The definition of fractional linear motion only really makes sense when $H \neq \frac{1}{\alpha}$. What is the natural extension of the definition to the case when $H = \frac{1}{\alpha}$?

If $a = 1$ and $b = 0$ (ie. the case of **nonanticipating** motion), then as $H \to \frac{1}{\alpha}$, the kernel

$$\phi_H(T,t) \;=\; \left((T-t)_+^{H-1/\alpha} - (t)_+^{H-1/\alpha}\right)$$

converges, pointwise, to the kernel

$$\phi_{1/\alpha}(T,t) \;=\; \mathbb{1}_{[0,T]}(t)$$

Hence, the **α-stable Lévy process** seems the natural limit of nonanticipating fractional motion as $H \to \frac{1}{\alpha}$.

However, in the **well-balanced** case (when $a = 1 = b$), this argument no longer works. Instead, as $H \to \frac{1}{\alpha}$, the kernel

$$\phi_H(T,t) \;=\; |T-t|^{H-1/\alpha} - |t|^{H-1/\alpha}$$

converges pointwise to zero.

However, suppose that $1 < \alpha < 2$, and suppose we "renormalize" $\phi_H(T,t)$, multiplying by the value of $c_H = \frac{1}{\phi_H(1,0)}$. Then visually, the kernel "seems" to converge to the function

$$\psi(T,t) \;=\; \log|T-t| - \log|t|$$

(see, for example, Figure 7.10, p. 354 of [154]).

This motivates the definition of **Log-fractional stable motion**, which is the process:

$$\mathcal{X}(T) \;=\; \int_{-\infty}^{\infty} (\log|T-t| - \log|t|)\; d\mathcal{W}_\alpha.$$

When $\alpha \in (1,2)$, the process $\mathcal{X}$ is α-stable, and $\frac{1}{\alpha}$-self affine.

D.4 Simulation

Given a stochastic integral representation

$$\mathcal{X}(T) \;=\; \int_{-\infty}^{\infty} \phi(T,t)\; d\mathcal{W}_\alpha^\Gamma[t]$$

of a stochastic process, it is straightforward to simulate the process on a computer. First, we must choose some "mesh size", $\delta > 0$, which determines the resolution of our discrete representation of $\mathbb{R}$. To generate an "instance" of the process $\mathcal{X}(T)$, we must first simulate the α-stable noise $\mathcal{W}_\alpha^\Gamma$. For every $n \in \mathbb{Z}$, let $\mathbf{W}_n \in \mathbb{R}^D$ be an α-stable random vector, with spectral measure Γ, so that $\{\mathbf{W}_n \; ; \; n \in \mathbb{Z}\}$ are independent (see Section B.5 on page 151 for the simulation of α-stable variables). Then, for all $T \in \delta.\mathbb{Z}$, define:

$$\mathbf{X}(T) \;=\; \sum_{n=-\infty}^{\infty} \phi(T, \delta.n) \; \mathbf{W}_n$$

Of course, in reality, we cannot compute an infinite sum on a computer, so we instead compute:

$$\mathsf{X}(T) \;=\; \sum_{n=-N}^{N} \phi(T, \delta.n) \; \mathbf{W}_n$$

where N is "large", and $\{\mathbf{W}_n \; ; \; -N \leq n \leq N\}$ are independent α-stable variables Assume that $|\phi(t)|$ decays rapidly as $t \to \pm\infty$. Then, as long as T is far from $\pm\delta N$, this approximation should be fairly accurate.

Appendix E

Tensor Algebra

[An excellent introduction to tensor algebra and its applications to Riemannian geometry, is [147]]

E.1 Tensor Products

If $\mathbb{V}$ and $\mathbb{W}$ are two vector spaces over a field $\mathbb{F}$, then the **tensor product** of $\mathbb{V}$ and $\mathbb{W}$ is defined:

$$\mathbb{V} \otimes \mathbb{W} \;=\; \{f : \mathbb{V}^* \times \mathbb{W}^* \longrightarrow \mathbb{F} \;;\; f \text{ a bilinear map}\}$$

Inductively, if $\mathbb{V}_1, \ldots, \mathbb{V}_N$ are vector spaces, then

$$\mathbb{V}_1 \otimes \mathbb{V}_2 \otimes \ldots \otimes \mathbb{V}_N \;=\; \{f : \mathbb{V}_1^* \times \ldots \times \mathbb{V}_N^* \longrightarrow \mathbb{F} \;;\; f \text{ an } N\text{-linear map}\}$$

If $\mathbf{v}_n \in \mathbb{V}_n$, for all $n \in [1..N]$, then the element $\mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \ldots \otimes \mathbf{v}_n \in \mathbb{V}_1 \otimes \mathbb{V}_2 \otimes \ldots \otimes \mathbb{V}_N$ is defined to be the N-linear map:

$$\begin{array}{rccl} \mathbf{v}_1 \otimes \ldots \otimes \mathbf{v}_n : \mathbb{V}_1^* \times \ldots \times \mathbb{V}_N^* & \longrightarrow & \mathbb{F} \\ (\mathbf{w}_1^*, \ldots, \mathbf{w}_N^*) & \mapsto & \mathbf{w}_1^*(\mathbf{v}_1) \cdot \ldots \cdot \mathbf{w}_N^*(\mathbf{v}_N) \end{array}$$

The operator "$\otimes$" itself is bilinear: if $\mathbf{v}_1, \mathbf{w}_1 \in \mathbb{V}_1$ and $\mathbf{v}_2, \mathbf{w}_2 \in \mathbb{V}_2$, and $c_1, c_2 \in \mathbb{C}$, then

$$(\mathbf{v}_1 + c_1.\mathbf{w}_1) \otimes (\mathbf{v}_2 + c_2.\mathbf{w}_2) \;=\; (\mathbf{v}_1 \otimes \mathbf{v}_2) + c_1.(\mathbf{w}_1 \otimes \mathbf{v}_2) + c_2.(\mathbf{v}_1 \otimes \mathbf{w}_2) + c_1.c_2.(\mathbf{w}_1 \otimes \mathbf{w}_2)$$

If $\mathbb{V}_1, \ldots, \mathbb{V}_N$ are **inner product spaces** (or, equivalently, if we have selected a particular basis of each spaces and defined that basis to be "orthonormal"), then there is a canonical isomorphism $\mathbb{V} \cong \mathbb{V}^*$, and therefore,

a canonical isomorphism;

$$\mathbb{V}_1 \otimes \mathbb{V}_2 \otimes \ldots \otimes \mathbb{V}_N \;\cong\; \{f : \mathbb{V}_1 \times \ldots \times \mathbb{V}_N \longrightarrow \mathbb{F}\,;\; f \text{ an } N\text{-linear map}\}$$

In this case, the element $\mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \ldots \otimes \mathbf{v}_n$ is defined to be the N-linear map:

$$\begin{array}{rcl} \mathbf{v}_1 \otimes \ldots \otimes \mathbf{v}_n : \mathbb{V}_1 \times \ldots \times \mathbb{V}_N & \longrightarrow & \mathbb{F} \\ (\mathbf{w}_1, \ldots, \mathbf{w}_N) & \mapsto & \langle \mathbf{w}_1, \mathbf{v}_1 \rangle \cdot \ldots \cdot \langle \mathbf{w}_N, \mathbf{v}_N \rangle \end{array}$$

Let's look specifically at $\mathbb{V}^{\otimes N} := \bigotimes_{n=1}^{N} \mathbb{V}$.

Suppose that $\mathcal{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_D$ is an orthonormal basis for $\mathbb{V}$.

$$\text{Define } \mathbb{D} = \underbrace{[1..D] \times \ldots \times [1..D]}_{N},$$

and for each $\mathbf{d} \in \mathbb{D}$, define

$$\mathbf{e}_{[\otimes \mathbf{d}]} = \mathbf{e}_{d_1} \otimes \mathbf{e}_{d_2} \otimes \ldots \otimes \mathbf{e}_{d_n}$$

Then the set $\left\{\mathbf{e}_{[\otimes \mathbf{d}]}\,;\, \mathbf{d} \in \mathbb{D}\right\}$ is a linear basis of $\mathbb{V}^{\otimes N}$. We can induce a natural inner product structure on $\mathbb{V}^{\otimes N}$ by legislating this basis to be orthonormal.

We can identify each element $\mathbf{a} \in \mathbb{V}_1 \otimes \ldots \otimes \mathbb{V}_N$ with a $D_1 \times D_2 \times \ldots \times D_N$ **array** of numbers $\boxed{\mathrm{A}} = \left[a_{\mathbf{d}}|_{\mathbf{d} \in \mathbb{D}}\right]$, so that

$$\mathbf{a} = \sum_{\mathbf{d} \in \mathbb{D}} a_{\mathbf{d}} \mathbf{e}_{\mathbf{d}}$$

In other words, $\mathbf{a}$ is the N-linear form on $\mathbb{V}$ so that, for each $(d_1, \ldots, d_N) \in \mathbb{D}$,

$$\mathbf{a}(\mathbf{e}_{d_1}, \mathbf{e}_{d_2}, \ldots, \mathbf{e}_{d_N}) = a_{(d_1, \ldots, d_N)}$$

Now, suppose that $\mathbf{v}_1, \ldots, \mathbf{v}_N \in \mathbb{V}$ are arbitrary elements, with $\mathbf{v}_n = \sum_{d=1}^{D} v_{n,d} \mathbf{e}_d$. Then the **array** corresponding to $\mathbf{v}_1 \otimes \ldots \otimes \mathbf{v}_n$ is computed as follows:

$$\begin{array}{rcl} \mathbf{v}_1 \otimes \ldots \otimes \mathbf{v}_n & = & \left(\sum_{d_1=1}^{D} v_{1,d_1} \mathbf{e}_{d_1}\right) \otimes \left(\sum_{d_2=1}^{D} v_{2,d_2} \mathbf{e}_{d_2}\right) \otimes \ldots \otimes \left(\sum_{d_N=1}^{D} v_{N,d_N} \mathbf{e}_{d_N}\right) \\ & = & \sum_{d_1=1}^{D} \sum_{d_2=1}^{D} \ldots \sum_{d_N=1}^{D} \left(v_{1,d_1} \cdot v_{2,d_2} \cdot \ldots \cdot v_{N,d_N}\right)\; \mathbf{e}_{d_1} \otimes \mathbf{e}_{d_2} \otimes \ldots \otimes \mathbf{e}_{d_N} \\ & = & \sum_{\mathbf{d} \in \mathbb{D}} \left(\prod_{n=1}^{N} v_{n,d_n}\right) \cdot \mathbf{e}_{[\otimes \mathbf{d}]} \end{array}$$

In other words, $\mathbf{v}_1 \otimes \ldots \otimes \mathbf{v}_n$ has the array $\boxed{\mathrm{A}} = \left[a_{\mathbf{d}}|_{\mathbf{d}\in\mathbb{D}}\right]$, where, for all $\mathbf{d} \in \mathbb{D}$,

$$a_{\mathbf{d}} = v_{1,d_1} \cdot v_{2,d_2} \cdot \ldots \cdot v_{N,d_N}$$

In other words, $\boxed{\mathrm{A}}$ can be imagined as the "product" of N one-dimensional arrays $\boxed{A_1}, \ldots, \boxed{A_N}$, where

$$\boxed{A_n} = [v_{n,1}, \ldots, v_{n,D}] .$$

E.2 Anti-Symmetric Tensors: The Exterior Product

Let $\boldsymbol{\sigma}(N)$ denote the permutation group on $[1..N]$. A tensor $\mathbf{a} \in \mathbb{V}^{\otimes N}$ is called **antisymmetric** (or **alternating**) if, for any $\mathbf{v}_1, \ldots, \mathbf{v}_N$, and any $\sigma \in \boldsymbol{\sigma}(N)$,

$$\mathbf{a}\left(\mathbf{v}_{\sigma(1)}, \ldots, \mathbf{v}_{\sigma(N)}\right) = \mathbf{sign}[\sigma] \cdot \mathbf{a}(\mathbf{v}_1, \ldots, \mathbf{v}_N)$$

The linear subspace of antisymmetric tensors in $\mathbb{V}^{\otimes N}$ is denoted by $\bigwedge^N \mathbb{V}$, and is called the N-fold **exterior product** of $\mathbb{V}$. For any $\mathbf{v}_1, \ldots, \mathbf{v}_N \in \mathbb{V}$, define

$$\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \ldots \wedge \mathbf{v}_N = \sum_{\sigma\in\boldsymbol{\sigma}(N)} \mathbf{sign}[\sigma] \cdot \mathbf{v}_{\sigma(1)} \otimes \ldots \otimes \mathbf{v}_{\sigma(N)}$$

then $\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \ldots \wedge \mathbf{v}_N$ is an antisymmetric tensor. Notice that permuting $\mathbf{v}_1, \ldots, \mathbf{v}_N$ does not change the value of $\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \ldots \wedge \mathbf{v}_N$, except perhaps in sign.

$$\text{Let } \mathbb{D}^{(<)} = \{(d_1, \ldots, d_N) \; ; \; 1 \le d_1 < d_2 < \ldots < d_N \le D\}$$

and for each $\mathbf{d} \in \mathbb{D}^{(<)}$, define

$$\mathbf{e}_{[\wedge\mathbf{d}]} = \mathbf{e}_{d_1} \wedge \mathbf{e}_{d_2} \wedge \ldots \wedge \mathbf{e}_{d_N}$$

then $\left\{\mathbf{e}_{[\wedge\mathbf{d}]} \; ; \; \mathbf{d} \in \mathbb{D}^{(<)}\right\}$ is a basis for $\bigwedge^N \mathbb{V}$. If we restrict the natural inner product on $\mathbb{V}^{\otimes N}$ (where the elements of $\left\{\mathbf{e}_{[\otimes\mathbf{d}]} \; ; \; \mathbf{d} \in \mathbb{D}^{(<)}\right\}$ are orthonormal), to an inner product on $\bigwedge^N \mathbb{V}$, then this is an **orthonormal** basis.

E.3 Multilinear Forms

Let $\mathbb{V}$ be a D-dimensional vector space, and let $N \in [0..D]$.

An **N-linear form** on $\mathbb{V}$ is an element of $\bigwedge^N \mathbb{V}^*$. In other words, an N-linear form is a function

$$\omega : \underbrace{\mathbb{V} \times \mathbb{V} \times \ldots \times \mathbb{V}}_{N} \longrightarrow \mathbb{C}$$

which is linear in each variable, and antisymmetric under permutation of the variables.

Think of an element of $\mathbb{V}^N$ as "frame" of N vectors, and imagine that this frame "spans" an N-dimensional parallellipiped within $\mathbb{V}$. An N-form provides a way of measuring the (oriented) N-dimensional "area" of this parallellipiped.

For example, suppose $\mathbb{V}$ is an inner product space, with orthonormal basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_D\}$ and corresponding dual basis $\{\mathbf{e}_1^*, \ldots, \mathbf{e}_D^*\}$, and suppose $\omega = \mathbf{e}_1^* \wedge \ldots \mathbf{e}_N^*$. Thus, if $(\vec{v}_1, \ldots, \vec{v}_N)$ is an N-frame spanning parallellipiped $\mathbf{P}$, then $\omega(\vec{v}_1, \ldots, \vec{v}_N)$ measures the area of the "shadow" $\mathbf{P}$ casts when projects down onto the subspace spanned by the first N coordinates.

E.4 Hodge Duality

Suppose that $\mathbb{V}$ is an D-dimensional vector space equipped with an inner-product, with orthonormal basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_D\}$ and corresponding dual basis $\{\mathbf{e}_1^*, \ldots, \mathbf{e}_D^*\}$ for $\mathbb{V}^*$. The D-linear form

$$\Omega \;:=\; \mathbf{e}_1^* \wedge \mathbf{e}_2^* \wedge \ldots \wedge \mathbf{e}_D^*$$

is a natural analog to the **determinant** on $\mathbb{V}$. Suppose $\vec{v}_1, \ldots, \vec{v}_D \in \mathbb{V}$, where, for each $d \in [1..D]$,

$$\vec{v}_d \;=\; v_{1d}\,\mathbf{e}_1 + v_{2d}\,\mathbf{e}_2 + \ldots + v_{Dd}\,\mathbf{e}_D$$

Then

$$\Omega(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_D) \;=\; \det \begin{bmatrix} v_{11} & v_{12} & \ldots & v_{1D} \\ v_{21} & v_{22} & \ldots & v_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ v_{D1} & v_{D2} & \ldots & v_{DD} \end{bmatrix}$$

In the same fashion, an N-form can be thought of as a scalar multiple of a "determinant" type function on some suitably chosen N-dimensional subspace of $\mathbb{V}$.

Definition 105: *Norm*

Let $\mathbb{V}$ be a D-dimensional inner product space, and let ω be a linear N-form on $\mathbb{V}$. The **norm** *of ω is defined:*

$$\|\omega\| \ = \ \sup\left\{|\omega(\mathbf{v}_1,\ldots,\mathbf{v}_N)|^{1/N} \ ; \ (\mathbf{w}_1,\ldots,\mathbf{w}_N)\in\mathbb{V}^N \ \text{ an orthonormal frame}\right\}$$

The reason for the exponent $1/N$ in the definition of $\|\omega\|$ is so that $\|\bullet\|$ will actually be a norm....

Proposition 106: *Properties of the Form Norm*

1. *$\|\bullet\|$ is a norm on the vector space $\bigwedge^N \mathbb{V}$ of linear N-forms on $\mathbb{V}$. In other words:*
   - *If ω is an N-linear form and $c>0$, then*
   $$\|c.\omega\| \ = \ c\cdot\|\omega\| \, .$$
   - *If α is another N-linear form, then*
   $$\|\omega+\alpha\| \ \leq \ \|\omega\|+\|\alpha\| \, .$$
2. *There is an orthonormal frame $(\mathbf{w}_1,\ldots,\mathbf{w}_N)\in\mathbb{V}^N$ so that*
   $$\omega(\mathbf{v}_1,\ldots,\mathbf{v}_N) \ = \ \|\omega\|$$
3. *Let $\mathbb{W}=span\{\mathbf{w}_1,\ldots,\mathbf{w}_N\}$, with the orientation induced by this ordering of the basis elements, and with the inner product structure it inherits from $\mathbb{V}$. Let $\mathbb{SO}\,[\mathbb{W}]$ be the group of orthogonal transformations of $\mathbb{W}$ with determinant 1. Then*
   - *If $g\in\mathbb{SO}\,[\mathbb{W}]$, then $(g.\mathbf{w}_1,\ldots,g.\mathbf{w}_N)$ is another orthonormal frame in $\mathbb{W}$ such that $\omega(g.\mathbf{w}_1,\ldots,g.\mathbf{w}_N) \ = \ 1$.*
   - *Conversely, if $(\mathbf{w}'_1,\ldots,\mathbf{w}'_N)$ is any orthonormal frame in $\mathbb{V}$ so that $\omega(\mathbf{w}'_1,\ldots,\mathbf{w}'_N)=1$, then $(\mathbf{w}'_1,\ldots,\mathbf{w}'_N) \ = \ (g.\mathbf{w}_1,\ldots,g.\mathbf{w}_N)$ for some $g\in\mathbb{SO}\,[\mathbb{W}]$.*
4. *In particular, if ω is any linear D-form, then*
   $$\omega \ = \ \pm c.\Omega$$
   where Ω is the determinant D-form on $\mathbb{V}$, and $c=\|\omega\|$.

□

Example 107:

Suppose that $\omega = c_1 \mathbf{e}_1^* \wedge \ldots \wedge c_N \mathbf{e}_N^*$. Then $\|\omega\| = (c_1 \cdot c_2 \cdot \ldots \cdot c_N)^{1/N}$, and ω attains this value when evaluated on the orthonormal frame $(\mathbf{e}_1, \ldots, \mathbf{e}_N)$, or on any image of $(\mathbf{e}_1, \ldots, \mathbf{e}_N)$ under an orthogonal transformation of the space $\mathbb{W} = span\{\mathbf{e}_1, \ldots, \mathbf{e}_N\}$. ———————————

Now, if $N \in [0..D]$, and $\omega \in \bigwedge^N \mathbb{V}^*$ is an N-form, then there is a unique $(D-N)$-form, $(\star\omega) \in \bigwedge^{(D-N)} \mathbb{V}^*$ so that

$$\omega \wedge (\star\omega) \;=\; \|\omega\| \cdot \Omega$$

For example, if $\omega = \mathbf{e}_1^* \wedge \ldots \wedge \mathbf{e}_N^*$, then $\star\omega \;=\; \mathbf{e}_{N+1}^* \wedge \ldots \wedge \mathbf{e}_D^*$.

The operation $\star : \bigwedge^N \mathbb{V}^* \longrightarrow \bigwedge^{(D-N)} \mathbb{V}^*$ is called the **Hodge Star Operator**, and is a canonical linear isomorphism between $\bigwedge^N \mathbb{V}^*$ and $\bigwedge^{(D-N)} \mathbb{V}^*$.

If course, there is actually a different Hodge Star operator for every $N \in [0..D]$; however normally, all are refered to simply as "the" Hodge star.

Since $N = D - (D-N)$, we have dual Hodge star operators

$$\star_N : \bigwedge^N \mathbb{V}^* \;\longrightarrow\; \bigwedge^{(D-N)} \mathbb{V}^*$$

$$\text{and } \star_{(D-N)} : \bigwedge^{(D-N)} \mathbb{V}^* \;\longrightarrow\; \bigwedge^N \mathbb{V}^*$$

It is straightforward to verify:

$$\star_{(D-N)} \circ \star_N \;=\; (-1)^{N(D-N)} \cdot \mathbf{Id}.$$

In a sense, the Hodge star satisfies a certain sort of self-adjointness property...

Proposition 108: *Self-Adjointness of Hodge Star*

Let $\mathbb{V}$ be a D-dimensional inner product space. Let α andβ be linear N-form, where $N \leq D$. Then, as linear D-forms,

$$(\star\alpha) \wedge \beta \;=\; \alpha \wedge (\star\beta)$$

Proof: Since the operator "$\star$" is linear, and the operation "$\wedge$" is bilinear, it suffices to show that this is true for elements from a basis of $\bigwedge^N \mathbb{V}$. So, let $\mathbf{e}_1, \ldots, \mathbf{e}_D$ be an orthonormal basis of $\mathbb{V}$. For any subset $S \subset [1..D]$, with $S = \{s_1 < s_2 < \ldots < s_N\}$, define

$$\mathbf{e}_{[\wedge S]} \;:=\; \mathbf{e}_{s_1} \wedge \mathbf{e}_{s_2} \wedge \ldots \wedge \mathbf{e}_{s_N}$$

Then $\left\{ \mathbf{e}_{[\wedge S]} \;;\; \mathcal{C}\!\mathit{ard}\,[S] = N \right\}$ is a basis for $\bigwedge^N \mathbb{V}$. Furthermore, for any such S,

$$\star \mathbf{e}_{[\wedge S]} \;=\; (-1)^P \mathbf{e}_{[\wedge S^c]}$$

where $S^c = [1..D] \setminus S$, and $P = \sum_{n=1}^{N} (s_n - n)$.

Thus, for any sets $S, T \subset [1..D]$ with $\mathcal{C}\!\mathit{ard}\,[S] = \mathcal{C}\!\mathit{ard}\,[T] = N$, notice that, if S^c is *not* disjoint from T, then

$$\left(\star \mathbf{e}_{[\wedge S]}\right) \wedge \mathbf{e}_{[\wedge T]} \;=\; (-1)^P \mathbf{e}_{[\wedge S^c]} \wedge \mathbf{e}_{[\wedge T]} \;=\; 0$$

and in this case, T^c is not disjoint from S either, so that

$$\mathbf{e}_{[\wedge S]} \wedge \left(\star \mathbf{e}_{[\wedge T]}\right) \;=\; (-1)^P \mathbf{e}_{[\wedge S]} \wedge \mathbf{e}_{[\wedge T^c]} \;=\; 0$$

On the other hand, since $\mathcal{C}\!\mathit{ard}\,[S] = \mathcal{C}\!\mathit{ard}\,[T] = N$, it is clear that

$$\left(\; S^c \text{ is disjoint from } T \; \right) \iff \left(\; S = T \; \right)$$

and in this case,

$$\begin{aligned} \left(\star \mathbf{e}_{[\wedge S]}\right) \wedge \mathbf{e}_{[\wedge S]} &= (-1)^P \mathbf{e}_{[\wedge S^c]} \wedge \mathbf{e}_{[\wedge S]} \\ &= (-1)^{P+Q} \mathbf{e}_{[\wedge S^c \sqcup S]} \\ &= \Omega \end{aligned}$$

where Ω is the determinant D-form, and similarly,

$$\mathbf{e}_{[\wedge S]} \wedge \left(\star \mathbf{e}_{[\wedge S]}\right) \;=\; \Omega$$

□

Appendix F

Differential Geometry

[For an excellent introduction to differential geometry, see [179]. An elegant (though more abstract) approach can be found in [32], which also contains an excellent introduction to Riemannian geometry. Another good discussion of Riemannian geometry, as well as an excellent introduction to tensor algebra, is [147]. For another approach, see [27]]

F.1 Manifolds

Definition 109: *Topological Manifold*

A **topological manifold** *of* **dimension** D *is a topological space* $\mathcal{M}$*, so that, for every* $m \in \mathcal{M}$*, there is an open neighbourhood* $\mathcal{U}_m \subset \mathcal{M}$ *and an open set* $\mathcal{V}_m \subset \mathbb{R}^D$*, and a homeomorphism* $\phi_m : \mathcal{V}_m \longrightarrow \mathcal{U}_m$ *(called a* **chart** *around* m*).*

The collection $\{\phi_m : \mathcal{V}_m \longrightarrow \mathcal{U}_m \; ; \; m \in \mathcal{M}\}$ *is called the* **atlas** *of the manifold.*

Remark 110: Actually, we don't even need the set $\mathcal{M}$ to be a topological space when we begin. If $\{\phi_m : \mathcal{V}_m \longrightarrow \mathcal{U}_m \; ; \; m \in \mathcal{M}\}$ is some collection of "charts" (where, for each $m \in \mathcal{M}$, $\mathcal{V}_m \subset \mathbb{R}^D$ is open, and $\mathcal{U}_m \subset \mathcal{M}$ is just some subset containing m), then we can *define* the topology on $\mathcal{M}$ to be the smallest topology so that:

- For every open subset $\mathcal{W} \subset \mathcal{V}_m$, the image $\phi_m(\mathcal{W}) \subset \mathcal{M}$ is open.

- For every $m \in \mathcal{M}$, $\mathcal{U}_m \subset \mathcal{M}$ is open.

This topology defines a topological manifold structure if and only if the maps $\phi_m : \mathcal{V}_m \longrightarrow \mathcal{U}_m$ are all homeomorphisms relative to it. Since the maps are open by construction, we need only to check continuity. To do this, it is sufficient to check that, for every $m, n \in \mathcal{M}$, the set $\phi_m^{-1}(\mathcal{U}_n)$ is open in $\mathbb{R}^D$.

Remark 111: Normally we impose the additional axiom that the space $\mathcal{M}$ be **second countable**. This is to exclude "extremely large" spaces such as the "long line" $\mathbb{R} \times \boldsymbol{\omega}$.

Suppose $\mathcal{M}$ is a topological manifold with atlas $\{\phi_m : \mathcal{V}_m \longrightarrow \mathcal{U}_m \; ; \; m \in \mathcal{M}\}$. For any $m, n \in \mathcal{M}$, with $\mathcal{U} = \mathcal{U}_m \cap \mathcal{U}_n \neq \emptyset$, let $\mathcal{V}_{m,n} := \phi_m^{-1}(\mathcal{U}) \subset \mathcal{V}_m$ and $\mathcal{V}_{n,m} := \phi_n^{-1}(\mathcal{U}) \subset \mathcal{V}_n$. Then

$$\phi_{nm} := \left(\phi_n^{-1} \circ \phi_m\right)_| : \mathcal{V}_{m,n} \longrightarrow \mathcal{V}_{n,m}$$

is a homeomorphism, by construction.

Definition 112: *Smooth Manifold*

A **smooth manifold** *is a topological manifold equipped with an atlas* $\{\phi_m : \mathcal{V}_m \longrightarrow \mathcal{U}_m \; ; \; m \in \mathcal{M}\}$ *so that, for any* $n, m \in \mathcal{M}$ *with* $\mathcal{U}_m \cap \mathcal{U}_n \neq \emptyset$, *the map* $\phi_{nm} : \mathcal{V}_{m,n} \longrightarrow \mathcal{V}_{n,m}$ *is a* diffeomorphism.

Such an atlas defines a **differentiable structure** on $\mathcal{M}$ in the following sense.

Definition 113: *Smooth Functions*

If $f : \mathcal{M} \longrightarrow \mathbb{C}$ *is continuous, then say that* f *is* **smooth** *if* $f \circ \phi_m : \mathcal{V}_m \longrightarrow \mathbb{C}$ *is smooth for all* $m \in \mathcal{M}$. *Then the set of smooth functions forms a* $\mathbb{C}$*-algebra under pointwise addition and multiplication —call this algebra* $\mathcal{C}^\infty(\mathcal{M})$.

If $\gamma : (-\epsilon, \epsilon) \longrightarrow \mathcal{M}$ *is a continuous function, then say* γ *is* **smooth** *if* $\phi_m^{-1} \circ \gamma : (-\epsilon, \epsilon) \longrightarrow \mathbb{R}^N$ *is smooth for all* $m \in \mathcal{M}$. *It is easy to check that, if* γ *and* f *are smooth, then* $f \circ \gamma$ *is smooth.*

If $\mathcal{M}$ *and* $\mathcal{N}$ *are two manifolds, then a function* $\psi : \mathcal{M} \longrightarrow \mathcal{N}$ *is* **smooth** *if, for every smooth chart* $\phi_m : \mathcal{V}_m \longrightarrow \mathcal{U}_m \subset \mathcal{M}$, *and every*

smooth chart $\phi_n : \mathcal{V}_n \longrightarrow \mathcal{U}_n \subset \mathcal{N}$, *if* $\mathcal{W} := \psi(\mathcal{U}_m) \cap \mathcal{U}_n \neq \emptyset$, *and if* $\mathcal{X} := \phi_m^{-1}\psi^{-1}(\mathcal{W}) \subset \mathcal{V}_m$, *then the function*

$$\left(\phi_n^{-1} \circ \phi \circ \phi_m\right)_| : \mathcal{X} \longrightarrow \mathbb{R}^{D_\mathcal{N}}$$

is smooth (where $D_\mathcal{N}$ *is the dimension of* $\mathcal{N}$*).*

F.2 Tangent Vectors

F.2.1 Tangent Spaces

Definition 114: *Differentiation functional*

A **differentiation functional** *on* $\mathcal{M}$ *is a* $\mathbb{C}$*-linear functional*

$$\partial : \mathcal{C}^\infty(\mathcal{M}) \longrightarrow \mathbb{C}$$

which satisfies the **Liebniz product rule***:*

$$\partial_\gamma[f \cdot g](m) \;=\; f(m) \cdot \partial[g](m) \;+\; g(m) \cdot \partial[f](m).$$

If $m \in \mathcal{M}$*, then we say* ∂ *is* **located at** m *if, for any* f *and* g*, if* f *and* g *agree in some neighbourhood of* m*, then* $\partial f = \partial g$*.*

The set of all differentiation functionals located at m *will be denoted by* $\mathsf{T}_m\,\mathcal{M}$*. It is not hard to see that* $\mathsf{T}_m\,\mathcal{M}$ *is a vector space.*

For example suppose $\gamma : (-\epsilon, \epsilon) \longrightarrow \mathcal{M}$ is a smooth curve, with $\gamma(0) = m$. The **derivative** of f at m, in the **direction** of γ, is then defined to be the real number

$$\partial_\gamma[f](m) \;:=\; (f \circ \gamma)'(0)$$

For fixed γ and m, the operation

$$\partial_\gamma[\bullet](m) : \mathcal{C}^\infty(\mathcal{M}) \longrightarrow \mathbb{C}$$

is easily seen to be a **differentiation** on the algebra $\mathcal{C}^\infty(\mathcal{M})$, located at m.

If $\gamma_1, \gamma_2 : (-\epsilon, \epsilon) \longrightarrow \mathcal{M}$, with $\gamma_1(0) = m = \gamma_2(0)$, then we say that γ_1 and γ_2 are **infinitesimally equivalent**, and write $\gamma_1 \sim \gamma_2$, if

$$(\phi_m^{-1} \circ \gamma_1)'(0) \;=\; (\phi_m^{-1} \circ \gamma_2)'(0)$$

It is easy to see that $\gamma_1 \sim \gamma_2$ if and only if the differentiation operations $\partial_{\gamma_1}[\bullet](m)$ and $\partial_{\gamma_1}[\bullet](m)$ are identical. It is also straightforward to show that *every* differentiation functional located at m arises in this manner.

Definition 115: *Tangent Space*

If $m \in \mathcal{M}$, then the **tangent space** *of $\mathcal{M}$ at m is the set of infinitesmal equivalence classes of smooth paths $\gamma : (-\epsilon, \epsilon) \longrightarrow \mathcal{M}$, with $\gamma(0) = m$. Equivalently, it is the set of all differentiation functions located at m, which we have already denoted by $\mathsf{T}_m \mathcal{M}$. This set has a natural vector space structure.*

Elements of $\mathsf{T}_m \mathcal{M}$ will be referred to as **tangent vectors**

Definition 116: *Tangent Bundle*

The **tangent bundle** *of $\mathcal{M}$ is a disjoint union containing all tangent spaces:*

$$\mathsf{T}\mathcal{M} \;=\; \{(m, \vec{v}) \;;\; m \in \mathcal{M} \text{ and } \vec{v} \in \mathsf{T}_m \mathcal{M}\}$$

The tangent bundle of $\mathcal{M}$ can be made into a $2D$-dimensional smooth manifold in a natural way. Suppose $\{\phi_m : \mathcal{V}_m \longrightarrow \mathcal{U}_m \;;\; m \in \mathcal{M}\}$ is a smooth atlas for $\mathcal{M}$. For every $m \in \mathcal{M}$, define

$$\mathsf{T}\mathcal{U} \;=\; \{(u, \vec{v}) \;;\; u \in \mathcal{U} \text{ and } \vec{v} \in \mathsf{T}_u \mathcal{M}\}$$

and then define $\psi_m : \mathcal{V}_m \times \mathbb{R}^D \longrightarrow \mathsf{T}\mathcal{U}$ so that, for any $(\mathbf{x}, \mathbf{v}) \in \mathcal{V}_m \times \mathbb{R}^D$,

$$\psi_m(\mathbf{x}, \mathbf{v}) \;=\; (\phi_m(\mathbf{x}), \partial)$$

where $\partial : \mathcal{C}^\infty(\mathcal{M}) \longrightarrow \mathcal{C}^\infty(\mathcal{M})$ is the differentiation functional defined

$$\partial(f) \;=\; \partial_{\vec{v}}(f \circ \phi_m)(\mathbf{x})$$

(where $\partial_{\vec{v}}$ indicates the directional derivative in along the vector $\vec{v}$ in $\mathbb{R}^D$)

We have not yet even defined a topology on $\mathsf{T}\mathcal{M}$. One can use atlas $\{\psi_m : \mathcal{V}_m \times \mathbb{R}^D \longrightarrow \mathsf{T}\mathcal{U}_m \;;\; m \in \mathcal{M}\}$ to define the topology on $\mathsf{T}\mathcal{M}$, as described by Remark 110 on page 207, and then use it to define a smooth structure.

Definition 117: *Derivative*

Let $f : \mathcal{M} \longrightarrow \mathcal{N}$ be a smooth map, with $m \in \mathcal{M}$ and $n = f(m) \in \mathcal{N}$. f induces a linear map

$$\mathsf{D}_m f : \mathsf{T}_m \mathcal{M} \longrightarrow \mathsf{T}_n \mathcal{N}$$

as follows: if $\gamma : (-\epsilon, \epsilon) \longrightarrow \mathcal{M}$ *is any smooth path with* $\gamma(0) = m$*, then* $f \circ \gamma : (-\epsilon, \epsilon) \longrightarrow \mathcal{N}$ *is a smooth path with* $\gamma(0) = n$*. It is not hard to check that:*

- *The map* $\gamma \mapsto f \circ \gamma$ *preserves infinitesmal equivalence, and thus maps elements of* $\mathsf{T}_m \mathcal{M}$ *into elements of* $\mathsf{T}_n \mathcal{N}$*.*
- *This map is linear.*

$\mathsf{D}_m f$ *is called the* **derivative** *of* f *at* m*.*

F.2.2 Vector Fields

Prerequisites:

- Tangent Spaces [F.2.1 on page 209]

Definition 118: *(I) Smooth vector field*

A **smooth vector field** *is a smooth function* $\mathbf{X} : \mathcal{M} \longrightarrow \mathsf{T}\mathcal{M}$ *so that, for all* $m \in \mathcal{M}$*,* $\mathbf{X}(m) \in \mathsf{T}_m \mathcal{M}$*.*

The equivalence between tangent vectors and differentiation functionals, along with the aforementioned smooth structure on $\mathsf{T}\mathcal{M}$, immediately implies that this definition is equivalent to the following one:

Definition 119: *(II) Smooth vector field*

A **smooth vector field** *is an operator*

$$\partial : \mathcal{C}^\infty(\mathcal{M}) \longrightarrow \mathcal{C}^\infty(\mathcal{M})$$

so that for any fixed $m \in \mathcal{M}$*, the map* $f \mapsto \partial f[m]$ *is a differentiation functional, located at* m*.*

The set of all smooth vector fields on $\mathcal{M}$ will be denoted $\mathfrak{S}\mathsf{T}\mathcal{M}$. It is not hard to see that

- $\mathfrak{S}\mathsf{T}\mathcal{M}$ is a $\mathbb{C}$-vector space, where, for all $\partial_1, \partial_2 \in \mathfrak{S}\mathsf{T}\mathcal{M}$, $c \in \mathbb{C}$, and $f \in \mathcal{C}^\infty(\mathcal{M})$, we define $(\partial_1 + c.\partial_2)[f] = \partial_1[f] + c.\partial_2[f]$

- $\mathfrak{S}\mathsf{T}\mathcal{M}$ is a $\mathcal{C}^\infty(\mathcal{M})$-module, where, for all $\partial_1, \partial_2 \in \mathfrak{S}\mathsf{T}\mathcal{M}$, and $f, g \in \mathcal{C}^\infty(\mathcal{M})$, we define $(\partial_1 + g.\partial_2)[f] = \partial_1[f] + g.\partial_2[f]$

It will often be useful to extend a tangent vector at a point to a vector field in a neighbourhood of that point.

Lemma 120: *Extension Lemma* *Let $m \in \mathcal{M}$. For any $\vec{v} \in \mathsf{T}_m\mathcal{M}$, there is a smooth vector field $\vec{V} \in \mathfrak{S}\mathsf{T}\mathcal{M}$ so that $\vec{V}(m) = \vec{v}$.* _____□

F.3 Differential Forms and Integration on Manifolds

F.3.1 Differential Forms

Prerequisites:

- Smooth vector fields [F.2.2 on the preceding page]

Let $\mathcal{M}$ be a smooth manifold, and for any $m \in \mathcal{M}$, let $\bigwedge^N \mathsf{T}^*_m\mathcal{M}$ denote the vector space of **linear N-forms** on $\mathsf{T}_m\mathcal{M}$ (see Appendix E.2 on page 201). Then define the **N-form bundle**:

$$\bigwedge^N \mathsf{T}^*\mathcal{M} := \left\{ (m, \omega) \; ; \; m \in \mathcal{M}, \;\; \omega \in \bigwedge^N \mathsf{T}^*_m\mathcal{M} \right\}$$

Definition 121: *(I) Differential Form*

A **differential N-form** *on $\mathcal{M}$ is a map*

$$\omega : \mathcal{M} \longrightarrow \bigwedge^N \mathsf{T}^*\mathcal{M}$$

so that, for all $m \in \mathcal{M}$, $\omega(m) \in \bigwedge^N \mathsf{T}^*_m\mathcal{M}$, and so that, if $\vec{V}_1, \ldots, \vec{V}_N \in \mathfrak{S}\mathsf{T}\mathcal{M}$ are smooth vector fields, then the function

$$\begin{array}{rccl} \omega[\vec{V}_1, \ldots, \vec{V}_N] : & \mathcal{M} & \longrightarrow & \mathbb{C} \\ & m & \mapsto & \omega(m)\left[\vec{V}_1(m), \ldots, \vec{V}_N(m)\right] \end{array}$$

is in $\mathcal{C}^\infty(\mathcal{M})$.

It is straightforward to show that this is equivalent to:

Definition 122: *(II)Differential Form*

Let $\mathcal{M}$ be a manifold of dimension D, and let $N \in [0..D]$. A **differential N-form** *on $\mathcal{M}$ is an N-linear map*

$$\omega : \underbrace{\mathfrak{S}\mathsf{T}\mathcal{M} \times \ldots \times \mathfrak{S}\mathsf{T}\mathcal{M}}_{N} \longrightarrow \mathcal{C}^\infty\mathcal{M}$$

so that:

1. *For any vector fields $\vec{V}_1, \ldots, \vec{V}_N \in \mathfrak{S}\mathsf{T}\mathcal{M}$, and any $m \in \mathcal{M}$, the value of*

   $$\omega[\vec{V}_1, \ldots, \vec{V}_N](m)$$

   depends only upon the values of $\vec{V}_1(m), \ldots, \vec{V}_D(m)$. Thus, if we define $\vec{v}_1 := \vec{V}_1(m), \ \ldots, \vec{v}_D := \vec{V}_D(m)$, we can define

   $$\omega_m[\vec{v}_1, \ldots, \vec{v}_N] \ := \ \omega[\vec{V}_1, \ldots, \vec{V}_N](m)$$

2. *For every $m \in \mathcal{M}$, the function*

   $$\omega_m : \underbrace{\mathsf{T}_m\mathcal{M} \times \ldots \times \mathsf{T}_m\mathcal{M}}_{N} \longrightarrow \mathbb{C}$$

   is an alternating N-linear form

The set of all differentiable N-forms will be writen as:

$$\mathfrak{S}\bigwedge^N \mathsf{T}^*\mathcal{M}$$

and is clearly a vector space under pointwise addition and scalar multiplication.

Note that the space $\mathfrak{S}\bigwedge^0 \mathsf{T}^*\mathcal{M}$ is basically the same as $\mathcal{C}^\infty(\mathcal{M})$.

A linear N-form provides a way of measuring the "area" of N-dimensional parallellipipeds. A differential N-form defines a linear N-form at every point along a manifold; thus, it defines a way of measuring the "area" of N-dimensional submanifolds.

Definition 123: *Integration of differential N-forms*

Let $\mathcal{M}$ be a D-dimensional manifold, and let ω be a differential N form, with $N \in [0..D]$. Let $\mathcal{V} \subset \mathbb{R}^N$, and let $\phi : \mathcal{V} \longrightarrow \mathcal{M}$ be a smooth embedding, so that $\mathcal{N} := \phi(\mathcal{V})$ is a N-dimensional submanifold.

Let $\{\mathbf{e}_1, \ldots, \mathbf{e}_N\}$ be the standard orthonormal basis of $\mathbb{R}^N$.

The **integral** *of ω on $\mathcal{N}$ is defined:*

$$\omega(\mathcal{N}) \quad = \quad \int_{\mathcal{N}} \omega \quad := \quad \int_{\mathcal{V}} \omega \left[\mathsf{D}_{\mathbf{x}}\phi(\mathbf{e}_1), \ldots, \mathsf{D}_{\mathbf{x}}\phi(\mathbf{e}_N)\right] \; d\mathcal{L}^{\text{leb}\,N}[\mathbf{x}]$$

If $\mathcal{N}$ is an arbitrary N-dimensional submanifold of $\mathcal{M}$, then we define the integral of ω on $\mathcal{N}$ by partitioning $\mathcal{N}$ into locally Euclidean patches in the obvious way. It is straightforward to show that the resulting definition is independent of the choice of partition.

F.3.2 Exterior Derivatives

Prerequisites:

- Differential Forms [F.3.1 on page 212]

Definition 124: *Exterior Derivative*

Let $\mathcal{M}$ be a D-dimensional manifold. For each $N \in [0..D)$, we define the linear maps $\mathfrak{d}_{[N]} : \mathfrak{S} \bigwedge^N \mathsf{T}^*\mathcal{M} \longrightarrow \mathfrak{S} \bigwedge^{N+1} \mathsf{T}^*\mathcal{M}$, to be the unique maps satisfying the following axioms:

1. *If $N = 0$ (ie. ω is an element of $\mathcal{C}^\infty(\mathcal{M})$), then $\mathfrak{d}_{[0]}\omega$ is the 1-form defined:*
$$(\mathfrak{d}_{[0]}\omega)(\vec{V}) \quad = \quad \vec{V}\omega$$
(where we treat $\vec{V} \in \mathfrak{S}\mathsf{T}\mathcal{M}$ as a differentiation operator acting on ω.)
2. *For any $\omega_1 \in \mathfrak{S} \bigwedge^{N_1} \mathsf{T}^*\mathcal{M}$ and $\omega_2 \in \mathfrak{S} \bigwedge^{N_2} \mathsf{T}^*\mathcal{M}$, with $N = N_1 + N_2$,*
$$\mathfrak{d}_{[N]}(\omega_1 \wedge \omega_2) \quad = \quad (\mathfrak{d}_{[N_1]}\omega_1) \wedge \omega_2 \; + \; (-1)^N \omega_1 \wedge (\mathfrak{d}_{[N_2]}\omega_2).$$
3. $\mathfrak{d}_{[N+1]} \circ \mathfrak{d}_{[N]} \equiv 0.$

Normally we suppress the subscript "$_{[N]}$", and just write "$\mathfrak{d}$". The map $\mathfrak{d}$ is called the **exterior derivative**.

Remark 125: The previous axioms completely characterise $\mathfrak{d}$ because they force it to behave in a unique fashion relative to any coordinate system. Suppose that $\mathcal{V} \subset \mathbb{R}^D$ and $\phi : \mathcal{V} \longrightarrow \mathcal{U} \subset \mathcal{M}$ is some smooth chart, inducing local coordinate vector fields $\vec{\mathbf{E}}_1, \ldots, \vec{\mathbf{E}}_D \in \mathfrak{S}\mathsf{T}\mathcal{U}$ and corresponding covector fields $\vec{\mathbf{E}}^*_1, \ldots, \vec{\mathbf{E}}^*_D \in \mathfrak{S}\mathsf{T}^*\mathcal{U}$. Suppose that $\omega \in \mathfrak{S}\bigwedge^N \mathsf{T}^*\mathcal{U}$ is defined:

$$\omega \;=\; \sum_{\mathbf{k}:=(k_1,\ldots,k_N)} f_{\mathbf{k}} \cdot \vec{\mathbf{E}}^*_{k_1} \wedge \ldots \wedge \vec{\mathbf{E}}^*_{k_N}$$

where $\mathbf{k}$ ranges over some suitable set of multiindices, and $f_{\mathbf{k}} \in \mathcal{C}^\infty(\mathcal{U})$ for all $\mathbf{k}$. Then the previous axioms imply:

$$\mathfrak{d}\omega \;=\; \sum_{\mathbf{k}:=(k_1,\ldots,k_N)} (\mathfrak{d}f_{\mathbf{k}}) \wedge \vec{\mathbf{E}}^*_{k_1} \wedge \ldots \wedge \vec{\mathbf{E}}^*_{k_N}$$

where, for all $\mathbf{k}$, $\mathfrak{d}f_{\mathbf{k}}$ is the 1-form:

$$\mathfrak{d}f_{\mathbf{k}} \;=\; \sum_{d=1}^{D} (\mathfrak{d}_d f_{\mathbf{k}}) \cdot \vec{\mathbf{E}}^*_d$$

In the case when ω is a 0-form (ie. smooth scalar field), the exterior derivative of ω (as a form) corresponds to the **gradient** of ω (as a scalar field). When ω is a vector field, the exterior derivative (in $\mathbb{R}^3$) creates an object analogous to the **curl**. By composing with the **Hodge star** (see Appendix E.4 on page 202), we can also induce an object analogous to the **divergence** (see Appendix G.5 on page 223 for more on this).

F.3.3 Gauss, Green, and Stokes

Prerequisites:

- Exterior Derivatives [F.3.2 on the preceding page]

The utility of exterior differentiation comes from the following far-reaching generalization of a result of classical integral calculus:

Theorem 126: *Gauss, Green, Stokes Theorem*
Let $\mathcal{M}$ be a D-dimensional manifold, and let ω be a differential N form, with $N \in [0..D]$.

Suppose that $\mathcal{N}$ is an $(N+1)$-dimensional submanifold of $\mathcal{M}$, with boundary $\mathfrak{d}\mathcal{N}$ (an N-dimensional submanifold). Then

$$\mathfrak{d}\omega(\mathcal{N}) \;=\; \omega(\partial\mathcal{N})$$

□

Corollary 127: *Integration By Parts on Manifolds*

Let $\mathcal{M}$ be a D-dimensional manifold, and let α and β be a differential A-form and B-form, respectively, where $A+B=D-1$. Thus, $(\mathfrak{d}\alpha)\wedge\beta$ and $\alpha\wedge(\mathfrak{d}\beta)$ are both differential D-forms, and we have:

$$\int_{\mathcal{M}}(\mathfrak{d}\alpha)\wedge\beta \;=\; -\int_{\partial\mathcal{M}}\alpha\wedge\beta \;+\; (-1)^A\int_{\mathcal{M}}\alpha\wedge(\mathfrak{d}\beta).$$

In particular, if $\mathcal{M}$ is a manifold without boundary, then

$$\int_{\mathcal{M}}(\mathfrak{d}\alpha)\wedge\beta \;=\; (-1)^A\int_{\mathcal{M}}\alpha\wedge(\mathfrak{d}\beta)$$

Proof: First note that

$$\mathfrak{d}(\alpha\wedge\beta) \;=\; (\mathfrak{d}\alpha)\wedge\beta \;+\; (-1)^A\alpha\wedge(\mathfrak{d}\beta)$$

Thus, applying Stoke's theorem,

$$\begin{aligned}\int_{\partial\mathcal{M}}\alpha\wedge\beta &= \int_{\mathcal{M}}\mathfrak{d}(\alpha\wedge\beta)\\ &= \int_{\mathcal{M}}(\mathfrak{d}\alpha)\wedge\beta \;+\; (-1)^A\int_{\mathcal{M}}\alpha\wedge(\mathfrak{d}\beta)\end{aligned}$$

□

Appendix G

Riemannian Geometry

G.1 Riemann Metrics

Prerequisites:

- Smooth vector fields [F.2.2 on page 211]

If $\mathbb{V}$ is any vector space, let $\bigotimes^2 \mathbb{V}^*$ denote the set of all **bilinear forms** from $\mathbb{V} \times \mathbb{V} \longrightarrow \mathbb{R}$.

If $\mathcal{M}$ is a manifold, define the **bilinear form bundle:**

$$\bigotimes^2 \mathsf{T}^* \mathcal{M} \ := \ \left\{ (m, \mathbf{g}) \ ; \ m \in \mathcal{M} \text{ and } \mathbf{g} \in \bigotimes^2 \mathsf{T}^*_m \mathcal{M} \right\}$$

Definition 128: *(I) Bilinear Form Field*

A **bilinear form field** *is a function* $\mathbf{g} : \mathcal{M} \longrightarrow \bigotimes^2 \mathsf{T}^* \mathcal{M}$ *so that, for all* $m \in \mathcal{M}$, $\mathbf{g}(m) \in \bigotimes^2 \mathsf{T}^*_m \mathcal{M}$, *and so that, for any smooth vector fields* $\vec{\mathbf{X}}_1, \vec{\mathbf{X}}_2 : \mathcal{M} \longrightarrow \mathsf{T}\mathcal{M}$, *the function*

$$\begin{array}{rccl} \mathbf{g}(\vec{\mathbf{X}}_1, \vec{\mathbf{X}}_2) : \mathcal{M} & \longrightarrow & \mathbb{R} \\ m & \mapsto & \mathbf{g}[m] \left(\vec{\mathbf{X}}_1[m], \vec{\mathbf{X}}_2[m] \right) \end{array}$$

is smooth.

Again, it is straightforward to show this is equivalent to

Definition 129: *(II) Bilinear Form Field*

A **bilinear form field** *is a $\mathcal{C}^\infty(\mathcal{M})$-bilinear function:*

$$\mathbf{g} : \mathfrak{S}\mathsf{T}\mathcal{M} \times \mathfrak{S}\mathsf{T}\mathcal{M} \longrightarrow \mathcal{C}^\infty(\mathcal{M})$$

Definition 130: *Riemann Metric*

Riemann Metric *is a bilinear form field* $\mathbf{g}$ *which is*

- **Symmetric:** $\mathbf{g}(\vec{\mathbf{X}}_1, \vec{\mathbf{X}}_2) = g(\vec{\mathbf{X}}_2, \vec{\mathbf{X}}_1)$
- **Positive Definite:** $g(\vec{\mathbf{X}}, \vec{\mathbf{X}}) > 0$ *for any* $\vec{\mathbf{X}} \in \mathfrak{S}\mathsf{T}\mathcal{M}$ *that is not zero everywhere.*

A **Riemannian Manifold** *is a manifold equipped with a Riemann metric.*

A Riemann metric endows the tangent space $\mathsf{T}_m\mathcal{M}$ with an **inner product** in a natural way. If $\vec{v}, \vec{w} \in \mathsf{T}_m\mathcal{M}$ are two vectors, we can find smooth vector fields $\vec{V}, \vec{W} : \mathcal{M} \longrightarrow \mathsf{T}\mathcal{M}$ so that $\vec{V}(m) = \vec{v}$ and $\vec{W}(m) = \vec{w}$. Then we define the **inner product** of $\vec{v}$ and $\vec{w}$ (relative to $\mathbf{g}$ by:

$$\langle \vec{v}, \vec{w} \rangle_{\mathbf{g}} = \mathbf{g}(\vec{V}, \vec{W})[m]$$

it is straightforward to check that this definition does not depend upon the choice of "extensions" $\vec{V}$ and $\vec{W}$, and that it defines a positive-definite, symmetric bilinear form at $\mathsf{T}_m\mathcal{M}$.

Thus, $\mathbf{g}$ endows each point in the manifold with a "local geometric structure", by providing a way to measure the "lengths" and "angles" of tangent vectors. This, in turn, allows us to define a natural metric structure on $\mathcal{M}$, by defining a way of measuring path length.

G.2 Distance and Geodesics

Prerequisites:

- Riemann metrics [G.1 on the page before]

Definition 131: *Velocity*

Let $\gamma : (-\epsilon, \epsilon) \longrightarrow \mathcal{M}$ be some smooth path, with $\gamma(0) = m$. The **velocity** *of γ at 0 is defined:*

$$\left\|\gamma'(0)\right\|_{\mathbf{g}} = \sqrt{\langle \gamma'(0), \gamma'(0) \rangle_{\mathbf{g}}}$$

Definition 132: *Path Length*

If $\gamma : (A, B) \longrightarrow \mathcal{M}$ is a smooth path, and $[a, b] \subset (A, B)$, then we define the **length** *of γ between $\gamma(a)$ and $\gamma(b)$ to be:*

$$\ell_{ength}\left[\gamma(a, b)\right] = \int_a^b \left\|\gamma'(t)\right\|_{\mathbf{g}} \, dt$$

Definition 133: *The Path-Length Metric*

The **path-length metric** *on $\mathcal{M}$ is then defined as follows: For any $m_0, m_1 \in \mathcal{M}$,*

$$\mathbf{dist}\left[m_0,\ m_1\right] = \inf\left\{ \ell_{ength}\left[\gamma(0,1)\right];\quad \begin{array}{r} \gamma : (A, B) \longrightarrow \mathcal{M} \text{ a smooth path, with} \\ \gamma(0) = m_0,\ \ \gamma(1) = m_1 \end{array} \right\}$$

This metric is compatible with the locally Euclidean topology induced upon $\mathcal{M}$ by the smooth atlas.

Definition 134: *Geodesic*

A **geodesic** *is a path $\gamma : (A, B) \longrightarrow \mathcal{M}$ which has* constant velocity, *and which* locally minimizes *path length. In other words,*

- $\left\|\gamma'(t)\right\|_{\mathbf{g}} = v$ *for some constant $v > 0$.*
- *For every $t \in (A, B)$, there is some open interval $(a, b) \subset (A, B)$, with $t \in (a, b)$, so that*

$$\ell_{ength}\left[\gamma(a, b)\right] = \mathbf{dist}\left[\gamma(a),\ \ \gamma(b)\right]$$

(In other words, γ is the "shortest possible path" from $\gamma(a)$ to $\gamma(b)$)

Lemma 135: *Let $m \in \mathcal{M}$ and $\vec{v} \in \mathsf{T}_m \mathcal{M}$. Then there is a unique geodesic $\gamma_{\vec{v}} : (-\epsilon, \epsilon) \longrightarrow \mathcal{M}$ so that $\gamma_{\vec{v}}(0) = m$ and $\gamma'_{\vec{v}}(0) = \vec{v}$.*

Definition 136: *Exponential Map*

Let $m \in \mathcal{M}$ and, for each $\vec{v} \in \mathsf{T}_m \mathcal{M}$, define $\gamma_{\vec{v}} : (-\epsilon, \epsilon) \longrightarrow \mathcal{M}$ as before. We define the **exponential map** $\exp_m : \mathsf{T}_m \mathcal{M} \longrightarrow \mathcal{M}$ *by:*

$$\exp_m(\vec{v}) \;=\; \gamma_{\vec{v}}(1)$$

whenever $\gamma_{\vec{v}}(1)$ is well-defined.

(hence, in general, $\exp_m$ may only be well-defined on some subset of $\mathsf{T}_m \mathcal{M}$.)

Proposition 137: *Properties of the Exponential Map*

- *There exists an open neighbourhood $\mathcal{V} \subset \mathsf{T}_m \mathcal{M}$ around 0, so that, then $\exp_{m|\mathcal{V}}$ is well-defined and injective.*
- *If $\mathcal{U} = \exp(\mathcal{V})$, then the map $\exp_{m|} : \mathcal{V} \longrightarrow \mathcal{U}$ is a diffeomorphism.*
- *Although the map $\exp_{m|} : \mathcal{V} \longrightarrow \mathcal{U}$ is not an isometry, it* is *true that, for all $\vec{v} \in \mathsf{T}_m \mathcal{M}$,*

$$\mathbf{dist}\left[\exp_m(\gamma_{\vec{v}}),\; m\right] \;=\; \|\vec{v}\|_{\mathbf{g}}\,.$$

Definition 138: *Geodesically Complete*

A Riemannian manifold $\mathcal{M}$ is called **geodesically complete** *if, for every $m_0, m_1 \in \mathcal{M}$, there is a geodesic $\gamma : [0,1] \longrightarrow \mathcal{M}$ so that $\gamma(0) = m_0$ and $\gamma(1) = m_1$.*

Equivalentely, $\mathcal{M}$ is geodesically complete if, for any $m \in \mathcal{M}$, the exponential map $\exp_m : \mathsf{T}_m \mathcal{M} \longrightarrow \mathcal{M}$ is surjective.

Theorem 139: *Hopf-Rinow theorem*

A Riemannian manifold is geodesically complete if and only if it is metrically complete, relative to the metric defined by geodesic distance between points.

In particular, any compact manifold is geodesically complete.

Proof: See [32]. □

G.3 Integration on Riemannian Manifolds

Prerequisites:

- Riemann metrics [G.1 on page 217]
- Differential forms [F.3.1 on page 212]

The Riemann metric also induces a natural measure on $\mathcal{M}$.

Definition 140: *Canonical volume-form*

Let $\mathcal{M}$ be a D-dimensional manifold

For any $m \in \mathcal{M}$, pick a linear isomorphism between $\phi_m : \mathsf{T}_m \mathcal{M} \longrightarrow \mathbb{R}^D$, which is an isometry relative to the inner product structure $\langle \bullet, \bullet \rangle_{\mathbf{g}}$ on $\mathsf{T}_m \mathcal{M}$ and the standard inner product on $\mathbb{R}^D$.

Let $\Omega \in \bigotimes^D \mathbb{R}^D$ be the determinant D-linear form on $\mathbb{R}^D$.

Define the **canonical volume form** *$\Omega_m \in \bigotimes^D \mathsf{T}_m \mathcal{M}^*$ as follows: for any $\vec{v}_1, \ldots, \vec{v}_D \in \mathsf{T}_m \mathcal{M}$,*

$$\Omega_m[\vec{v}_1, \ldots, \vec{v}_D] \;=\; \Omega\left[\phi_m(\vec{v}_1),\; \ldots, \phi_m(\vec{v}_D)\right].$$

Although this definition involves a specific coordinate system, the volume form is actually well-defined, independent of the choice of coordinates.

The linear D-forms $\{\Omega_m \;;\; m \in \mathcal{M}\}$ together define a differential D-form

$$\boldsymbol{\Omega} \in \mathfrak{S} \bigotimes^D \mathsf{T}^* \mathcal{M}$$

called the **canonical volume form** *of $\mathcal{M}$.*

Definition 141: *Canonical Measure*

The **canonical volume measure** *on $\mathcal{M}$ is the measure $\mathcal{L}^{\mathrm{leb}}{}_{\mathcal{M}}$ defined as follows.*

Suppose that $\mathcal{M}$ has smooth atlas $\{\phi_m : \mathcal{V}_m \longrightarrow \mathcal{U}_m \;;\; m \in \mathcal{M}\}$. Fix $m \in \mathcal{M}$; we will first define a "local measure", $\mathcal{L}^{\mathrm{leb}}{}_m$, on the set $\mathcal{U}_m$.

Let $\mathbf{e}_1, \ldots, \mathbf{e}_D$ be the standard orthonormal basis of $\mathbb{R}^D$, and define $\mathbf{L} : \mathcal{V}_m \longrightarrow \mathbb{R}$ so that, for any $v \in \mathcal{V}_m$, with $u = \phi_m(v)$,

$$\mathbf{L}(v) \;=\; \Omega_u\left[\mathsf{D}_v\phi_m(\mathbf{e}_1),\; \ldots,\; \mathsf{D}_v\phi_m(\mathbf{e}_D)\right]$$

Let $\mathcal{B} \subset \mathcal{U}_m$ be Borel-measurable. Let $\mathcal{B}_0 := \phi_m^{-1}(\mathcal{B})$, and define

$$\mathcal{L}^{bsg}{}_m[\mathcal{B}] \;=\; \int_{\mathcal{B}_0} \mathbf{L}\; d\mathcal{L}^{bsg\,D}$$

where $\mathcal{L}^{bsg\,D}$ is the standard Lebesgue measure on $\mathbb{R}^D$.

Now, if $\mathcal{B} \subset \mathcal{M}$ is arbitrary, break $\mathcal{B}$ up into disjoint peices which lie inside different chart domains, and define its measure to be the sums of the measures of the peices, in terms of the "local measures" just defined.

It is straightforward to check that this definition is independent of the manner in which $\mathcal{B}$ is broken up, and independent of which charts and local measures we use.

G.4 Hodge Duality

Prerequisites:

- Riemann metrics [G.1 on page 217]
- Hodge Duality [E.4 on page 202]

If $\mathcal{M}$ is a D-dimensional Riemannian manifold, then every tangent space is an inner product space. Hence, for any $m \in \mathcal{M}$, and every $D \in [0..D]$ there is a well-defined **Hodge star operator**

$$\star : \bigwedge^{N} \mathsf{T}^*_m \mathcal{M} \longrightarrow \bigwedge^{(D-N)} \mathsf{T}^*_m \mathcal{M}$$

(see Appendix E.4 on page 202).

We can extend this to a Hodge star operator

$$\star : \bigwedge^{N} \mathsf{T}^* \mathcal{M} \longrightarrow \bigwedge^{(D-N)} \mathsf{T}^* \mathcal{M}$$

which is a linear isomorphism from each fibre of the tensor bundle $\bigwedge^{N} \mathsf{T}^* \mathcal{M}$ to the corresponding fibre of $\bigwedge^{(D-N)} \mathsf{T}^* \mathcal{M}$.

Proposition 142: *The Hodge star operator $\star$ defines a canonical isomorphism between the vector bundles $\bigwedge^N \mathsf{T}^*\mathcal{M}$ and $\bigwedge^{(D-N)} \mathsf{T}^*\mathcal{M}$.*

□

Next, we can extend the Hodge star to operate on any differential N-form on the manifold. This defines a canonical $\mathcal{C}^\infty(\mathcal{M})$-linear map:

$$\star : \mathfrak{S} \bigwedge^N \mathsf{T}^*\mathcal{M} \longrightarrow \mathfrak{S} \bigwedge^{(D-N)} \mathsf{T}^*\mathcal{M}$$

Proposition 143: *The Hodge star operator $\star$ defines a canonical isomorphism between the $\mathcal{C}^\infty(\mathcal{M})$-modules bundles $\mathfrak{S}\bigwedge^N \mathsf{T}^*\mathcal{M}$ and $\mathfrak{S}\bigwedge^{(D-N)} \mathsf{T}^*\mathcal{M}$.*

□

G.5 Divergence, Gradient, and Laplacian

Prerequisites:

- Hodge Duality [G.4 on the preceding page]
- Exterior derivatives [124 on page 214]

Definition 144: *Gradient covector Field*

If $\mathcal{M}$ is an arbitrary smooth manifold, and $f \in \mathcal{C}^\infty(\mathcal{M})$, then we can define the **gradient** *of f to be the covector field*

$$\nabla^* f \in \mathfrak{S}\mathsf{T}^*\mathcal{M}$$

so that, for any vector field $\vec{\mathbf{V}} \in \mathfrak{S}\mathsf{T}\mathcal{M}$,

$$(\nabla^* f)(\vec{\mathbf{V}}) \;=\; \vec{\mathbf{V}} f$$

where we interpret $\vec{\mathbf{V}}$ as a **differentiation** *operator acting on f.*

If we interpret f as a differential 0-form on $\mathcal{M}$, then it turns out that

$$\nabla^* f \;=\; \mathfrak{d} f$$

where $\mathfrak{d}$ represents exterior-differentiation of f as a differential form.

Definition 145: *Gradient Vector Field*

If $\mathcal{M}$ is a Riemannian manifold, there is a canonical isomorphism $\mathsf{T}_m\,\mathcal{M} \cong \mathsf{T}_m^*\,\mathcal{M}$, given by the Riesz Representation theorem. Thus, the gradient covector field in this case can be transformed into a **gradient vector field**.

Formally, for any $f \in \mathcal{C}^\infty(\mathcal{M})$, define $\nabla f \in \mathfrak{S}\mathsf{T}\mathcal{M}$ to be the unique vector field such that, for any other vector field $\vec{\mathbf{V}} \in \mathfrak{S}\mathsf{T}\mathcal{M}$,

$$\left\langle \nabla f,\ \vec{\mathbf{V}} \right\rangle \ =\ (\nabla^* f)(\vec{\mathbf{V}})$$

Definition 146: *Divergence of a covector field*

If $\mathcal{M}$ is any manifold, and $\vec{\mathbf{V}}^* \in \mathfrak{S}\mathsf{T}^*\,\mathcal{M}$ is a smooth covector field, then define the **divergence** *of $\vec{\mathbf{V}}^*$:*

$$\mathbf{div}\,\vec{\mathbf{V}}^* \ =\ \star\mathfrak{d}\star\,\vec{\mathbf{V}}^*$$

Thus, $\mathbf{div}\,\vec{\mathbf{V}}^*$ is a zero-form —ie. another element of $\mathcal{C}^\infty(\mathcal{M})$.

Definition 147: *Divergence of a vector field*

If $\mathcal{M}$ is a Riemannian manifold, and $\vec{\mathbf{V}} \in \mathfrak{S}\mathsf{T}\mathcal{M}$ is a smooth vector field, then we can identify $\vec{\mathbf{V}}$ with a smooth covector field $\vec{\mathbf{V}}^* \in \mathfrak{S}\mathsf{T}^*\,\mathcal{M}$, and then define the **divergence** *of $\vec{\mathbf{V}}$:*

$$\mathbf{div}\,\vec{\mathbf{V}} \ =\ \mathbf{div}\,(\vec{\mathbf{V}}^*)$$

Intuitively, the Laplacian is supposed to measure the rate at which the vector field $\vec{\mathbf{V}}$ is "spreading" at each point in $\mathcal{M}$. The vector field $\vec{\mathbf{V}}$ defines a flow on $\mathcal{M}$, and this flow transports the canonical volume form $\mathbf{\Omega}$ on $\mathcal{M}$. Thus, the rate at which $\vec{\mathbf{V}}$ is "spreading" should be related to the rate at which the canonical volume form is being "inflated" or "compressed" by the flow of $\vec{\mathbf{V}}$. This is the content of the next theorem.

Proposition 148: *Let $\vec{\mathbf{V}}$ be a smooth vector field on $\mathcal{M}$, and let $\mathbf{\Omega}$ be the canonical volume form. Then*

$$\mathbf{div}\,\vec{\mathbf{V}} \ =\ \mathcal{L}_{\vec{\mathbf{V}}}\mathbf{\Omega}.$$

□

Definition 149: *Laplacian, Laplace-Beltrami Operator*

Let $f \in \mathcal{C}^\infty(\mathcal{M})$, and treat f as a zero-form. The **Laplacian** *of f is defined*

$$\triangle f \;=\; \mathbf{div}\,\nabla^* f \;=\; \star\mathfrak{d}\star\mathfrak{d} f$$

Thus, $\triangle f$ is another zero-form —ie. another element of $\mathcal{C}^\infty(\mathcal{M})$.

(In a similar fashion, we can define the Laplacian of any differential N-form ω to be

$$\triangle\omega \;=\; \star\mathfrak{d}\star\mathfrak{d}\omega \;-\; \mathfrak{d}\star\mathfrak{d}\star\omega$$

If ω is a zero-form, the second term on the right is trivial, so this agrees with the earlier definition of $\triangle\omega$. In this context, $\triangle$ is sometimes called the **Laplace-Beltrami Operator**.*)*

For the concrete formulae of Laplacians on spheres of various dimensions, see Example 39 on page 62.

Proposition 150: *Self-Adjointness of the Laplacian*
Let $\mathcal{M}$ be a smooth manifold without boundary.
If $f, g \in \mathcal{C}^\infty(\mathcal{M})$, then

$$\int_{\mathcal{M}} (\triangle f)\cdot g \; d\mathcal{L}^{bsg} \;=\; \int_{\mathcal{M}} f\cdot\triangle g \; d\mathcal{L}^{bsg}$$

Proof:

$$\begin{aligned}
\int_{\mathcal{M}} f\cdot\triangle g \; d\mathcal{L}^{bsg} \;&=\; \int_{\mathcal{M}} f\cdot(\star\mathfrak{d}\star\mathfrak{d}g)\;\; d\mathcal{L}^{bsg} \\
&=_{(1)}\; \int_{\mathcal{M}} f\cdot(\mathfrak{d}\star\mathfrak{d}g) \\
&=_{(2)}\; (-1)\cdot\int_{\mathcal{M}} (\mathfrak{d}f)\cdot(\star\mathfrak{d}g) \\
&=_{(3)}\; (-1)\cdot\int_{\mathcal{M}} (\star\mathfrak{d}f)\cdot(\mathfrak{d}g) \\
&=_{(4)}\; \int_{\mathcal{M}} (\mathfrak{d}\star\mathfrak{d}f)\cdot g \\
&=_{(5)}\; \int_{\mathcal{M}} (\star\mathfrak{d}\star\mathfrak{d}f)\cdot g \; d\mathcal{L}^{bsg} \\
&=\; \int_{\mathcal{M}} (\triangle f\cdot)\cdot g \; d\mathcal{L}^{bsg}
\end{aligned}$$

(1) Recall that $\mathfrak{d} \star \mathfrak{d} g$ is a linear D-form.

(2), (4) By the Integration by Parts theorem for differential forms (Proposition 127 on page 216), using the fact that $\mathcal{M}$ is a manifold without boundary.

(3) By the self-adjointness of the Hodge operator (Proposition 108 on page 204)

(5) Again, $\mathfrak{d} \star \mathfrak{d} f$ is a linear D-form.

□

Appendix H

Representations of Lie Groups

H.1 Preliminaries

Definition 151: *Linear Representation*

Let $\mathbb{V}$ *be a Banach space*

Let $\mathbb{G}$ *be a topological group. A* **(linear) representation** *of* $\mathbb{G}$ *is a continuous group homomorphism:*

$$\phi : \mathbb{G} \longrightarrow \mathbb{GL}\ [\mathbb{V}]$$

where $\mathbb{GL}\ [\mathbb{V}]$ *is the set of bounded linear maps on* $\mathbb{V}$*. The* **dimension** *of the representation is the dimension of* $\mathbb{V}$*:*

$$\dim[\phi]\ =\ \dim[\mathbb{V}].$$

If $g \in \mathbb{G}$ *and* $\mathbf{v} \in \mathbb{V}$*, then the action of* g *on* $\mathbf{v}$ *via* ϕ *will sometimes be indicated by "*$g\,{}_\phi\,\mathbf{v}$*".*

Definition 152: *Invariant Subspace, Reducible, Irreducible*

Let $\mathbb{V}$ *be a Banach space. If* $\phi : \mathbb{G} \longrightarrow \mathbb{GL}\ [\mathbb{V}]$ *is a linear representation, an* **invariant subspace** *of* ϕ *is a closed linear subspace* $\mathbb{W}$ *so that, for all* $g \in \mathbb{G}$*,*

$$g\,{}_\phi\,[\mathbb{W}] \subset \mathbb{W}$$

The representation ϕ is **reducible** *if there is a nontrivial invariant subspaces $\mathbb{W} \subset \mathbb{V}$.*

Otherwise, ϕ is called **irreducible**.

If $\mathbb{W} \subset \mathbb{V}$ is an invariant subspace, then for all $g \in G$, the map

$$\phi(g) : \mathbb{V} \longrightarrow \mathbb{V}$$

restricts to a map

$$\phi(g)_| : \mathbb{W} \longrightarrow \mathbb{W}$$

hence, the representation $\phi : \mathbb{G} \longrightarrow \mathbb{GL}\ [\mathbb{V}]$ induces a **subrepresentation** *$\phi_{\mathbb{W}} : \mathbb{G} \longrightarrow \mathbb{GL}\ [\mathbb{W}]$.*

One consequence of this irreducibility is

Theorem 153: *(Schur's Lemma)*
Let $\mathbb{V}$ be a complex Banach space. If $\mathbb{V}$ is an irreducible $\mathbb{G}$-module, and $f : \mathbb{V} \longrightarrow \mathbb{V}$ is a continuous, complex-linear map that commutes with the $\mathbb{G}$-action, then f is multiplication by some scalar.

Proof: Since $\mathbb{C}$ is algebraically complete, the characteristic polynomial of f has a root —ie. f has some eigenvalue $\lambda \in \mathbb{C}$. Let $\vec{v} \in \mathbb{V}$ be a corresponding eigenvalue: $f(\vec{v}) = \lambda.\vec{v}$. Now, let $\mathbb{G}_\phi\, \vec{v} = \{g_\phi\, \vec{v}\,;\ g \in \mathbb{G}\}$. Since $\mathbf{clspan}[\mathbb{G}_\phi\, \vec{v}]$ is a nontrivial, closed, $\mathbb{G}$-invariant subspace, we conclude that $\mathbf{clspan}[\mathbb{G}_\phi\, \vec{v}] = \mathbb{V}$. Thus, any $\vec{w} \in \mathbb{V}$ can be written as a $\mathbb{C}$-linear combination:

$$\vec{w} = \sum_{g \in \mathbb{G}_0} w_g g_\phi\, \vec{v}$$

where $\mathbb{G}_0 \subset \mathbb{G}$ is some suitably chosen countable subset, and $w_g \in \mathbb{C}$, $\forall g \in \mathbb{G}_0$. But then

$$\begin{aligned} f(\vec{w}) &= f\left(\sum_{g \in \mathbb{G}_0} w_g g_\phi\, \vec{v}\right) \\ &= \sum_{g \in \mathbb{G}_0} w_g f\,(g_\phi\, \vec{v}) \;=\; \sum_{g \in \mathbb{G}_0} w_g g_\phi\, f\,(\vec{v}) \\ &= \sum_{g \in \mathbb{G}_0} w_g g_\phi\, \lambda.\vec{v} \;=\; \lambda.\left(\sum_{g \in \mathbb{G}_0} w_g g_\phi\, \vec{v}\right) \\ &= \lambda.\vec{w} \end{aligned}$$

□

Definition 154: *Unitary Representation*

Let $\mathbb{U}^D = \{f \in \mathbb{GL}\ [\mathbb{C}^D]\ ;\ f\ \textit{unitary}\}$.

Let $\mathbb{G}$ *be a topological group. A* **unitary representation** *of* $\mathbb{G}$ *is a continuous group homomorphism:*

$$\phi : \mathbb{G} \longrightarrow \mathbb{U}^D$$

where $D = \dim[\phi]$.

Example 155:

If $\mathbb{G}$ is an abelian group, then all irreducible unitary representations are 1-dimensional; in other words, they are homomorphisms

$$\phi : \mathbb{G} \longrightarrow \mathbb{U}^1 \ = \ \mathbb{T}^1$$

Normally, these are called the **characters** of the group. (In the more general context of nonabelian representation theory, these would properly be called the **irreducible unitary characters**). ———————

Notation 156: The set of all **irreducible unitary representations** of $\mathbb{G}$ will be denoted by $\mathfrak{U}[\mathbb{G}]$.

Proposition 157: *Suppose* $\mathbb{G}$ *is compact, and* $\mathbb{V}$ *is any finite dimensional complex vector space, and* $\phi : \mathbb{G} \longrightarrow \mathbb{GL}\ [\mathbb{V}]$ *is any linear representation of* $\mathbb{G}$ *on* $\mathbb{V}$. *There is an inner product* $\langle \bullet, \bullet \rangle$ *on* $\mathbb{V}$ *so that* ϕ *is a* **unitary representation** *relative to* $\langle \bullet, \bullet \rangle$.

Proof: Let $\mathcal{H}^{oar}_{\mathbb{G}}$ be the Haar measure on $\mathbb{G}$. Since $\mathbb{G}$ is compact, $\mathcal{H}^{oar}_{\mathbb{G}}$ is finite —assume it has total mass 1. Furthermore, the image $\phi(\mathbb{G})$ is some compact subgroup of $\mathbb{GL}\ [\mathbb{V}]$; thus, for any $\mathbf{v} \in \mathbb{V}$, the orbit of $\mathbf{v}$ under $\phi(\mathbb{G})$ is bounded.

Let $[\bullet, \bullet]$ be an arbitrary inner product on $\mathbb{V}$, and define the inner product $\langle \bullet, \bullet \rangle$ as follows:

$$\langle \mathbf{v}, \mathbf{w} \rangle \ = \ \int_{\mathbb{G}} [\phi(g)(\mathbf{v}),\ \phi(g)(\mathbf{w})]\ d\mathcal{H}^{oar}_{\mathbb{G}}[g]$$

Since all elements in $\mathbb{V}$ have bounded orbits, this integral converges. The resulting function, $\langle \bullet, \bullet \rangle$, is also an inner product, and is invariant under

the action of $\mathbb{G}$. Thus, if we consider $\mathbb{V}$ to be an inner-product space relative to $\langle \bullet, \bullet \rangle$ then all elements of $\mathbb{G}$ act in a unitary fashion on $\mathbb{V}$ —in other words, ϕ has become a unitary representation. ——————□

Remark 158: If $\Phi : \mathbb{G} \longrightarrow \mathbb{U}^D$ is a unitary representation, and $\mathbb{W} \subset \mathbb{C}^D$ is an invariant subspace, then so is $\mathbb{W}^\perp$, and we can write

$$\mathbb{C} \;=\; \mathbb{W} \oplus \mathbb{W}^\perp.$$

This is indicative of a general pattern.

Proposition 159: *If $\mathbb{V}$ is a finite-dimensional vector space, and $\phi : \mathbb{G} \longrightarrow \mathbb{GL}\ [\mathbb{V}]$ is a a reducible representation, then we can write:*

$$\mathbb{V} = \mathbb{V}_1 \oplus \mathbb{V}_2 \oplus \ldots \oplus \mathbb{V}_N$$

where $\mathbb{V}_1, \ldots, \mathbb{V}_N$ are invariant subspaces, and the action of ϕ on each $\mathbb{V}_n$ is irreducible.

Proof: Let $\langle \bullet, \bullet \rangle$ be a $\mathbb{G}$-invariant inner product on $\mathbb{V}$. Thus, if $\mathbb{V}_1 \subset \mathbb{V}$ is a nontrivial invariant subspace, then so is $\mathbb{V}_1^\perp$. If $\mathbb{V}_1$ and $\mathbb{V}_1^\perp$ are irreducible, we are done. Otherwise, proceed inductively. Since $\mathbb{V}$ is finite dimensional, this induction terminates after finitely many steps. ——————□

H.2 Noncommutative Harmonic Analysis

If $\mathbb{G}$ is a compact abelian group, then classical harmonic analysis tells us that the (irreducible unitary) characters of $\mathbb{G}$ form an orthonormal basis for $\mathbf{L}^2(\mathbb{G})$. We now seek to generalize this result to the nonabelian setting....

Definition 160: *Coefficient Function*

Suppose $\phi : \mathbb{G} \longrightarrow \mathbb{U}^D$ is an irreducible unitary representation, and let $\mathcal{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_D\}$ be the standard orthonormal basis for $\mathbb{C}^D$. Thus, for all $g \in \mathbb{G}$, the unitary transformation $\phi(g)$ has a matrix relative to $\mathcal{E}$, say:

$$\begin{bmatrix} \phi_{1,1}(g) & \ldots & \phi_{1,N}(g) \\ \vdots & \ddots & \vdots \\ \phi_{N,1}(g) & \ldots & \phi_{N,N}(g) \end{bmatrix}$$

For any such representation, and for all $i, j \in [1...D]$, the functions $\phi_{i,j} : \mathbb{G} \longrightarrow \mathbb{C}$ are called **coefficient functions**.

The coefficient functions are measurable and bounded, and therefore, are elements of $\mathbf{L}^2(\mathbb{G},\ \mathcal{H}^{aar}_{\mathbb{G}})$.

If $\mathbb{G}$ is abelian, then all irreducible unitary representations are one-dimensional, and the coefficient functions are just the classical characters. Hence, the following results are somewhat expected.

Theorem 161: *Orthogonality relations*
Let $\mathbb{G}$ be a compact group, and let $\phi, \psi : \mathbb{G} \longrightarrow \mathbb{U}^D$ be irreducible unitary representations.

1. *For any i, j, k, ℓ,*

$$\langle \phi_{i,j},\ \psi_{k,\ell} \rangle \ = \ \begin{cases} \frac{1}{\dim \phi} & \text{if} \quad \phi = \psi,\ i = k,\ \text{and}\ j = \ell \\ 0 & \text{otherwise} \end{cases}$$

2. *Thus, for i, j, $\|\phi_{i,j}\|_2 \ = \ \frac{1}{\sqrt{\dim \phi}}$*

Proof: See [172] p. 79 □

Theorem 162: *Peter-Weyl Theorem*
Let

$$\hat{\mathbb{G}} \ = \ \left\{ \sqrt{\dim[\phi]} \cdot \phi_{i,j} \ ;\ \phi \ \text{ any irreducible unitary representation}, 1 \leq i, j \leq \dim[\phi] \right\}.$$

Then

1. *$\hat{\mathbb{G}}$ is an* **orthonormal basis** *for $\mathbf{L}^2[\mathbb{G},\ \mathcal{H}^{aar}_{\mathbb{G}}]$,*

2. *$\hat{\mathbb{G}}$ is a linear basis for the Banach space $\mathcal{C}(\mathbb{G},\ \mathbb{C})$.*

Proof: See [172] p. 133 □

Definition 163: *Fourier Coefficients*

If $f \in \mathbf{L}^2[\mathbb{G},\ \mathcal{H}^{oar}_{\mathbb{G}}]$, *then the Peter-Weyl theorem says that*

$$f \ = \ \sum_{\chi \in \widehat{\mathbb{G}}} \widehat{f}_\chi \cdot \chi$$

where $\widehat{f}_\chi \ = \ \langle f,\ \chi \rangle$. *The coefficients* $\left\{ \widehat{f}_\chi \ ; \ \chi \in \widehat{\mathbb{G}} \right\}$ *are the* **(scalar) Fourier coefficients** *of* f.

We can group these coefficients together into matrices; for every $\phi \in \mathfrak{U}[\mathbb{G}]$, *define* $\widehat{f}_\phi$ *to be the matrix:*

$$\widehat{f}_\phi \ = \ \begin{bmatrix} \widehat{f}_{\phi_{1,1}} & \cdots & \widehat{f}_{\phi_{1,N}} \\ \vdots & \ddots & \vdots \\ \widehat{f}_{\phi_{N,1}} & \cdots & \widehat{f}_{\phi_{N,N}} \end{bmatrix}$$

where, for all $i, j \in [1..N]$,

$$\widehat{f}_{\phi_{i,j}} \ = \ \sqrt{N} \cdot \langle f,\ \phi_{ij} \rangle \, .$$

The coefficients $\left\{ \widehat{f}_\phi \ ; \ \phi \in \mathfrak{U}[\mathbb{G}] \right\}$, *are the* **matrix-valued Fourier coefficients** *of* f. *Notice that another way to define* $\widehat{f}_\phi$ *is as an integral-linear combination of matrices:*

$$\widehat{f}_\phi \ = \ \sqrt{N} \cdot \int_{\mathbb{G}} f(\mathbf{g}) \cdot \Phi(\mathbf{g}) \ d\mathcal{H}^{oar}_{\mathbb{G}}[\mathbf{g}]$$

Where $\Phi(\mathbf{g}) \in \mathbb{C}^{N \times N}$ *is the* $N \times N$ *matrix corresponding to the unitary transformation* $\phi(\mathbf{g})$.

Example 164:

In classical harmonic analysis of abelian groups, **convolution** of functions on the group corresponds to **pointwise multiplication** of their Fourier coefficients. There is an analogous result in the nonabelian case, except that now the Fourier coefficients are matrix-valued.

Definition 165: *Convolution*

If $\mathbb{G}$ is a compact group, and $\phi, \eta \in \mathbf{L}^2(\mathbb{G}, \mathcal{H}_{\mathbb{G}}^{aar})$, then we define $\phi * \eta : \mathbb{G} \longrightarrow \mathbb{C}$ by:

$$\phi * \eta(g) \;=\; \int_{\mathbb{G}} \phi(g \cdot h^{-1}) \eta(h) \; d\mathcal{H}_{\mathbb{G}}^{aar}[h]$$

In Chapter 6, we used a rather bizare definition (Definition 61 on page 84) of convolution. The two definitions are equivalent:

Lemma 166: *With $\mathbb{G}$, ϕ, and η as before,*

$$\phi * \eta(g) \;=\; \int_{\mathbb{G}} \eta(h^{-1} \cdot g) \phi(h) \; d\mathcal{H}_{\mathbb{G}}^{aar}[h]$$

Proof:

$$\begin{aligned}
\int_{\mathbb{G}} \eta(h^{-1} \cdot g) \phi(h) \; d\mathcal{H}_{\mathbb{G}}^{aar}[h] &= \int_{\mathbb{G}} \phi(g \cdot g^{-1} \cdot h) \eta(h^{-1} \cdot g) \; d\mathcal{H}_{\mathbb{G}}^{aar}[h] \\
&= \int_{\mathbb{G}} \phi(g \cdot (h^{-1} \cdot g)^{-1}) \eta(h^{-1} \cdot g) \; d\mathcal{H}_{\mathbb{G}}^{aar}[h] \\
&=_{(1)} \int_{\mathbb{G}} \phi(g \cdot k^{-1}) \eta(k) \; d\mathcal{H}_{\mathbb{G}}^{aar}[h] \\
&= \phi * \eta(g)
\end{aligned}$$

(1) Making the change of variables $k = h^{-1}g$, since the Haar measure is invariant under translation and inversion. □

Theorem 167: *Convolution of Coefficient Functions*

Let $\mathbb{G}$ be a compact group, and let $\phi, \psi : \mathbb{G} \longrightarrow \mathbb{U}$ be irreducible unitary representations. For any i, j, k, ℓ,

$$\phi_{i,j} * \psi_{k,\ell} \;=\; \begin{cases} \frac{1}{\dim \phi} \phi_{i,\ell} & \text{if} \quad \phi = \psi, \text{ and } j = k, \\ 0 & \text{otherwise} \end{cases} .$$

Proof: See [172] p. 83. □

Corollary 168: *If $f, g \in \mathbf{L}^2(\mathbb{G})$ have matrix-valued Fourier Coefficients $\left\{\widehat{f}_\phi|_{\phi\in\mathfrak{U}[\mathbb{G}]}\right\}$ and $\left\{\widehat{g}_\phi|_{\phi\in\mathfrak{U}[\mathbb{G}]}\right\}$, respectively, then the function $f * g$ has matrix-valued Fourier coefficients $\left\{\widehat{f}_\phi \cdot \widehat{g}_\phi|_{\phi\in\mathfrak{U}[\mathbb{G}]}\right\}$.*

Definition 169: *Fourier Coefficients of a Measure*

If Γ is a measure on $\mathbb{G}$, then we can define the **(scalar) Fourier coefficients** *of Γ as follows: for every $N \in \mathbb{N}$, $\chi \in \widehat{\mathbb{G}}$,*

$$\widehat{\Gamma}_\chi \;=\; \int_{\mathbb{G}} \chi \; d\Gamma$$

The **matrix valued Fourier coefficients** *are defined in the obvious way.*

Remark 170: Suppose that $S_1 \subset S_2 \subset \ldots \subset \mathfrak{U}[\mathbb{G}]$ is a sequence of finite subsets whose union is all of $\mathfrak{U}[\mathbb{G}]$. Then the partial Fourier sums

$$\Gamma_n \;=\; \sum_{\phi\in S_n} \sum_{i,j} \widehat{\Gamma}_{\phi_{i,j}} \cdot \phi_{i,j}$$

are elements of $\mathbf{L}^2(\mathbb{G}, \mathcal{H}_{\mathbb{G}}^{aar}) \cap \mathcal{C}(\mathbb{G}, \mathbb{C})$, and, as linear functionals on $\mathcal{C}(\mathbb{G}, \mathbb{C})$ they converge in the weak* topology to Γ. To see this, it is sufficient to show that, for any $f \in \mathbf{L}^2(\mathbb{G}, \mathcal{H}_{\mathbb{G}}^{aar}) \cap \mathcal{C}(\mathbb{G}, \mathbb{C})$,

$$\langle f, \Gamma_n \rangle \;\xrightarrow[n\to\infty]{}\; \int_{\mathbb{G}} f \; d\Gamma$$

Through a similar weak* convergence argument, it is possible to show:

Proposition 171: *If $\eta \in \mathbf{L}^2(\mathbb{G})$ and Γ is a measure on $\mathbb{G}$, having matrix-valued Fourier coefficients $\left\{\widehat{\eta}_\phi|_{\phi\in\mathfrak{U}[\mathbb{G}]}\right\}$ and $\left\{\widehat{\Gamma}_\phi|_{\phi\in\mathfrak{U}[\mathbb{G}]}\right\}$, respectively, then the function $\Gamma * \eta$ has matrix-valued Fourier coefficients $\left\{\widehat{\Gamma}_\phi \cdot \widehat{\eta}_\phi|_{\phi\in\mathfrak{U}[\mathbb{G}]}\right\}$.*

Corollary 172: *Suppose* $\mathbf{g}$, $\eta \in \mathbf{L}^2(\mathbb{G})$ *are functions with (known) matrix-valued Fourier transforms* $\left\{\widehat{f}_{\phi}|_{\phi \in \mathfrak{U}[\mathbb{G}]}\right\}$ *and* $\left\{\widehat{\eta}_{\phi}|_{\phi \in \mathfrak{U}[\mathbb{G}]}\right\}$, *respectively, and* Γ *is a measure on* $\mathbb{G}$, *having (unknown) matrix-valued Fourier coefficients* $\left\{\widehat{\Gamma}_{\phi}|_{\phi \in \mathfrak{U}[\mathbb{G}]}\right\}$, *and suppose furthermore that*

$$\mathbf{g} \;=\; \Gamma * \eta$$

Suppose that, for some $\phi \in \mathfrak{U}[\mathbb{G}]$, $\widehat{\eta}_{\phi}$ *is an* **invertible** *matrix. Then we can recover the* ϕ*th Fourier coefficient of* Γ *through the formula:*

$$\widehat{\Gamma}_{\phi} \;=\; \widehat{g}_{\phi} \cdot (\widehat{\eta}_{\phi})^{-1}.$$

H.3 Tensor Products of Group Representations

[See Appendix E.1 on page 199 for background on tensor products]

Suppose that $\mathbb{G}$ is a group, and $\phi_1 : \mathbb{G} \longrightarrow \mathbf{Aut}[\mathbb{V}_1]$ and $\phi_2 : \mathbb{G} \longrightarrow \mathbf{Aut}[\mathbb{V}_2]$ are two group representations. We can define the **tensor product** of these two representations as the map $\phi_1 \otimes \phi_2 : \mathbb{G} \longrightarrow \mathbf{Aut}[\mathbb{V}_1 \otimes \mathbb{V}_2]$, where, for all $g \in \mathbb{G}$, $\mathbf{v}_1 \in \mathbb{V}_1$ and $\mathbf{v}_2 \in \mathbb{V}_2$,

$$(\phi_1 \otimes \phi_2)[g](\mathbf{v}_1 \otimes \mathbf{v}_2) \quad = \quad (g\,{}_{\phi_1}\,\mathbf{v}_1) \otimes (g\,{}_{\phi_2}\,\mathbf{v}_2)$$

Now suppose $\mathbb{V}$ is an inner product space with orthonormal basis $\mathcal{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_D\}$, and we canonically identify elements of $\mathbb{V}^{\otimes N}$ with N-dimensional arrays of numbers in the manner described in Section E.1 on page 199. If $\phi_1, \ldots, \phi_N : \mathbb{G} \longrightarrow \mathbf{Aut}[\mathbb{V}]$ are N representations of $\mathbb{G}$ on $\mathbb{V}$, we would like to understand the action of $\phi = \phi_1 \otimes \ldots \otimes \phi_N$ in terms of these arrays. Let us adopt the same notation as in section E.1 on page 199.

Proposition 173: *Let* $g \in \mathbb{G}$. *Suppose that, relative to the basis* $\mathcal{E}$, *the transformation* $\phi_n[g]$ *has matrix* $\mathbf{g}^{[n]} = \left[g^{[n]}_{i,j}|^{D}_{i,j=1}\right]$. *(Thus, for any* $d \in [1..D]$, $g\,{}_{\phi_n}\,\mathbf{e}_d = \sum_{j=1}^{D} g^{[n]}_{j,d}\mathbf{e}_j$.*) Then, for any* $\mathbf{d} \in \mathbb{D}$,

$$g\,{}_{\phi}\,\mathbf{e}_{\otimes \mathbf{d}} \;=\; \sum_{\mathbf{j} \in \mathbb{D}} \left(\prod_{n=1}^{N} g^{[n]}_{j_n, d_n} \right) \cdot \mathbf{e}_{\otimes \mathbf{j}}$$

Proof:

$$\begin{aligned} g_{\,\phi}\,\mathbf{e}_{\otimes\mathbf{d}} &= (g_{\,\phi_1}\,\mathbf{e}_{d_1}) \otimes (g_{\,\phi_2}\,\mathbf{e}_{d_2}) \otimes \ldots \otimes (g_{\,\phi_N}\,\mathbf{e}_{d_N}) \\ &= \left(\sum_{j_1=1}^{D} g^{[1]}_{j_1,d_1}\mathbf{e}_{j_1}\right) \otimes \left(\sum_{j_2=1}^{D} g^{[2]}_{j_2,d_2}\mathbf{e}_{j_2}\right) \otimes \ldots \otimes \left(\sum_{j_N=1}^{D} g^{[N]}_{j_N,d_N}\mathbf{e}_{j_N}\right) \\ &= \sum_{\mathbf{j}\in\mathbb{D}} \left(g^{[1]}_{j_1,d_1} \cdot g^{[2]}_{j_2,d_2} \cdot \ldots \cdot g^{[N]}_{j_N,d_N}\right) \mathbf{e}_{j_1} \otimes \mathbf{e}_{j_2} \otimes \ldots \otimes \mathbf{e}_{j_N} \\ &= \sum_{\mathbf{j}\in\mathbb{D}} \left(\prod_{n=1}^{N} g^{[n]}_{j_n,d_n}\right) \cdot \mathbf{e}_{\otimes\mathbf{j}} \end{aligned}$$

□

Corollary 174: *If $\phi_1(g), \phi_2(g), \ldots, \phi_N(g)$ are orthogonal (or unitary) transformations on $\mathbb{V}$, then $\phi(g)$ is an orthogonal (resp. unitary) transformation on $\mathbb{V}^{\otimes N}$, relative to the canonical inner product on this space.*

Proof: The elements $\{\mathbf{e}_{\otimes\mathbf{d}}\ ;\ \mathbf{d}\in\mathbb{D}\}$ form an orthonormal basis for $\mathbb{V}^{\otimes N}$; it suffices to show that these elements remain orthonormal under transformation by $\phi(g)$.

Retaining the notation of the previous theorem, we have:

$$g_{\,\phi}\,\mathbf{e}_{\otimes\mathbf{d}} = \sum_{\mathbf{j}\in\mathbb{D}} \left(\prod_{n=1}^{N} g^{[n]}_{j_n,d_n}\right) \cdot \mathbf{e}_{\otimes\mathbf{j}}$$

Thus, for any $\mathbf{d}$ and $\widetilde{\mathbf{d}}$ in $\mathbb{D}$,

$$\begin{aligned} &\left\langle g_{\,\phi}\,\mathbf{e}_{\otimes\mathbf{d}},\ \ g_{\,\phi}\,\mathbf{e}_{\otimes\widetilde{\mathbf{d}}}\right\rangle \\ &\quad= \sum_{\mathbf{j}\in\mathbb{D}}\sum_{\widetilde{\mathbf{j}}\in\mathbb{D}} \left(\prod_{n=1}^{N} g^{[n]}_{j_n,d_n}\right) \cdot \left(\prod_{n=1}^{N} g^{[n]}_{\widetilde{j}_n,\widetilde{d}_n}\right) \cdot \left\langle \mathbf{e}_{\otimes\mathbf{j}},\ \mathbf{e}_{\otimes\widetilde{\mathbf{j}}}\right\rangle \\ &=_{(1)}\ \sum_{\mathbf{j}\in\mathbb{D}} \left(\prod_{n=1}^{N} g^{[n]}_{j_n,d_n}\right) \cdot \left(\prod_{n=1}^{N} g^{[n]}_{j_n,\widetilde{d}_n}\right) \\ &\quad= \sum_{\mathbf{j}\in\mathbb{D}} \left(\prod_{n=1}^{N} g^{[n]}_{j_n,d_n} \cdot g^{[n]}_{j_n,\widetilde{d}_n}\right) \end{aligned}$$

$$= \sum_{j_1=1}^{D}\sum_{j_2=1}^{D}\cdots\sum_{j_N=1}^{D}\left[\left(g^{[1]}_{j_1,d_1}\cdot g^{[1]}_{j_1,\widetilde{d}_1}\right)\cdot\left(g^{[2]}_{j_2,d_2}\cdot g^{[2]}_{j_2,\widetilde{d}_2}\right)\cdot\ldots\cdot\left(g^{[N]}_{j_N,d_N}\cdot g^{[N]}_{j_N,\widetilde{d}_N}\right)\right]$$

$$= \left[\sum_{j_1=1}^{D} g^{[1]}_{j_1,d_1}\cdot g^{[1]}_{j_1,\widetilde{d}_1}\right]\cdot\left[\sum_{j_2=1}^{D} g^{[2]}_{j_2,d_2}\cdot g^{[2]}_{j_2,\widetilde{d}_2}\right]\cdot\ldots\cdot\left[\sum_{j_N=1}^{D} g^{[N]}_{j_N,d_N}\cdot g^{[N]}_{j_N,\widetilde{d}_N}\right]$$

$$=_{(2)} \left\langle\begin{bmatrix}\uparrow\\ g^{[1]}_{d_1}\\ \downarrow\end{bmatrix},\begin{bmatrix}\uparrow\\ g^{[1]}_{\widetilde{d}_1}\\ \downarrow\end{bmatrix}\right\rangle\cdot\left\langle\begin{bmatrix}\uparrow\\ g^{[2]}_{d_1}\\ \downarrow\end{bmatrix},\begin{bmatrix}\uparrow\\ g^{[2]}_{\widetilde{d}_2}\\ \downarrow\end{bmatrix}\right\rangle\cdot\ldots\cdot\left\langle\begin{bmatrix}\uparrow\\ g^{[N]}_{d_N}\\ \downarrow\end{bmatrix},\begin{bmatrix}\uparrow\\ g^{[N]}_{\widetilde{d}_N}\\ \downarrow\end{bmatrix}\right\rangle$$

(1) Because $\left\langle \mathbf{e}_{\otimes\mathbf{j}}, \mathbf{e}_{\otimes\widetilde{\mathbf{j}}}\right\rangle = \begin{cases} 1 & if \quad \mathbf{j}=\widetilde{\mathbf{j}} \\ 0 & \text{otherwise}\end{cases}$.

(2) where $\begin{bmatrix}\uparrow\\ g^{[n]}_{d}\\ \downarrow\end{bmatrix}$ is the dth column vector of the matrix of $\phi^{[n]}[g]$.

But for any fixed $n \in [1...N]$, the column vectors of $\phi^{[n]}[g]$ are orthogonal. Thus,

$$\left\langle\begin{bmatrix}\uparrow\\ g^{[n]}_{d_n}\\ \downarrow\end{bmatrix},\begin{bmatrix}\uparrow\\ g^{[n]}_{\widetilde{d}_n}\\ \downarrow\end{bmatrix}\right\rangle = \begin{cases} 1 & if \quad d_n=\widetilde{d}_n \\ 0 & \text{otherwise}\end{cases}.$$

Hence, we conclude:

$$\left\langle \phi(g)[\mathbf{e}_{\otimes\mathbf{d}}],\ \ \phi(g)[\mathbf{e}_{\otimes\widetilde{\mathbf{d}}}]\right\rangle = \begin{cases} 1 & if \quad \mathbf{d}=\widetilde{\mathbf{d}} \\ 0 & \text{otherwise}\end{cases}.$$

In other words, the basis vectors remain orthogonal under the tensored action of g. □

H.4 Exterior Products of Group Representations

[See Appendix E.2 on page 201 for background on exterior products]

Now suppose that $\phi : \mathbb{G} \longrightarrow \mathbf{Aut}[\mathbb{V}]$ is a linear representation, and define the linear representation $\phi^{\otimes N} : \mathbb{G} \longrightarrow \mathbf{Aut}\left[\mathbb{V}^{\otimes N}\right]$ in the aforementioned fashion. Then $\bigwedge^{N}\mathbb{V}$ is an **invariant subspace** of $\mathbb{G}$, because

$$\phi^{\otimes N}[g](\mathbf{v}_1 \wedge \ldots \wedge \mathbf{v}_N) = (g \,_\phi\, \mathbf{v}_1) \wedge \ldots \wedge (g \,_\phi\, \mathbf{v}_N)$$

Hence we can consider the subrepresentation

$$\phi^{\wedge N} : \mathbb{G} \longrightarrow \mathbf{Aut}\left[\bigwedge^N \mathbb{V}\right]$$

where $\phi^{\wedge N}[g] = \phi^{\otimes N}[g]\Big|_{\bigwedge^N \mathbb{V}}$.

Suppose that $\mathbb{V}$ is an inner product space with orthonormal basis $\mathcal{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_D\}$, and let $\mathbb{D}^{(<)}$ and $\mathbf{e}_{\wedge\mathbf{d}}$ be defined as in section E.2 on page 201. We seek to describe the action of $\mathbb{G}$ on $\bigwedge^N \mathbb{V}$ in terms of the basis $\left\{\mathbf{e}_{\wedge\mathbf{d}} \; ; \; \mathbf{d} \in \mathbb{D}^{(<)}\right\}$.

Proposition 175: *Suppose that $g \in \mathbb{G}$, and that the transformation $\phi[g]$ has matrix $\mathbf{g} = \left[g_{ij}|_{i,j=1}^{D}\right]$. Then for any $\mathbf{d} \in \mathbb{D}^{(<)}$,*

$$\phi^{\wedge N}[g]\left(\mathbf{e}_{\wedge\mathbf{d}}\right) = \sum_{\mathbf{j}\in\mathbb{D}^{(<)}} \det\left[\mathbf{g}_{[\mathbf{j},\mathbf{d}]}\right] \cdot \mathbf{e}_{\wedge\mathbf{j}},$$

where, for all $\mathbf{j} \in \mathbb{D}$, we define $\mathbf{g}_{[\mathbf{j},\mathbf{d}]}$ to be the $N \times N$ submatrix of $\mathbf{g}$:

$$\mathbf{g}_{[\mathbf{j},\mathbf{d}]} = \begin{bmatrix} g_{i_1,d_1} & g_{i_1,d_2} & g_{i_1,d_3} & \cdots & g_{i_1,d_N} \\ g_{i_2,d_1} & g_{i_2,d_2} & g_{i_2,d_3} & \cdots & g_{i_2,d_N} \\ g_{i_3,d_1} & g_{i_3,d_2} & g_{i_3,d_3} & \cdots & g_{i_3,d_N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ g_{i_N,d_1} & g_{i_N,d_2} & g_{i_N,d_3} & \cdots & g_{i_N,d_N} \end{bmatrix}$$

Proof:

$$\begin{aligned}
&\phi^{\wedge N}[g]\left(\mathbf{e}_{\wedge\mathbf{d}}\right) \\
&= \phi^{\otimes N}[g]\left(\sum_{\sigma\in\boldsymbol{\sigma}(N)} \mathbf{sign}[\sigma] \cdot \mathbf{e}_{d_{\sigma(1)}} \otimes \ldots \otimes \mathbf{e}_{d_{\sigma(N)}}\right) \\
&= \sum_{\sigma\in\boldsymbol{\sigma}(N)} \mathbf{sign}[\sigma]\phi^{\otimes N}[g]\left(\mathbf{e}_{d_{\sigma(1)}} \otimes \ldots \otimes \mathbf{e}_{d_{\sigma(N)}}\right) \\
&=_{(1)} \sum_{\sigma\in\boldsymbol{\sigma}(N)} \mathbf{sign}[\sigma] \sum_{\mathbf{j}\in\mathbb{D}} \left(g_{j_1,d_{\sigma(1)}} \cdot \ldots \cdot g_{j_N,d_{\sigma(N)}}\right) \mathbf{e}_{j_1} \otimes \ldots \otimes \mathbf{e}_{j_N} \\
&= \sum_{\mathbf{j}\in\mathbb{D}} \left(\sum_{\sigma\in\boldsymbol{\sigma}(N)} \mathbf{sign}[\sigma] g_{j_1,d_{\sigma(1)}} \cdot g_{j_2,d_{\sigma(2)}} \cdot \ldots \cdot g_{j_N,d_{\sigma(N)}}\right) \mathbf{e}_{j_1} \otimes \mathbf{e}_{j_2} \otimes \ldots \otimes \mathbf{e}_{j_N}
\end{aligned}$$

$$=_{(2)} \quad \sum_{\mathbf{j}\in\mathbb{D}} \det\left[\mathbf{g}_{[\mathbf{j},\mathbf{d}]}\right] \cdot \mathbf{e}_{j_1} \otimes \mathbf{e}_{j_2} \otimes \ldots \otimes \mathbf{e}_{j_N}$$

(1) By Proposition 173 on page 235.

(2) This follows from the "permutation formula" for computing determinants.

Now, every element in $\mathbb{D}$ can be thought of as a particular element in $\mathbb{D}^{(<)}$, acted on by a particular permutation in $\boldsymbol{\sigma}(N)$. Hence, we can rewrite this expression as:

$$\begin{aligned}
\phi^{\wedge N}[g](\mathbf{e}_{\wedge\mathbf{d}}) &= \sum_{\mathbf{j}\in\mathbb{D}^{(<)}} \sum_{\sigma\in\boldsymbol{\sigma}(N)} \det\left[\mathbf{g}_{[\sigma(\mathbf{j}),\mathbf{d}]}\right] \cdot \mathbf{e}_{j_{\sigma(1)}} \otimes \mathbf{e}_{j_{\sigma(2)}} \otimes \ldots \otimes \mathbf{e}_{j_{\sigma(N)}} \\
&= \sum_{\mathbf{j}\in\mathbb{D}^{(<)}} \sum_{\sigma\in\boldsymbol{\sigma}(N)} \mathbf{sign}[\sigma] \cdot \det\left[\mathbf{g}_{[\mathbf{j},\mathbf{d}]}\right] \cdot \mathbf{e}_{j_{\sigma(1)}} \otimes \mathbf{e}_{j_{\sigma(2)}} \otimes \ldots \otimes \mathbf{e}_{j_{\sigma(N)}} \\
&= \sum_{\mathbf{j}\in\mathbb{D}^{(<)}} \det\left[\mathbf{g}_{[\mathbf{j},\mathbf{d}]}\right] \cdot \sum_{\sigma\in\boldsymbol{\sigma}(N)} \mathbf{sign}[\sigma] \cdot \mathbf{e}_{j_{\sigma(1)}} \otimes \mathbf{e}_{j_{\sigma(2)}} \otimes \ldots \otimes \mathbf{e}_{j_{\sigma(N)}} \\
&= \sum_{\mathbf{j}\in\mathbb{D}^{(<)}} \det\left[\mathbf{g}_{[\mathbf{j},\mathbf{d}]}\right] \cdot \mathbf{e}_{j_1} \wedge \mathbf{e}_{j_2} \wedge \ldots \wedge \mathbf{e}_{j_N}
\end{aligned}$$

In other words:

$$\phi^{\wedge N}[g](\mathbf{e}_{\wedge\mathbf{d}}) = \sum_{\mathbf{j}\in\mathbb{D}^{(<)}} \det\left[\mathbf{g}_{[\mathbf{j},\mathbf{d}]}\right] \cdot \mathbf{e}_{\wedge\mathbf{j}}$$

□

Proposition 176: *If ϕ is an* **orthogonal** *(***unitary***) representation of $\mathbb{G}$ on $\mathbb{V}$, then $\phi^{\wedge N}$ is an orthogonal (unitary) representation of $\mathbb{G}$ on $\bigwedge^N \mathbb{V}$.*

Proof: $\phi^{\wedge N}$ is just the restriction to $\bigwedge^N \mathbb{V}$ of the action of $\mathbb{G}$ on $\mathbb{V}^{\otimes N}$ via the representation $\phi^{\otimes N}$; we have already shown that this action is orthogonal/unitary. □

H.5 The Irreducible Unitary Representations of $\mathbb{SO}^D[\mathbb{R}]$

It is desirable to have explicit descriptions of the **irreducible, unitary representations** of $\mathbb{SO}^D[\mathbb{R}]$, in terms of orthonormal coordinate systems.

$\mathbb{SO}^D[\mathbb{R}]$ acts upon $\mathbb{R}^D$ in an obvious fashion. By "complexifying" (that is, tensoring with $\mathbb{C}$), we can extend this to a complex-linear action of $\mathbb{SO}^D[\mathbb{R}]$ upon $\mathbb{C}^D$. Since the elements of $\mathbb{SO}^D[\mathbb{R}]$ act orthogonally in $\mathbb{R}^D$, they act *unitarily* on $\mathbb{C}^D$, so this is a unitary representation. Since there clearly can be no invariant subspaces, this action is irreducible. Call this action ϕ.

For every $N \in [1..D]$, we can look at the Nth **wedge power** of ϕ with itself: $\phi^{\wedge N}$ is a linear representation of $\mathbb{SO}^D[\mathbb{R}]$ upon $\bigwedge^N \mathbb{C}^D$, defined:

$$\phi^{\wedge N}[g](\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \ldots \mathbf{v}_N) \;=\; \phi[g](\mathbf{v}_1) \wedge \phi[g](\mathbf{v}_2) \wedge \ldots \wedge \phi[g](\mathbf{v}_N) \wedge$$

These representations are also irreducible (see [91], p. 83), and are unitary relative to a naturally defined inner product structure on $\bigwedge^N \mathbb{C}^D$. An explicit description of this action in terms of coordinates is given in Section H.4 on page 237.

For every $N \in \mathbb{N}$, the action of $\mathbb{SO}^D[\mathbb{R}]$ on $\mathbb{C}^D$ induces a natural right-action on the space of **harmonic homogeneous polynomials** of degree N with complex coefficients, defined:

$$(p.g)(x) \;=\; p(g(x))$$

and these representations are also irreducible (see [172] p.88 for case $D = 3$, or [91], p. 81).

Appendix I

Plots of Gegenbauer Polynomials, for $D = 5, 10, 20$

As observed in Proposition 44 on page 68, the zonal eigenfunctions of the Laplacian on a sphere are the **Gegenbauer Polynomials.** The explicit expressions for these polynomials are complicated, and their qualitative properties are most easily observed graphically. Below are plots of the Gegenbauer polynomials $C_N^{(\nu)}(x_1)$, for $N \in [1..39]$ and $x_1 \in [-1, 1]$, in to in dimensions $D = 5$, 10, and 20.

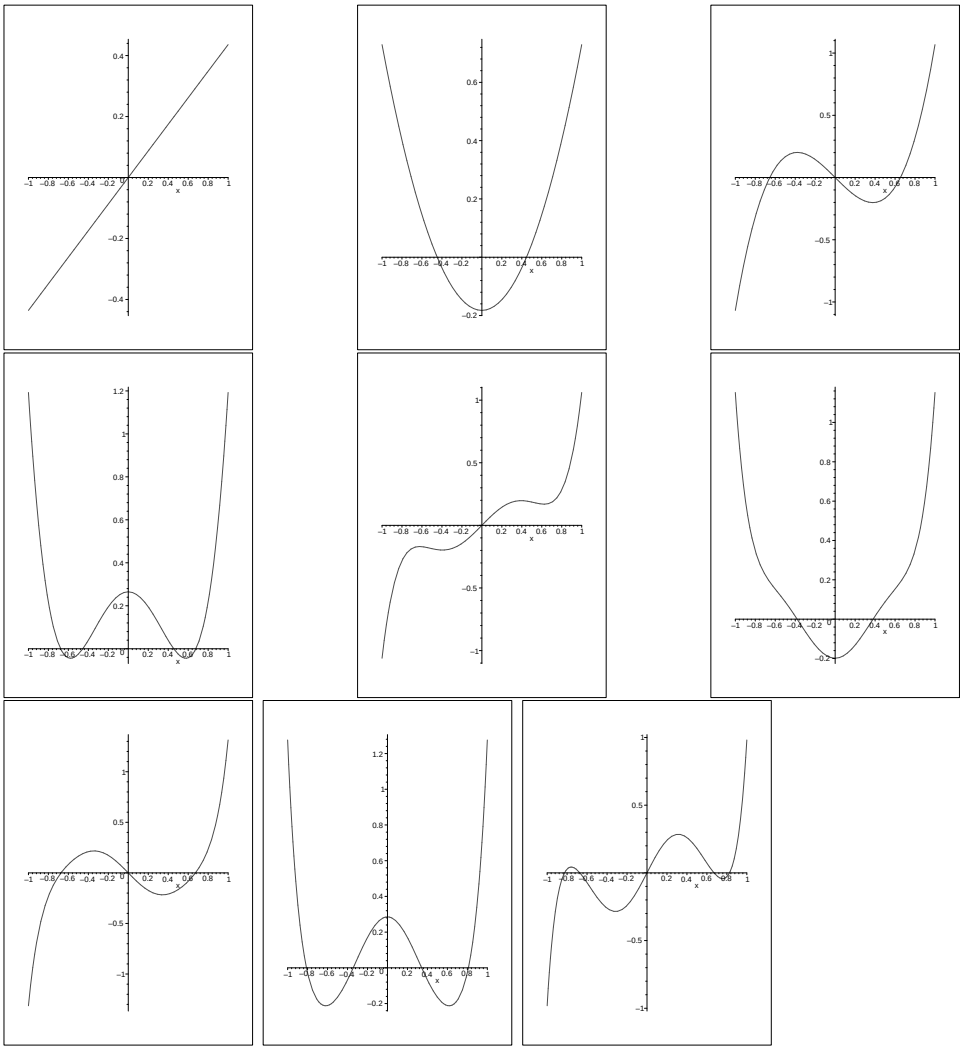

Figure I.1: Gegenbauer Polynomial, $D = 5$, $N = 1..9$

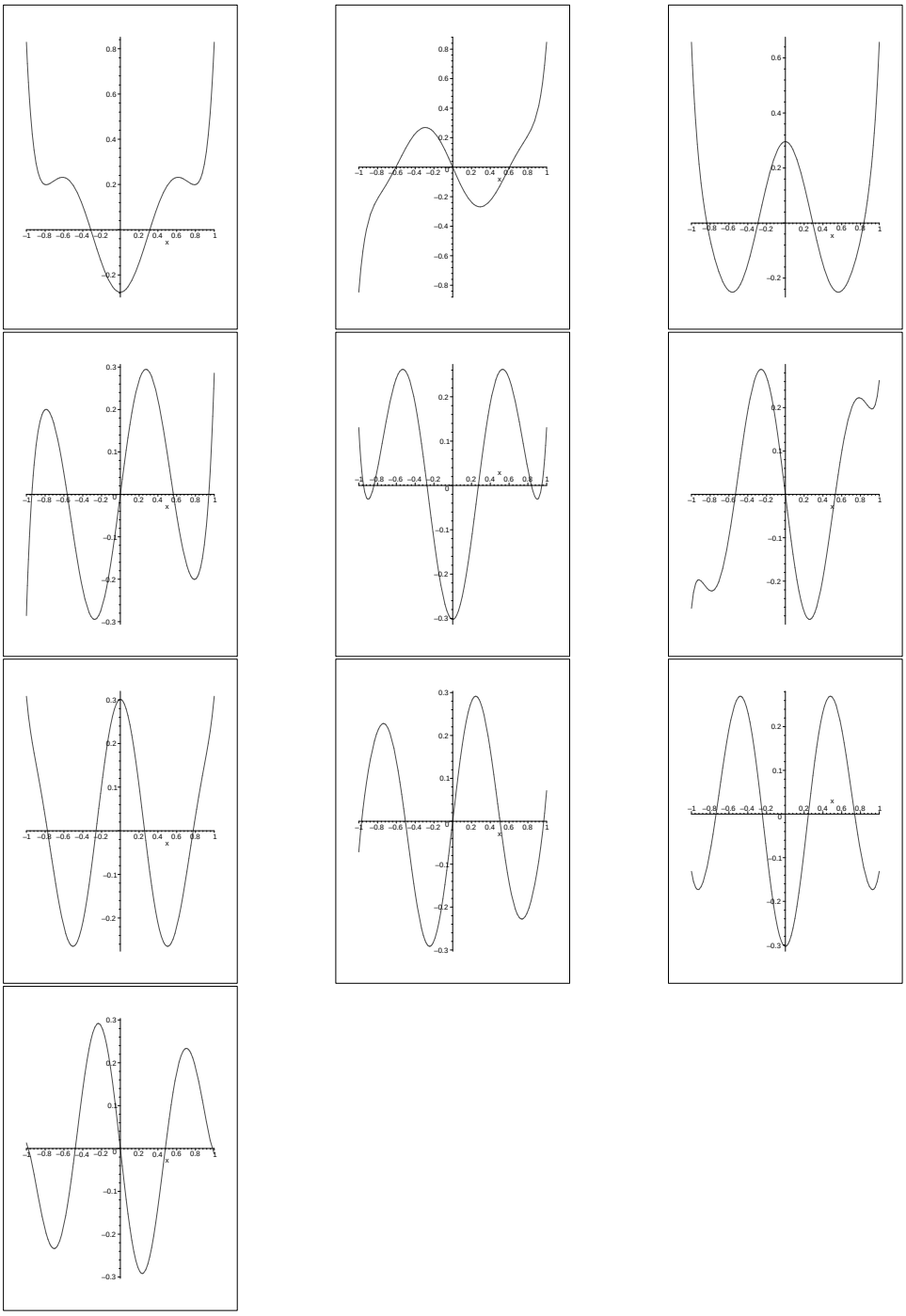

Figure I.2: Gegenbauer Polynomial, $D = 5$, $N = 10..19$

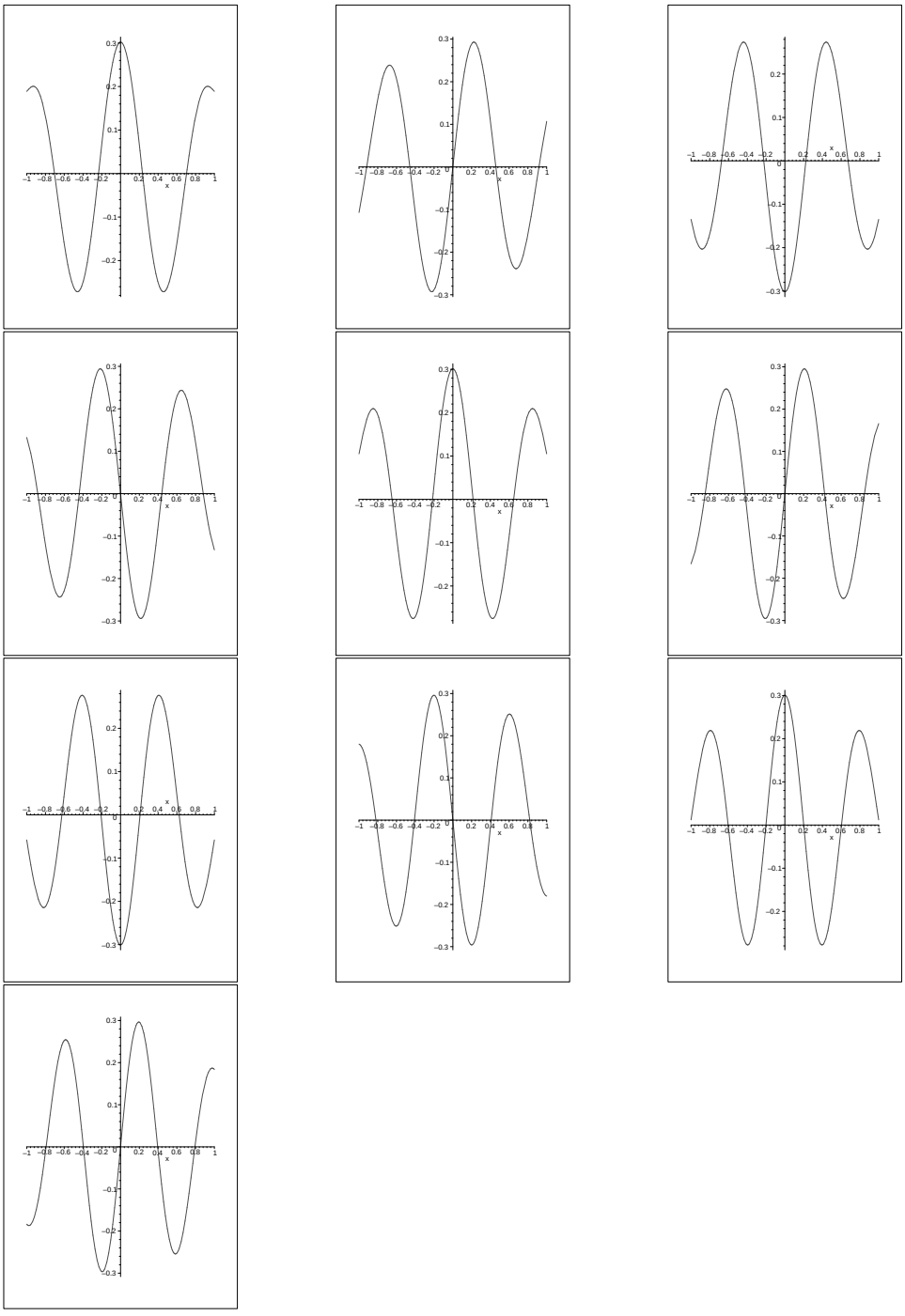

Figure I.3: Gegenbauer Polynomial, $D = 5$, $N = 20..29$

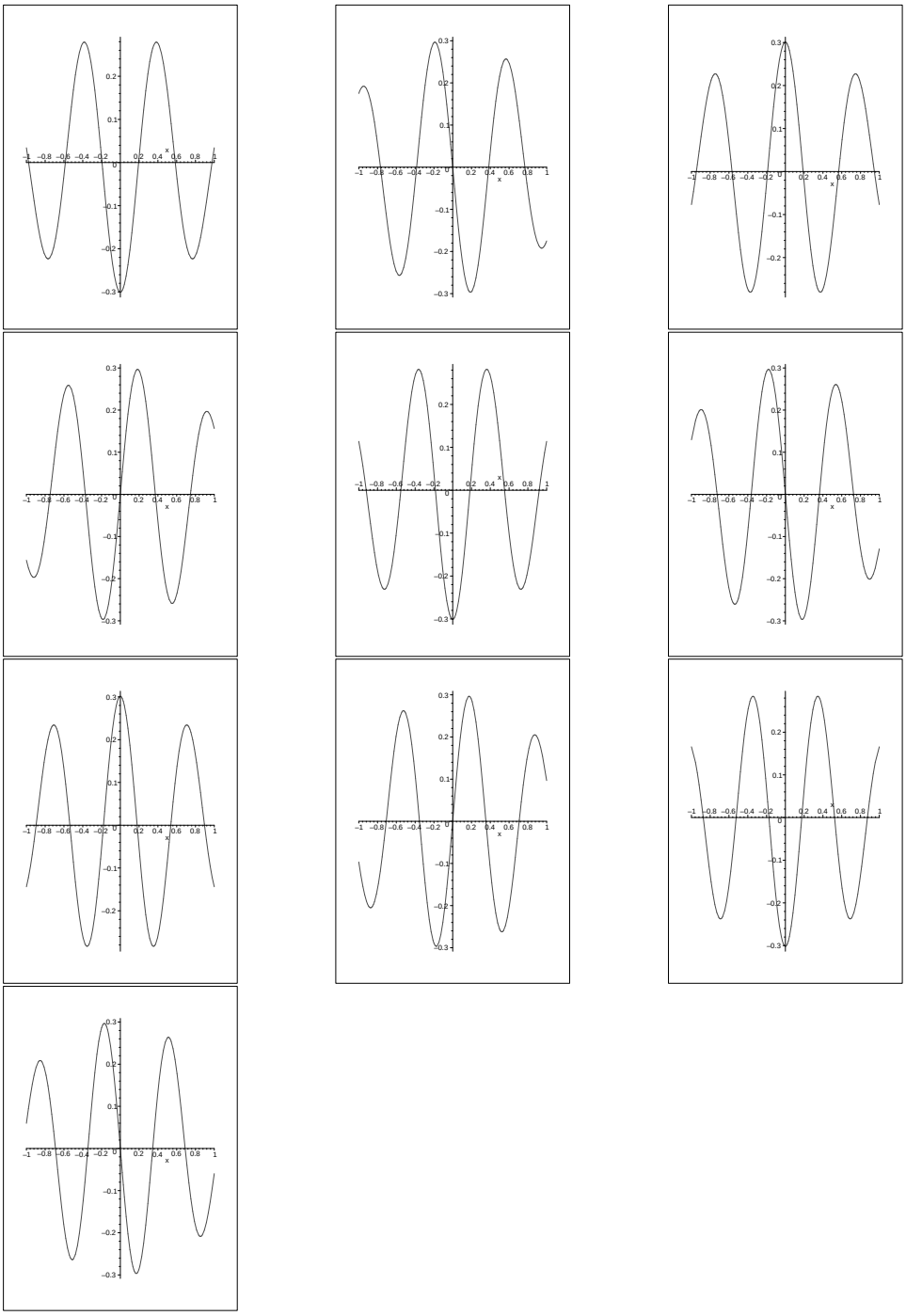

Figure I.4: Gegenbauer Polynomial, $D = 5$, $N = 30..39$

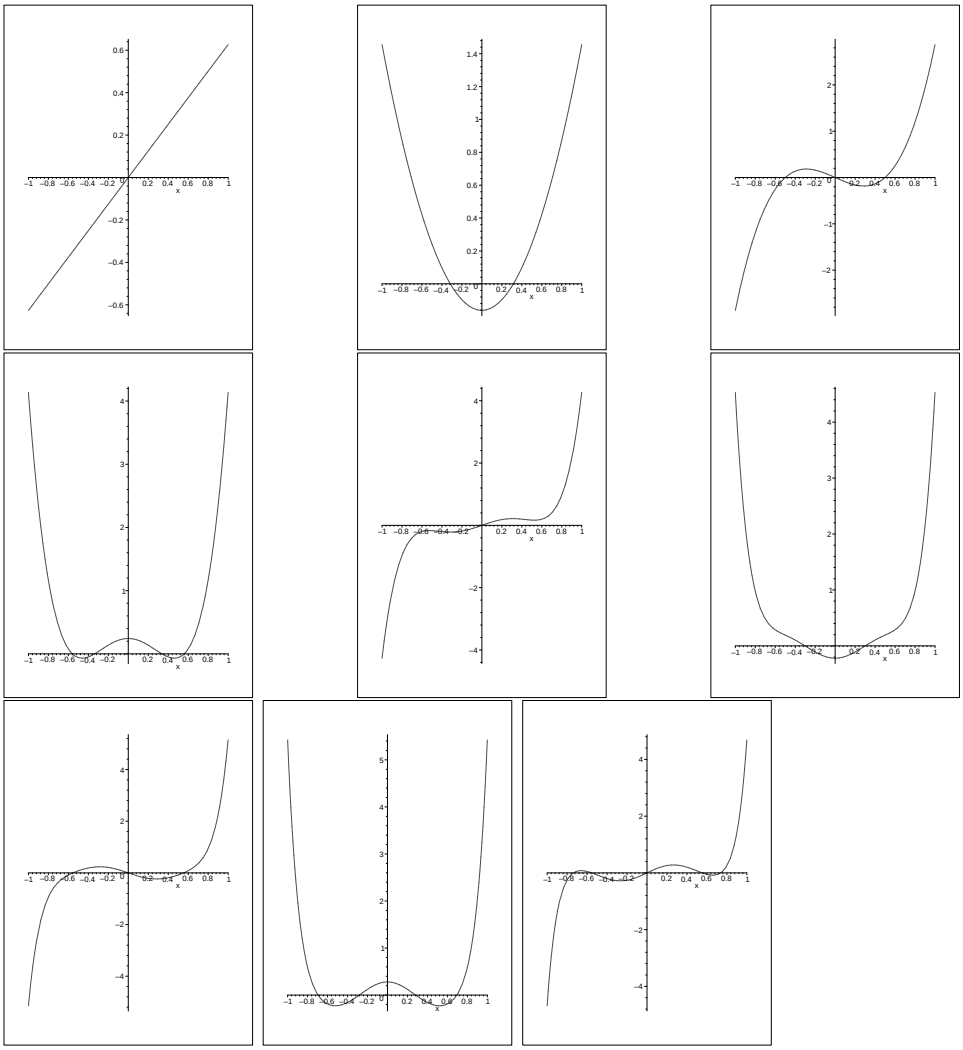

Figure I.5: Gegenbauer Polynomial, $D = 10$, $N = 1..9$

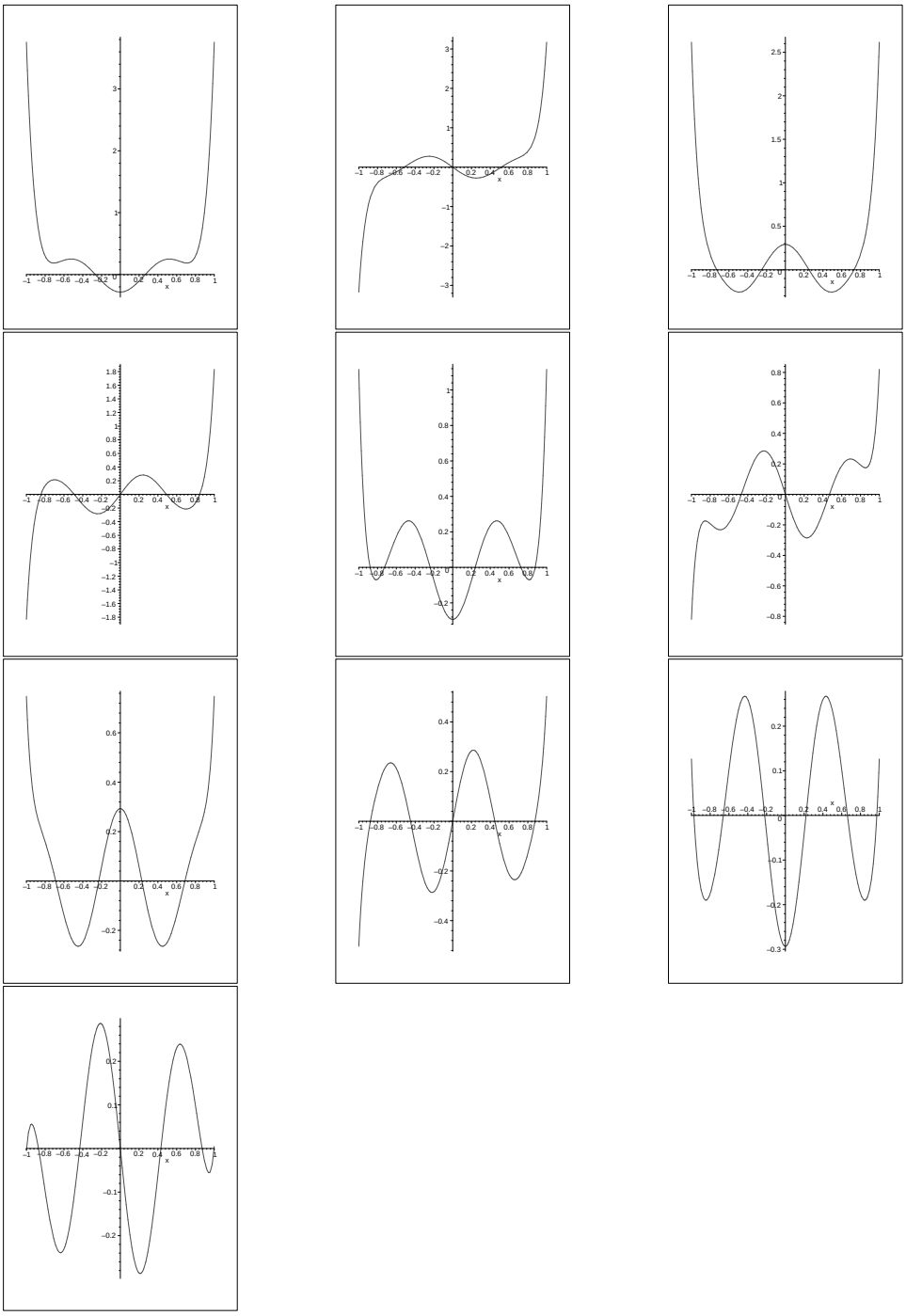

Figure I.6: Gegenbauer Polynomial, $D = 10$, $N = 10..19$

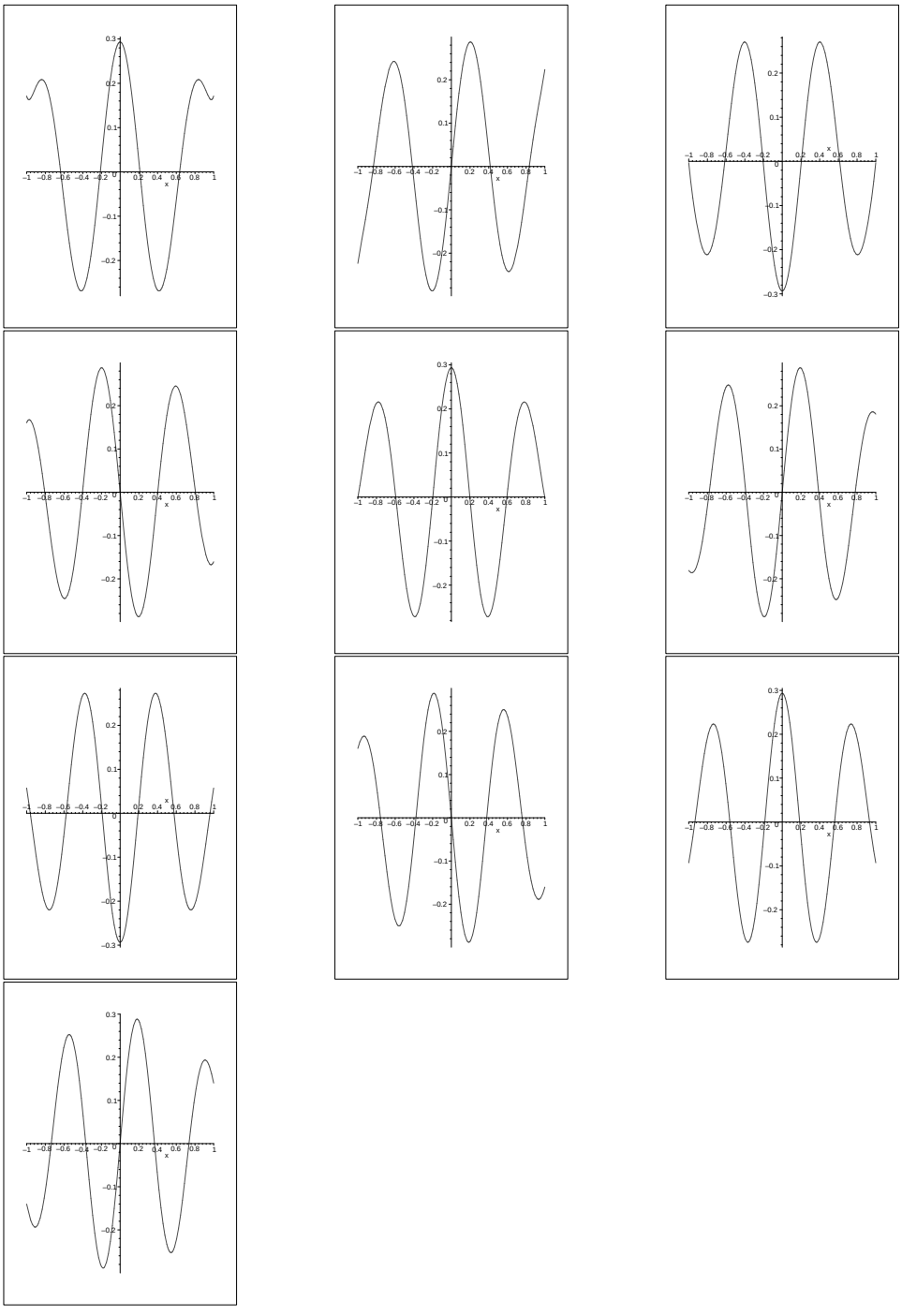

Figure I.7: Gegenbauer Polynomial, $D = 10$, $N = 20..29$

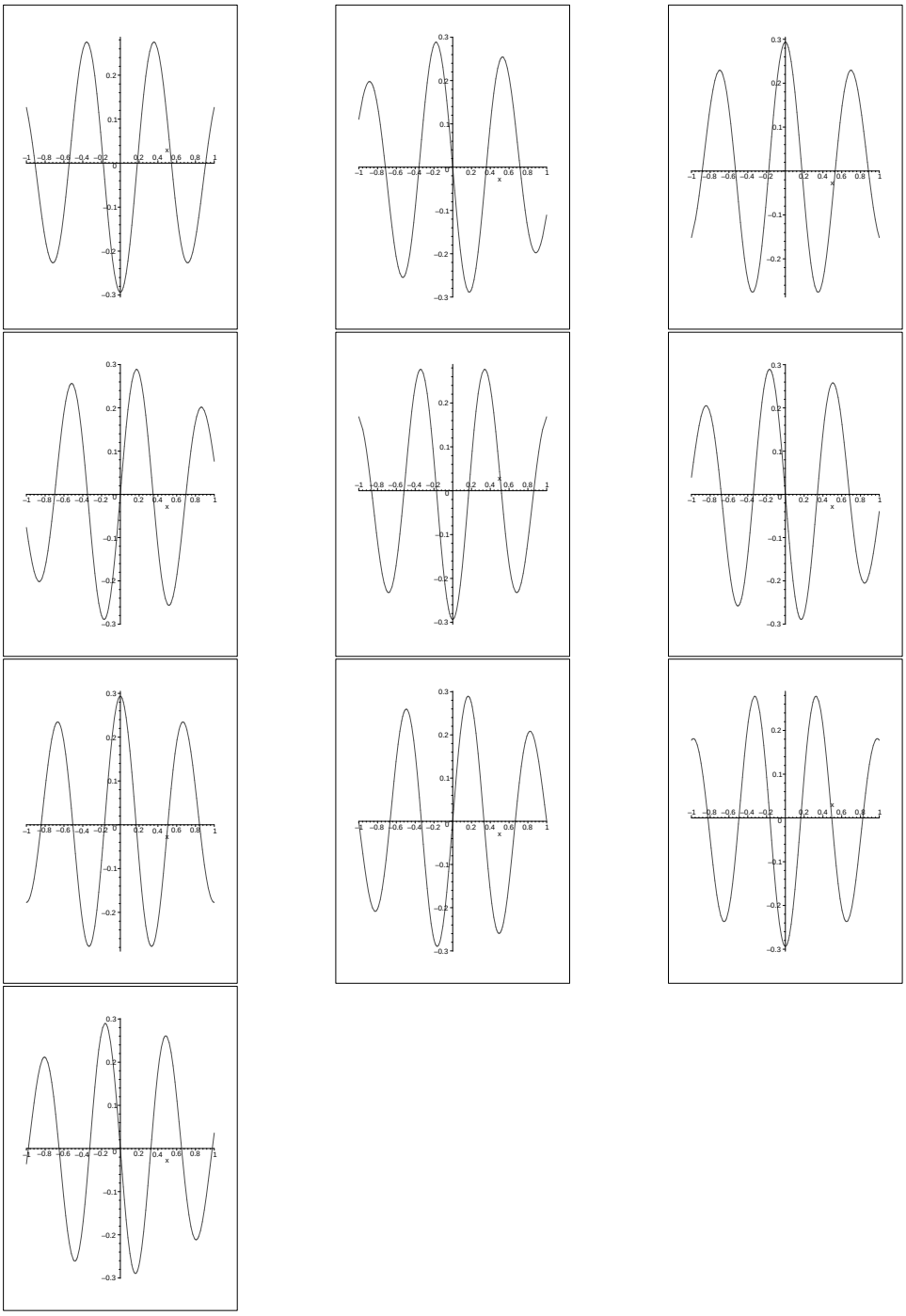

Figure I.8: Gegenbauer Polynomial, $D = 10$, $N = 30..39$

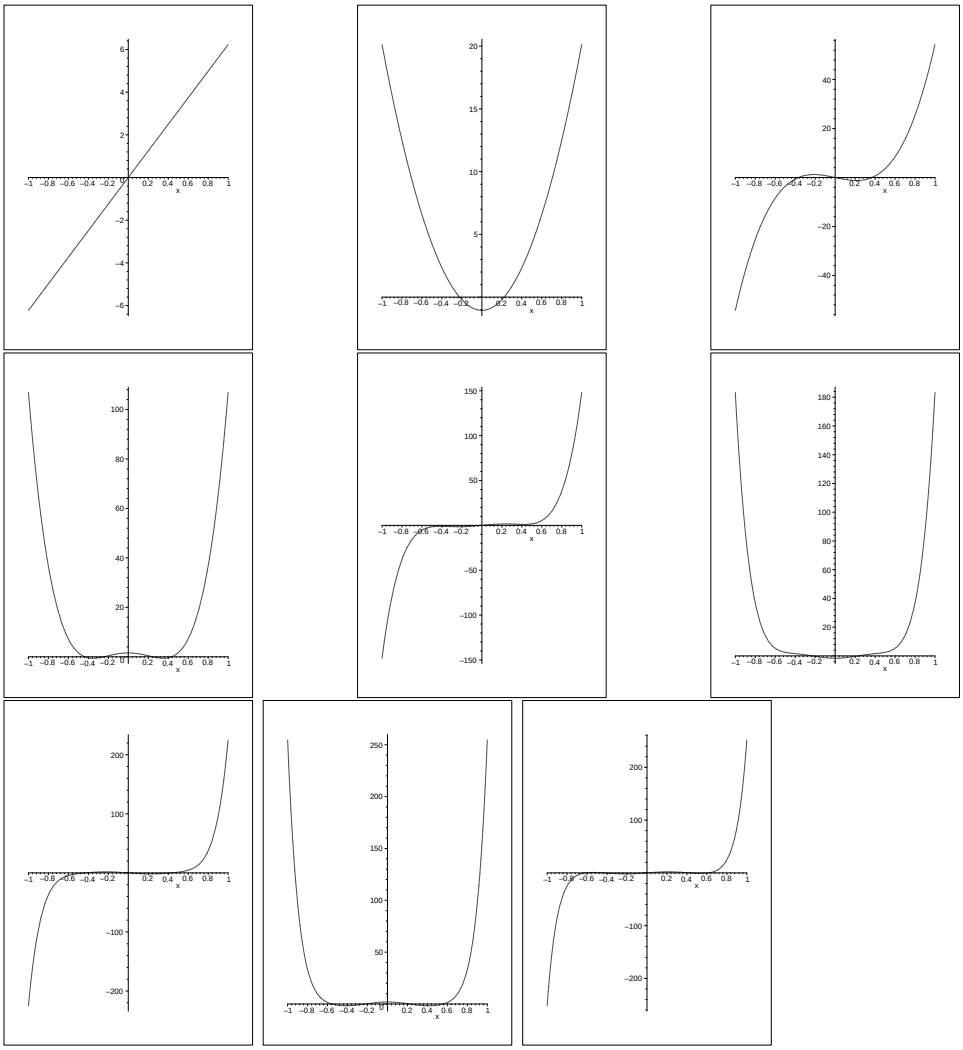

Figure I.9: Gegenbauer Polynomial, $D = 20$, $N = 1..9$

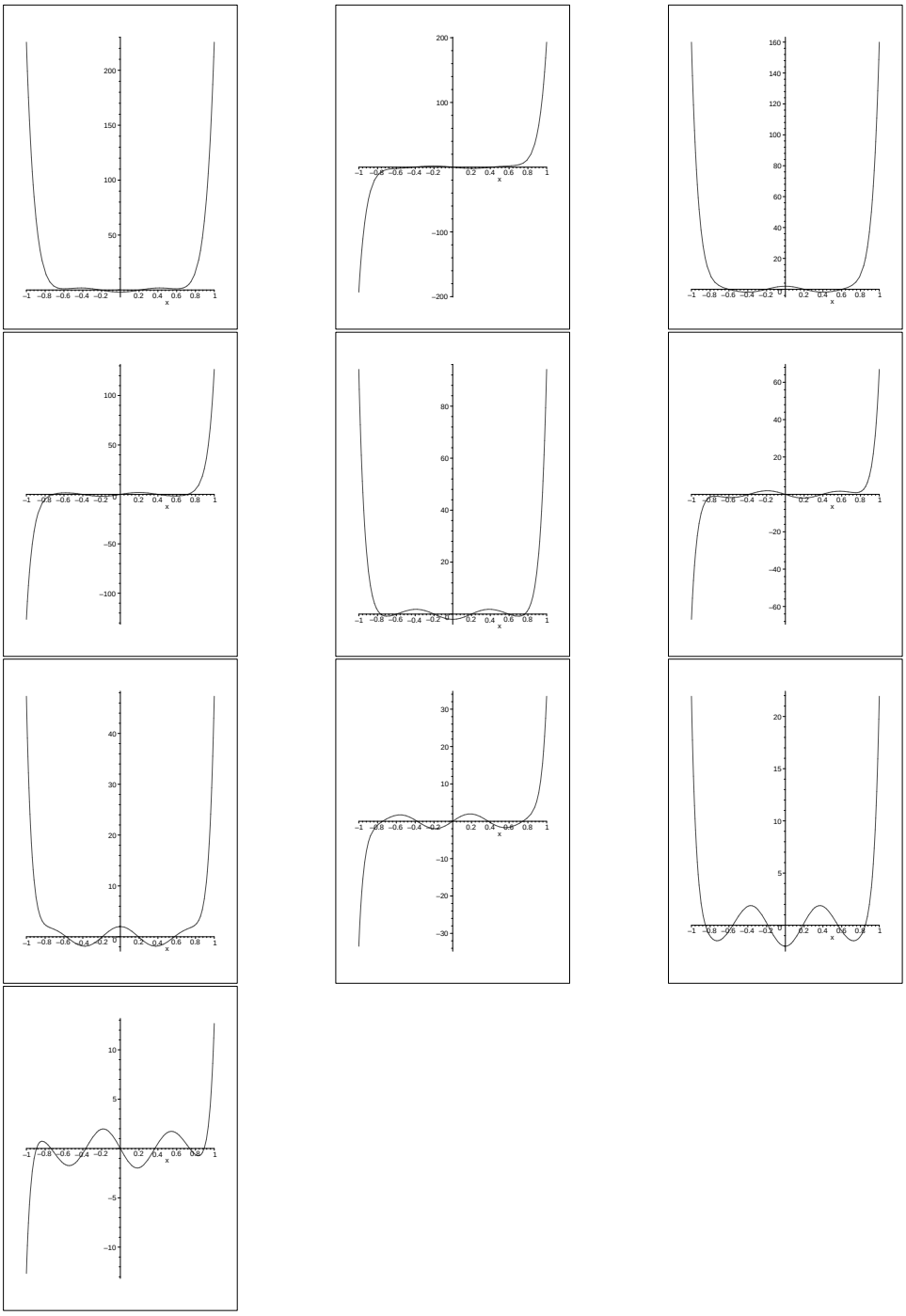

Figure I.10: Gegenbauer Polynomial, $D = 20$, $N = 10..19$

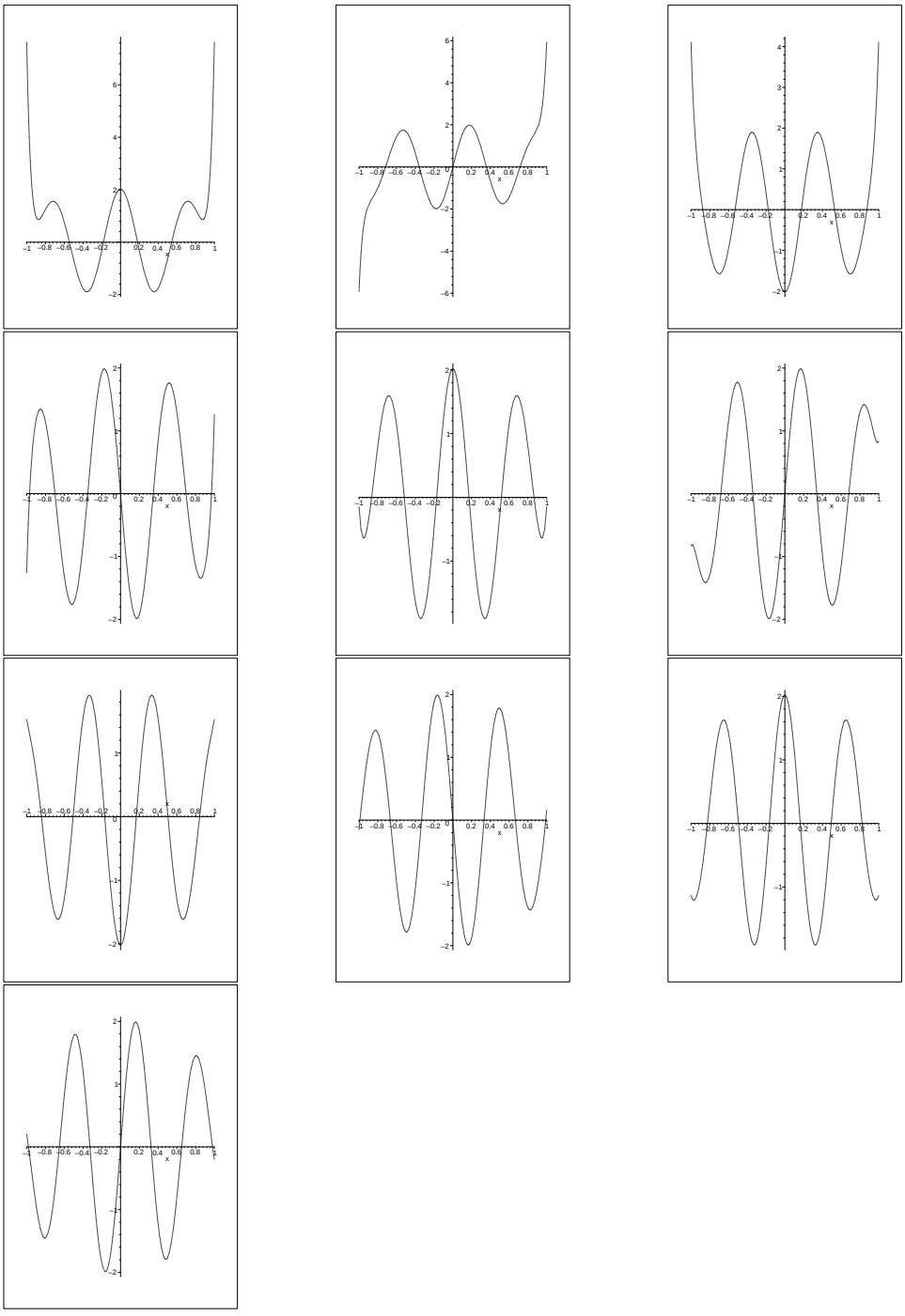

Figure I.11: Gegenbauer Polynomial, $D = 20$, $N = 20..29$

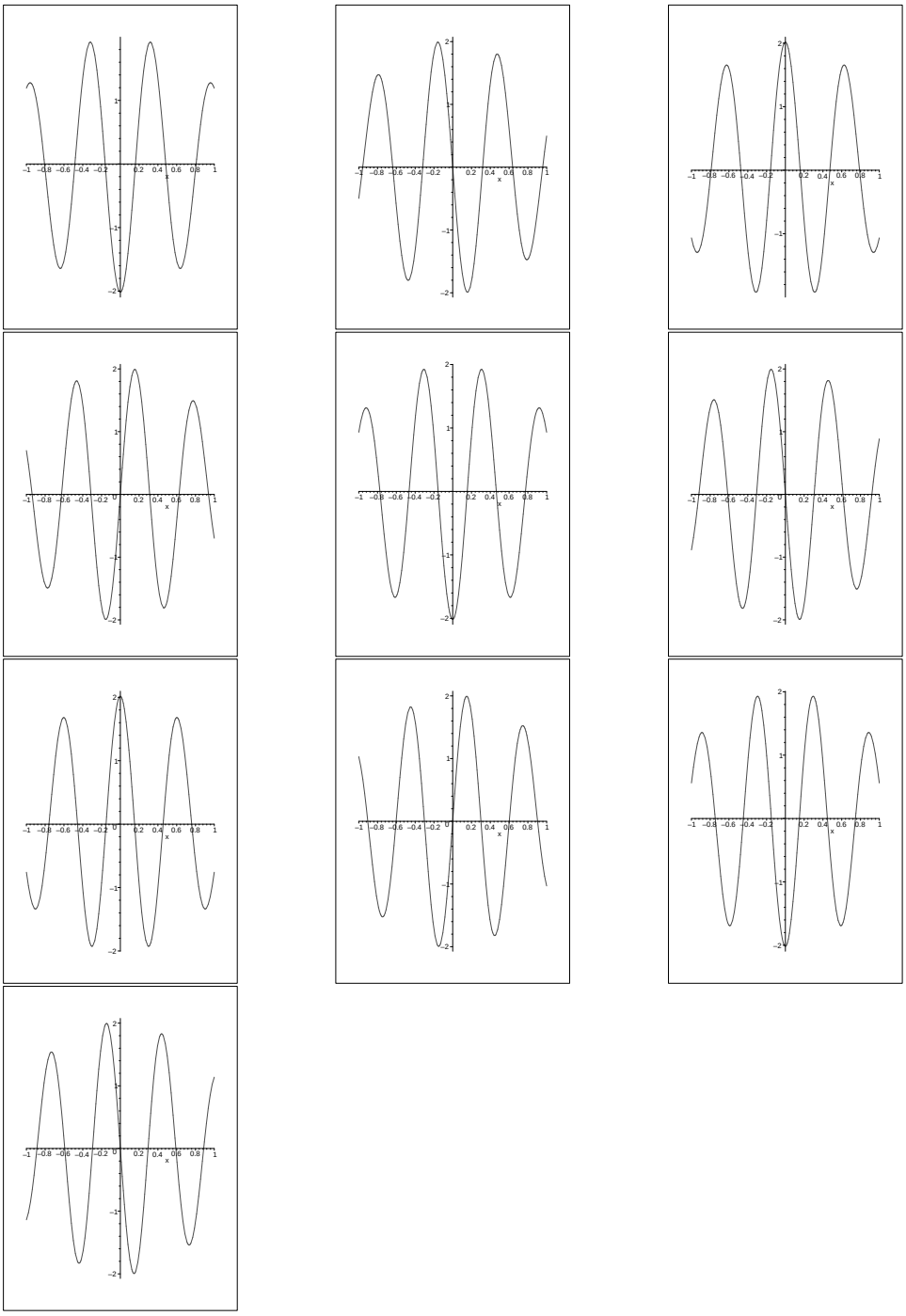

Figure I.12: Gegenbauer Polynomial, $D = 20$, $N = 30..39$

Bibliography

[1] A. Araujo and E. Giné. *The Central Limit Theorem for Real and Banach Valued Random Variables.* Wiley, New York, 1980.

[2] A. Erdelyi *et al. Higher Transcendental Functions*, volume I, II, III. McGraw-Hill, New York, 1953-1955.

[3] A. Janicki and A. Weron. *Simulation and Chaotic Behaviour of α-stable Stochastic Processes.* Marcel Dekker, New York, 1994.

[4] Husein Abdul-Hamid. *Density inequalities for multivariate stable distributions.* PhD thesis, American University, 1996.

[5] Richard Arratia A.D. Barbour and Simon Tavaré. Random combinatorial structures and prime factorizatons. *Notices of the American Mathematical Society*, 44(8):903–910, September 1997.

[6] J. F. Adams. On the nonexistence of elements of hopf invariant one. *Annals of Mathematics*, 72:20–104, 1960.

[7] John P. Nolan adnd Hyppolyte Fofack. Tail behaviour, modes, and other characteristics of stable distributions. *Extremes*, 2(1):39–58, 1999.

[8] Oliver Ledoit Anders Johansen and Didier Sornette. Crashes as critical points. *International Journal of Theoretical and Applied Finance*, 3(1), 2000.

[9] Andrey Feuerverger and Philip McDunnogh. On efficient inference in symmetric stable laws and processes. In M. Csörgö D.A. Dawson J.N.K. Rao A.K. Md. E. Saleh, editor, *Statistics and Related Topics*, Amsterdam, 1981. North-Holland.

[10] Ruth Arad. Parameter estimation for symmetric stable distributions. *International Economics Review*, 21:209–220, 1980.

[11] W. Ruth Arad. *The Implications of a Long-Tailed Distribution Structure to Selection and Capital Asset Pricing.* PhD thesis, Princeton University, January 1975.

[12] A.S. Paulson and T.A. Delahanty. Modified weighted squared error estimation procedures with special emphasis on the stable laws. *Communications in Statistics; Simulation and Computation*, 14(4):927–972, 1985.

[13] Audrey Terras. *Harmonic Analysis on Symmetric Spaces and Applications*, volume I. Springer-Verlag, New York, 1985.

[14] L. Bachelier. *Le jeu, la chance, et le hasard.* Flamarion, Paris, 1914.

[15] Per Bak. *How Nature Works.* Springer-Verlag, New York, 1996.

[16] Per Bak and K. Sneppen. Punctuated equilibrium and criticality in a simple model of evolution. *Physical Review Letters*, 71:4083–4086, 1993.

[17] B.D. Fielitz and E.W. Smith. Asymmetric stable distributions of stock price change. *Journal of the American Statistical Association (Applications)*, 67(340):813–814, 1972.

[18] F. Black and M. Scholes. *Journal of Political Economy*, 3:637, 1973.

[19] B.N. Cheng and S.T. Rachev. Multivariate stable securities in financial markets. *Mathematical Finance.*

[20] Boris V. Gnedenko and A.N. Kolmogorov. *Limit Distributions for Sums of Independent Random Variables.* Addison-Wesley, Reading, Massachusetts, revised edition, 1968.

[21] Jean-Phillipe Bouchaud and Marc Potters. *Theory of Financial Risk: From Data Analysis to Risk Management.* Science & Finance, Paris, France, 2000.

[22] J.P. Burg. *Maximum Entropy Spectral Analysis.* PhD thesis, Department of Geophysics, Stanford University, Stanford, CA, 1975.

[23] B.W. Stuck and B. Kleiner. A statistical analysis of telephone noise. *Bell Systems Technical Journal*, 53:1263–1320, 1974.

[24] P. Embrechts C. Klüppelberg and T. Mikosch. *Modelling Extremal Events for Insurance and Finance.* Springer-Verlag, Heidelberg, 1997.

[25] C. Müller. *Spherical Harmonics.* Number 17 in Lecture Notes in Mathematics. Springer-Verlag, New York, 1966.

[26] H.M. Jaeger C.H. Liu and S.R. Nagel. *Physical Review Letters*, 62:40, 1989.

[27] Isaac Chavel. *Riemannian Geometry: A modern introduction.* Cambridge UP, Cambridge, MA, 1993.

[28] Chrysostomos L. Nikias and A.P. Petropulu. *Higher order spectral analysis: a Nonlinear signal processing environment.* Prentice-Hall, Englewood Cliffs, New Jersey, 1993.

[29] Chrysostomos L. Nikias and M. Shao. *Signal Processing with Alpha-stable distributions and Applications.* John Wiley & Sons,, New York, 1995.

[30] Chrysostomos L. Nikias and Xinyu Ma. Parameter estimation and blind identification in non-gaussian impulsive environments. *IEEE Transactions on Signal Processing*, 43(12):2884–2897, December 1995.

[31] Chrysostomos L. Nikias and Xinyu Ma. Joint estimation of time delay and frequency delay in impulsive noise using fractional lower order statistics. *IEEE Transactions on Signal Processing*, 44(11):2669–2687, November 1996.

[32] Lawrence Conlon. *Differentiable manifolds: a first course.* Birkhäuser, Boston, 1993.

[33] John B. Conway. *A Course in Functional Analysis.* Springer-Verlag, New York, second edition, 1990.

[34] D. Weaire and N. Rivier. Soap, cells and statistics —random patterns in two dimensions. *Conhmp. Physics*, 25(1):59–99, 1984.

[35] Daniel Dennet. *Consciousness Explained.* Little, Brown, and Co., Toronto, 1991.

[36] W.H. DuMouchel. Stable distributions in statistical inference 2: Information from stably distributed samples. *Journal of the American Statistical Association (Theory and Methods)*, 70(350):386–393, 1975.

[37] Richard Durrett. *Lecture notes on particle systems and percolation.* Brooks/Cole Pub. Co., Pacific Grove, CA, 1988.

[38] V.S. Bawa E.J. Elton and M.J. Gruber. Simple rules for optimal portfolio selection in stable paretian markets. *The Journal of Finance*, 34:1041–1047, 1979.

[39] Eugene F. Fama and Richard Roll. Parameter estimates for symmetric stable distributions. *Journal of the American Statistical Association*, 66(334):331–338, 1971.

[40] Lawrence C. Evans. *Partial Differential Equations.* American Mathematical Society, Providence, Rhode Island, 1998.

[41] A.S. Paulson E.W. Holcomb and R.A. Leitch. The estimation of the parameters of the stable laws. *Biometrika*, 62:163–170, 1975.

[42] Eugene F. Fama. Mandelbrot and the stable Paretian hypothesis. *Journal of Business (Chicago)*, 36:420–429, 1963.

[43] Eugene F. Fama. The distribution of daily stock prices: A test of Mandelbrot's stable Paretian hypothesis. *Journal of Business (Chicago)*, 38:34–105, 1965.

[44] Eugene F. Fama. Portfolio analysis in a stable Paretian market. *Management Science*, 6:404–419, 1965.

[45] E. Feldheim. *Etude de la stabilité des lois de probabilitié.* PhD thesis, Thése de la Faculté des Sciences de Paris, 1937.

[46] William Feller. *An introduction to probability theory and its applications*, volume II. John Wiley and Sons, Inc., New York, second edition, 1966.

[47] Enrico Fermi. On the origin of cosmic radiation. *Physical Review*, 75:1169–1174, 1949.

[48] Gerald B. Folland. *Real Analysis.* John Wiley and Sons, New York, 1984.

[49] Malthe-Sorenssen A Feder J Christensen K Frette V and Jossang T. Surface fluctuations and correlations in a pile of rice. *PHYSICAL REVIEW LETTERS*, 83(4):764–767, July 1999.

[50] H. Furstenberg. *Ergodic Theory and Combinatorial Number Theory.* Princeton University Press, Princeton, New Jersey, 1981.

[51] G. Christoph and W. Wolf. *Convergence Theorems with a stable limit law*. Akadamie Verlag, Berlin, 1993.

[52] Andrea Gamba. Portfolio analysis with symmetric stable paretian returns. In Elio Canestrelli, editor, *Current Topics in Quantitative Finance*, pages 48–69. Physica-Verlag, Heidelberg, 1999.

[53] Andrea Gamba. Portfolio analysis with symmetric stable paretian returns. In *Current topics in quantitative finance*, Contributions in Management Science. Physica, Heidelberg, 1999.

[54] R. Gibrat. *Les inégalités économiques*. Sirey, Paris, 1932.

[55] Brian Goodwin. *How the Leopard Changed its spots: the evolution of complexity*. Touchstone, Rockefeller Center, New York, 1994.

[56] Geoffrey. Grimmett. *Percolation*. Springer-Verlag, New York, 2 edition, 1999.

[57] H. Dym and P. McKean. *Fourier Series and Integrals*. Academic Press, New York, 1972.

[58] Dennis M. Healy, Jr. Harrie Hendricks and Peter Kim. Spherical deconvolution. *Journal of Multivariate Analysis*, 67:1–22, 1998.

[59] Sigurdur Helgason. *Topics in Harmonic Analysis on Homogeneous Spaces*. Birkhäuser, Boston, Massachusetts, 1981.

[60] Nicholas Hernandez. *(unpublished)*. PhD thesis, University of Toronto, 2002.

[61] Z. Olami H.J.S. Feder and K. Christensen. *Physical Review Letters*, 68:1244, 1992.

[62] C.H. Liu H.M. Jaeger and S.R. Nagel. *Science*, 255:1523, 1992.

[63] John H. Holland. *Emergence: from chaos to order*. Perseus Books, Reading, Massachusetts, 1998.

[64] J. Holtsmark. Über die verbreiterung von spektrallinien. *Ann. Physik*, 4(58 (363)):577–630, 1919.

[65] Dale Husemoller. *Fibre Bundles*. Springer-Verlag, New York, 3rd edition, 1994.

[66] I. S. Gradshteyn and I. M. Ryzhik. *Tables of Integrals, Series, and Products.* Academic Press, Toronto, corrected and enlarged edition, 1980.

[67] Lotfi Belkacem Jacques Lévy Véhel and Christian Walter. Generalized market equlibrium: "stable" CAPM. Presented at AFFI — International Conference of Finance, Bordeaux, July, 1995.

[68] Lotfi Belkacem Jacques Lévy Véhel and Christian Walter. CAPM, risk and portfolio selection in "stable" markets. Technical Report 2776, Institut National de Recherche en Informatique at en Automatique, Domaine de Voluceau, Rocquencourt, BP 105, 78153 Le Chesnay, Cedex, France, January 1996.

[69] E.T. Jaynes. *Physical Review*, 108:171, 1957.

[70] E.T. Jaynes. *Physical Review*, 106:620, 1957.

[71] E.T. Jaynes. *The Maximum Entropy Formalism.* MIT Press, 1979.

[72] E.T. Jaynes. On the rationale of maximum entropy methods. *Proceedings of the IEEE*, 70:939–952, 1982.

[73] E.T. Jaynes. *Papers on Probability, Statistics, and Statistical Physics.* Reidel, Dordrecht, 1982.

[74] J.C.Willis. *Age and Area.* Cambridge University Press, 1922.

[75] Erica Jen, editor. *1989 Lectures in Complex Systems.* Addison-Wesley, Redwood City, CA, 1990.

[76] Henrik Jeldtoft Jensen. *Self-Organized Criticality.* Cambridge University Press, The Edinburgh Building, Cambridge, CB2 2RU, United Kingdom, 1998.

[77] B.W. Stuck J.M. Chambers, C. Mallows. A method for simulating stable random variables. *Journal of the American Statistical Association*, 71(354):340–344 (Theory and Methods Section), 1976.

[78] John P. Nolan and Anna K. Panorska. Data analysis for heavy tailed multivariate samples. *Communications in Statistics: Stochastic Models*, 13(4):687–702, 1997.

[79] John P. Nolan and Balram Rajput. Calculaton of multidimensional stable densities. *Communications in Statistics: Simulation and Computation*, 25(3):551–566, 1995.

[80] John P. Nolan and Husein Abdul-Hamid. Multivariate stable densities as functions of one-dimensional marginals. *Journal of Multivariate Analysis*, 67:80–89, 1998.

[81] J. Huston McCulloch John P. Nolan, Anna K. Panorska. Estimation of stable spectral measures. *(submitted)*, 1997.

[82] J.P. Bouchaud D. Sornette C. Walter J.P. Aguilar. Taming large events: Optimal portfolio theory for strongly fluctuating assets. *International Journal of Theoretical and Applied Finance*, 1(1):25–41, 1998.

[83] J.T. Lo and L.R. Eshelman. Exponential fourier densities on $\mathbb{SO}(3)$ and optimal estimation and detection for rotational processes. *SIAM Journal of Applied Mathematics*, 36:73–82, 1979.

[84] Yitzhak Katznelson. *An Introduction to Harmonic Analysis*. Dover, 180 Varick Street, New York, NY, 10014, USA, first edition, 1976.

[85] Stuart Kauffman. *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University PRess, New York, 1993.

[86] Stuart Kauffman. *At Home in the Universe*. Oxford University PRess, New York, 1995.

[87] Stuart Kauffman. *Investigations*. Oxford University PRess, New York, 2000.

[88] J.A. Scott Kelso. *Dynamic Patterns: the Self-organization of Brain and Behaviour*. MIT Press, Cambridge, Massachusetts, 1995.

[89] Harry Kesten. *Percolation theory for mathematicians*. Birkhauser, Boston, 1982.

[90] R. Kikuchi. *Journal of Chemical Physics*, 24:861, 1956.

[91] Anthony W. Knapp. *Representation Theory of Semisimple Groups: an overview based on examples*. Princeton University PRess, Princeton, New Jersey, 1986.

[92] Ioannis A. Koutrouvelis. Regression-type estimation of the parameters of stable laws. *Journal of the American Statistical Association, (Theory and Methods)*, 75(372):918–928, 1980.

[93] Ioannis A. Koutrouvelis. An iterative procedure for the estimation of the parameters of stable laws. *Communications in Statistics: Simulation and Computation*, B10:17–28, 1981.

[94] Paul Krugman. *The Self-Organizing Economy*. Blackwell Publishers, 228 Main Street, Cambridge, MA, 02142, 1 edition, 1996.

[95] J. Kuelbs. A representation theorem for symmetric stable processes and symmetric measures on h. *Zeitschrift für Wahrscheinlichtkeitstheorie und verwalte Gebiete*, 26:259–271, 1973.

[96] N. N. Lebedev. *Special Functions and Applications*. Dover, New York, 1972.

[97] L.G. Benckert and I. Sternberg. An attempt to find an expression for the distribution of fire damage amount. In *Transactions 15th International Congress of Actuaries*, volume 2, pages 288–294, New York, 1957.

[98] Thomas M. Liggett. *Interacting Particle Systems*. Springer-Verlag, New York, 1985.

[99] Lotfi Belkacem. How to select a portfolio in α-stable markets. Technical report, Institut National de Recherche en Informatique at en Automatique, Domaine de Voluceau, Rocquencourt, BP 105, 78153 Le Chesnay, Cedex, France, January 1997.

[100] Lotfi Belkacem and Jacques Lévy Véhel. Capm, risk, and portfolio selection in α-stable markets. *Fractals*, 8(1):99–115, 2000.

[101] A.J. Lotka. *Elements of Mathematical Biology*. Dover, New York, 1925, reprinted 1956.

[102] J. Rosendahl M. Vekić and J. Kelly. Persistent self-organization of sandpiles. *Physical Review E.*, 47(2):1401–1404, February 1993.

[103] F.R. Macaulay. Pareto's laws and the general problem of mathematically describing the frequency of income. In *Income in the United States, Its amount and distribution, 1909-1919*, volume 2, New York, 1922. National Bureau of Economic Research.

[104] Benoit B. Mandelbrot. Adaptation d'un message á la ligne de transmission I & II. *Comptes Rendus*, 232:1638–1740, 2003–2005, 1951.

[105] Benoit B. Mandelbrot. *Fractals and Scaling in Finance: Discontinuities, Concentration, Risk*. Springer Verlag, New York, 1997.

[106] Benoit B. Mandelbrot. Scaling distributions and income maximization. In *Fractals and Scaling in Finance: Discontinuities, Concentration, Risk*, (orig. 1962).

[107] Benoit B. Mandelbrot. The variation in the prices of cotton, wheat, and railroad stocks, and of some financial rates. In *Fractals and Scaling in Finance: Discontinuities, Concentration, Risk*, (orig. 1965).

[108] Benoit B. Mandelbrot. The variation of certain speculative prices. In *Fractals and Scaling in Finance: Discontinuities, Concentration, Risk*, New York, (orig. 1965). Springer Verlag.

[109] Benoit B. Mandelbrot. Proportional growth with or without diffusion, and other explanations of scaling. In *Fractals and Scaling in Finance: Discontinuities, Concentration, Risk*, (orig. 1996).

[110] Benoit B. Mandelbrot. Rank-size plots, Zipf's law, and scaling. In *Fractals and Scaling in Finance: Discontinuities, Concentration, Risk*, (orig. 1996).

[111] Benoit B. Mandelbrot. States of randomness from mild to wild, and concentration in the short, medium, and long term. In *Fractals and Scaling in Finance: Discontinuities, Concentration, Risk*, (orig. 1996).

[112] Rosario N. Mantegna and H. Eugene Stanley. Scaling behaviour in the dynamics of an economic index. *Nature*, 376(6):46–49, July 1995.

[113] Tom Hurd Jeff Boland Marcus Pivato and Luis Seco. Measures of dependence for multivariate lévy distributions. In *Conference Proceedings of the IAMP*, London, July 2000.

[114] H. Markovitz. *Portfolio Selection: Efficient Diversification of Investments*. John Wiley and Sons, New York, 1959.

[115] Per Bak Maya Paczuski Martin Shubik. Price variations in a stock market with many agents. *Physica A*, 246:430–453, 1997.

[116] J. Huston McCulloch. Simple consistent estimators of stable distribution parameters. *Communications in Statistics: Simulation and Computation*, 15(4):1109–1136, 1986.

[117] J. Huston McCulloch. Financial applications of stable distributions. In *Statistical Methods in Finance*, volume 14 of *Handbook of Statistics*, pages 393–425. North-Holland, Amsterdam, 1996.

[118] Elliot Middleton. Adaptation level, 'animal spirits' and expectations: the path to self-organized criticality. *Nonlinear dynamics, psychology, and Life Sciences*, under review: see `http://www.marketscience.com/chaos97.zip`, 1997.

[119] Elliot Middleton. Adaptation level, 'animal spirits' and self-organized criticality; `http://www.marketscience.com/socpaper.zip`. available at `http://www.marketscience.com/socpaper.zip`, 1997.

[120] G. Miller. Properties of certain symmetric stable distributions. *Journal of Multivariate Analysis*, 8:346–360, 1978.

[121] Walter Willinger Murad S. Taqqu and A. Erramilli. A bibliographical guide to self-similar traffic and performance modelling for modern high-speed networks. In F.P.Kelly S. Zachary and I. Ziedins, editors, *Stochastic Networks: Theory and Applications*, pages 339–366. Clarendon Press, Oxford, UK, 1996.

[122] W.E. Leland Murad S. Taqqu and Walter Willinger. On the self-similar nature of ethernet traffic. *IEEE/ACM Trans. Networking*, 2:1–15, 1994.

[123] N. Rivier and A. Lissowski. *Journal of Physics A*, 15:L143, 1982.

[124] Lynn Nadel and Daniel Stein, editors. *1990 Lectures in Complexity.* Addison-Wesley, Redwood City, CA, 1991.

[125] Lynn Nadel and Daniel Stein, editors. *1991 Lectures in Complexity.* Addison-Wesley, Redwood City, CA, 1992.

[126] John P. Nolan. Identification of stable measures. *(in preparation)*, 1996.

[127] John P. Nolan. Maximum likelihood estimation and diagnostics for stable distributions. *(preprint)*, 1999. Available from `http://www.cas.american.edu/~jpnolan`.

[128] John P. Nolan. Multivariate stable distributions: approximation, estimation, simulation, and identification. *(preprint)*, 1999. Available from `http://www.cas.american.edu/~jpnolan`.

[129] John P. Nolan. *Stable Distributions*. (preprint), April 1999. Available from `http://www.cas.american.edu/~jpnolan`.

[130] J.P. Nolan. Fitting data and assessing goodness of fit with stable distributions. *(preprint)*, June 1999. Available from `http://www.cas.american.edu/~jpnolan`.

[131] Z. Olami and K. Christensen. *Physical Review A*, 46:R1720, 1992.

[132] P. Tsakalides and C. Nikias. Maximum likelihood localization of sources in noise modelled as a stable process. *IEEE Transactions on Signal Processing*, 43, 1995.

[133] Vilfredo Pareto. *Cours d'économie politique*. Droz, Geneva, 1896,1965.

[134] Vilfredo Pareto. *Ouvres complètes*. Droz, Geneva, 1965.

[135] R.D. Peirce. Application of the positive α-stable distribution. In *Signal Processing Workshop on Higher-Order Statistics*, pages 420–424, Piscataway, New Jersey, 1997. IEEE Press.

[136] Chao Tang Per Bak and Kurt Wiesenfeld. Self-organized criticality. *Physical Review A*, 38(1):364–373, July 1988.

[137] José Scheinkman Per Bak, Kan Chen and Michael Woodford. Aggregate fluctuations from independent sectoral shocks: Self-organized criticality in a model of production and inventory dynamics. *Ricerche Economiche*, 57(1):3–30, March 1993.

[138] Kurt Wiesenfeld Per Bak, Chao Tang. Self-organized criticality: An explanation of $1/f$ noise. *Physical Review Letters*, 59(4):381–384, July 1987.

[139] M. Bretz J. B. Cunningham P.L. Kurczynski and F. Nori. *Physical Review Letters*, 69:2431, 1992.

[140] G.A. Held H.D.H. Solina D.T. Keane W.J. Haag P.M Horne and G. Grinstein. *Physical Review Letters*, 65:1120, 1990.

[141] James S. Press. Estimation in univariate and multivariate stable distributions. *Journal of the American Statistical Association (Theory and Methods)*, 67(340):842–846, 1972.

[142] James S. Press. Multivariate stable distributions. *Journal of Multivariate Analysis*, 2:444–462, 1972.

[143] R. Adler R. Feldman and M. Taqqu, editors. *A Practical Guide to Heavy Tailed Data.* Birkhauser, Boston, 1998.

[144] R. R. Coifman and G. Wiess. Representations of compact groups and spherical harmonics. *L'enseignment Math.*, 14:123–173, 1968.

[145] W. Willinger M. Taqqu R. Sherman and D. Wilson. Self-similarity through high variability: Statistical analysis of ethernet lan traffic at the source level. In *Proceedings of the ACM/SIGCOMM '95*, pages 100–113, 1995.

[146] R.Courant and D. Hilbert. *Methods of Mathematical Physics*, volume I. Wiley Interscience, New York, 1961.

[147] Richard L. Bishop and Samuel I. Goldberg. *Tensor Analysis on Manifolds.* Dover, New York, 1980.

[148] T. Riste and D. Sherrington, editors. *Spontaneous Formation of Space-Time Structures and Criticality.* Kluwer, Dordrecht, 1991.

[149] Robert Rosen. *Life Itself: A comprehensive inquiry into the nature, origin, and fabrication of life.* Complexity in Ecological Systems. Columbia University Press, New York, 1991.

[150] S. Cambanis and G. Miller. Linear problems in p-th order stable processes. *SIAM Journal of Applied Mathematics*, 43(1):43–69, August 1981.

[151] F. Nori S. Field, J. Witt. Superconducting vortex avalanches. *Physical Review Letters*, 74:1206, 1995.

[152] S. Mittnik and S.T. Rachev. Option pricing for stable and infinitely divisible asset returns. *Mth. Comput. Modelling*, 29(10-12):93–104, 1999.

[153] Saleh O. Badahdah and M.M. Siddiqui. Adaptive robust estimation of location for stable distributions. *Communications in Statistics: Theory and Methods*, 20(2):631–652, 1991.

[154] Gennady Samorodnitsky and Murad S. Taqqu. *Stable Non-Gaussian random processes: stochastic models with infinite variance.* Chapman and Hall, New York, 1994.

[155] P.A. Samuelson. Efficient portfolio selection for pareto-levy investments. *Journal of Financial and Quantitative Analysis*, 2:107–122, 1967.

[156] José Scheinkman and Michael Woodford. Self-organized criticality and economic fluctuations. *American Economics Association, Papers and Proceedings*, 84(2):417–421, May 1994.

[157] C.H. Scholz. *The Mechanics of Earthquakes and Faulting.* Cambridge University Press, Cambridge, UK, 1990.

[158] C.H. Scholz. Spontaneous formation of space-time structures and criticality. In Riste and Sherrington [148].

[159] Laurent Schwartz. *Lectures on disintegration of measures.* Tata Institute of Fundamental Research, Bombay, 1975.

[160] S.E. Luria and M. Delbrück. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics*, 28:491–511, 1943.

[161] W.F. Sharpe. *Management Science*, pages 9–2,277–293, 1963.

[162] H. Simon. On a class of skew distribution functions. *Biometrika*, 1955.

[163] Didier Sornette. Spontaneous formation of space-time structures and criticality. In Riste and Sherrington [148].

[164] Daniel L. Stein, editor. *Lectures in the Science of Complexity: Proceedings of the Santa Fe Summer School, 1988.* Addison-Wesley, Redwood City, CA, 1989.

[165] M. Sugiura. *Unitary Representations and Harmonic Analysis.* Wiley, New York, 1975.

[166] Svetlozar T. Rachev and Huang Xin. Test for association of random variables in the domain of attraction of multivariate stable law. *Probability and Mathematical Statistics*, 14(1):125–141, 1993.

[167] V. Frette K. Christensen A. Malthe-Sorensen J. Feder T. Jossang and P. Meakin. Avalanche dynamics in a pile of rice. *Nature*, 379(6560):49–52, January 1996.

[168] Masaru Takeuchi. *Modern Spherical Functions*, volume 135 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, Rhode Island, 1994.

[169] J.D. Talman. *Special Functions: A Group Theoretic Approach*. Benjamin, New York, 1968.

[170] M. Tanemura. *Proceedings of the Institute of Statistical Mathematics*, 26:107, 1979.

[171] Michael Eugene Taylor. *Noncommutative Harmonic Analysis*. American Mathematical Society, Providence, Rhode Island, 1986.

[172] Theodor Brocker and Tammo tom Dieck. *Representations of Compact Lie Groups*. Springer-Verlag, New York, 1 edition, 1985.

[173] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, New York, 1991.

[174] John P. Nolan Tomasz Byczkowski and Balram Rajput. Approximation of multidimensional stable densities. *Journal of Multivariate Analysis*, 46:13–31, 1993.

[175] van der Vaart, Aad. W. and Wellner, J. *Weak convergence and empirical processes: with applications to statistics*. Springer series in Statistics. Springer-Verlag, New York, 1996.

[176] N.J. Vilenkin. *Special Functions and the Theory of Group Representations*, volume 22 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, Rhode Island, 1968.

[177] Vivek S. Borkar. *Probability Theory: A Concise Course*. Universitext. Springer-Verlag, New York, 1995.

[178] G. Wahba. Spline interpolation and smoothing on the sphere. *SIAM Journal Sci.t Statist. Comput.*, 2:5–16, 1981.

[179] Frank M. Warner. *Foundations of Differentiable Manifolds and Lie Groups*. Springer-Verlag, New York, 1983.

[180] Hermann Weyl. *The Theory of Groups and Quantum Mechanics*. Dover, New York, 2nd edition, 1911.

[181] Walter Willinger and Vern Paxson. Where mathematics meets the internet. *Notices of the Americal Mathematical Society*, 45(8):961–970, September 1998.

[182] G. Udney Yule. A mathematical theory of evolution, based on the conclusions of Dr. J.C. Willis, F.R.S. *Philosophical Transactions of the Royal Society of London, Series B*, 213:21–87, 1925.

[183] D. Zajdenweber. *Hasard et prévision.* PhD thesis, Faculté de Droit de Paris, 1976.

[184] D. Zajdenweber. Self-similarity properties of the cac 40 index (french, with englsh summarry. *Revue d'Économie Politique*, 104:407–434, 1994.

[185] D. Zajdenweber. Business interuption insurance, a risky business. a study on some paretion risk phenomena. *Fractals*, 3:601–608, 1995.

[186] D. Zajdenweber. Extreme values in business interuption insurance. *The Journal of Risk and Insurance*, 63:95–110, 1995.

[187] W.T. Ziemba. Choosing investment portfolios when the returns have stable distributions. In *Mathematical Programming in theory and practice; Proceedings of the NATO Advanced Study Institute, Figueira da Foz*, pages 443–482, Amsterdam, 1974. North-Holland.

[188] George Kingsley Zipf. *Human behavior and the principle of least effort : an introduction to human ecology.* Hafner, New York, 1965.

[189] Vladimir M. Zolotarev. Statistical estimates of the parameters of stable laws. In *Mathematical Statistics: Banach Center Publications*, volume 6, pages 359–376, Warsaw, 1980. Polish Scientific Publishers.

[190] V.M. Zolotarev. *One-dimensional Stable Distributions.* American Mathematical Society, Providence, Rhode Island, 1986.

Index of Notation

$\mathbb{A} \subset \mathbb{S}^{D-1}$ usually designates a countable set of atoms for some spectral measure.

$\mathcal{A}$: The area of the unit sphere $\mathbb{S}^{D-1}$.

α: The **stability exponent** of a stable probability distribution, ranging from 0 to 2.

$\mathbb{B}(\mathbf{x};\epsilon)$: The open ball around $\mathbf{x}$ of radius ϵ.

$\mathbb{B}^D(\mathbf{x};\epsilon)$: The open ball around $\mathbf{x}$ of radius ϵ, in $\mathbb{R}^D$.

$\mathbb{B}^{\mathcal{M}}(\mathbf{x};\epsilon)$: The open ball around $\mathbf{x}$ of radius ϵ, on the Riemannian manifold $\mathcal{M}$.

$\mathsf{B}_e(0;r)$: The open ball of radius r around 0 in the tangent space $\mathsf{T}_e\,\mathcal{M}$.

$\mathcal{B}_\alpha = \tan\left(\frac{\pi\alpha}{2}\right)$, if $\alpha \neq 1$, while $\mathcal{B}_1 = -\frac{2}{\pi}$.

β: The **skewness parameter** of a univariate stable probability distribution, ranging from -1 to $+1$.

$\mathcal{C}(\mathbf{X})$: The space of continuous, complex-valued functions on topological space $\mathbf{X}$.

$\mathcal{C}^N(\mathcal{M})$: The space of N-times differentiable complex-valued functions on manifold $\mathcal{M}$.

$\mathcal{C}^\infty(\mathcal{M})$: The space of smooth, complex-valued functions on manifold $\mathcal{M}$.

χ: Usually designates the Fourier transform (ie. characteristic function) of a probability measure.

D: Usually the dimension of the Euclidean space under discussion, as in $\mathbb{R}^D$.

$\mathsf{D}_m f$: The **derivative** of a smooth function $f : \mathcal{M} \longrightarrow \mathcal{N}$ at $m \in \mathcal{M}$, where $\mathcal{M}, \mathcal{N}$ are manifolds and $\mathsf{D}_m : \mathsf{T}_m \mathcal{M} \longrightarrow \mathsf{T}_{f(m)} \mathcal{N}$ is a linear map (see Definition 117 on page 210).

$\mathbb{D} = \underbrace{[1..D] \times \ldots \times [1..D]}_{N}$ (§ E.1 on page 199).

$\mathbb{D}^{(<)} = \{(d_1, \ldots, d_N) \; ; \; 1 \leq d_1 < d_2 < \ldots < d_N \leq D\}$ (§ E.2 on page 201).

$d \in [1..D]$ usually indexes a coordinate in D-dimensional space.

$\mathfrak{d}$: The **exterior derivative** operator on differential forms (§ 124 on page 214).

$\delta_{\mathbf{a}}$: The point mass at $\mathbf{a}$.

$\triangle$: The Laplacian operator (see § G.5 on page 223).

dist $[x, y]$: The distance between two points x and y in a metric space.

$\mathbf{D}^{istr}[\mathbf{X}]$: The probability distribution of random variable $\mathbf{X}$.

$\mathbf{e}_d$: If $d \in [1..D]$, then $\mathbf{e}_d$ is the dth canonical basis vector of $\mathbb{R}^D$:

$$\mathbf{e}_d = \left(\underbrace{0, \ldots, 0}_{d-1}, 1, 0, \ldots, 0 \right)$$

$\mathbf{e}_{[\otimes \mathbf{d}]} = \mathbf{e}_{d_1} \otimes \mathbf{e}_{d_2} \otimes \ldots \otimes \mathbf{e}_{d_n}$, where $\mathbf{d} = (d_1, \ldots, d_N) \in \mathbb{D}$, (§ E.1 on page 199).

$\mathbf{e}_{[\wedge \mathbf{d}]} = \mathbf{e}_{d_1} \wedge \mathbf{e}_{d_2} \wedge \ldots \wedge \mathbf{e}_{d_N}$, where $\mathbf{d} = (d_1, \ldots, d_N) \in \mathbb{D}^{(<)}$ (§ E.2 on page 201).

$\mathcal{E}_{\vec{\xi}}$: If $\vec{\xi} \in \mathbb{R}^D$, then $\mathcal{E}_{\vec{\xi}} : \mathbb{R}^D \longrightarrow \mathbb{C}$ is the periodic function with multifrequency $\vec{\xi}$ –in other words, $\mathcal{E}_{\vec{\xi}}(\mathbf{x}) = \exp\left(2\pi \mathbf{i} \cdot \left\langle \mathbf{x}, \vec{\xi} \right\rangle\right)$.

$\eta^{(\alpha)}$: If $\mathbf{s}, \theta \in \mathbb{S}^{D-1}$, then $\eta^{(\alpha)}(\mathbf{s}, \theta) = |\langle \mathbf{s}, \theta \rangle|^{\alpha} + \mathcal{B}_\alpha \cdot \langle \mathbf{s}, \theta \rangle^{\langle \alpha \rangle} \cdot \mathbf{i}$.

$\exp_m(\vec{v})$: If $\mathcal{M}$ is a Riemannian manifold, and $m \in \mathcal{M}$, then exp is the **exponential map** from $\mathsf{T}_m \mathcal{M}$ into $\mathcal{M}$. (see Definition 136 on page 220).

$\mathbf{E}_{\mu}^{xpct}[f]$: The **conditional expectation** of the function f relative to the probability measure μ.

$\mathbf{E}^{expct}[\mathbf{X}]$: The **conditional expectation** of random variable $\mathbf{X}$.

Φ: Usually indicates the logarithm of the characteristic function of a probability measure.

Γ: Usually indicates the **spectral measure** of a multivariate stable probability distribution. (see Theorem 77 on page 143).

γ: usually the continuous part of spectral measure Γ.

$\gamma_{\mathbf{a}}$: If $\mathbf{a} \in \mathbb{S}^{D-1}$, then $\gamma_{\mathbf{a}}$ is usually the weight given to the point mass at $\mathbf{a}$, as part of spectral measure Γ.

$\mathbb{G}_e$: If $\mathbb{G}$ is a Lie group acting on manifold $\mathcal{M}$, and $e \in \mathcal{M}$, then $\mathbb{G}_e^*$ is the subgroup of $\mathbb{G}$ fixing e, and $\mathbb{G}_e$ is the connected component of $\mathbb{G}_e^*$ containing the identity (see Chapter 5).

$\mathcal{H}_{\mathbb{G}}^{haar}$: The Haar measure on compact group $\mathbb{G}$.

$\mathbf{i}$: The square root of -1.

$\mathbf{im}[z]$: The imaginary part of complex number z.

$\mathcal{K}(\rho) = \min_{\theta \in \mathbb{S}^{D-1}} \left| \log |\chi(\theta)| \right|$, where ρ is a probability measure, and χ its characteristic function (Definition 6 on page 19).

$\mathcal{L}^{leb}$: The Lebesgue measure, on $\mathbb{R}^D$ or a Riemannian manifold.

$\mathbf{L}^p(\mathbf{X}, \mu) = \left\{ f : \mathbf{X} \longrightarrow \mathbb{C} \, ; \; f \text{ is measurable, and } \int_{\mathbf{X}} |f|^p \; d\mu < \infty \right\}$ (for any $p \in [1, \infty)$).

$\mathbf{L}^{\langle \alpha \rangle}(\mathbf{X}, \mu) = \left\{ f : \mathbf{X} \longrightarrow \mathbb{R} \, ; \; f \text{ is measurable, and } \int_{\mathbf{X}} |f|^\alpha \; d\mu < \infty \right\}$, for any $\alpha \in [0, 2]$, $\alpha \neq 1$ (§ D.2 on page 181).

$\langle f \rangle^{\langle \alpha \rangle} = \int_{\mathbf{X}} f(x)^{\langle \alpha \rangle} \; d\mu[x]$, for any $\alpha \in [0, 2]$, $\alpha \neq 1$ (§ D.2 on page 181).

$\|f\|_\alpha = \left(\int_{\mathbf{X}} |f(x)|^\alpha \; d\mu[x] \right)^{1/\alpha}$, for any $\alpha \in [0, 2]$, although this is obviously not really a norm if $\alpha < 1$ (Corollary 101 on page 191).

$\mathbf{L}^{\langle 1\rangle}(\mathbf{X},\mu) = \left\{ f : \mathbf{X} \longrightarrow \mathbb{R};\ \ f \text{ is measurable, } \int_{\mathbf{X}} |f|\, d\mu < \infty, \text{ and } \int_{\mathbf{X}} |f| \log |f|\, d\mu < \infty \right\}$ (Corollary 102 on page 191).

$\langle f \rangle_{\langle 1\rangle} = \int_{\mathbf{X}} f(x) \log |f(x)|\, d\mu[x]$. (Proposition 100 on page 189).

$\Lambda(\mathcal{M}) = \{\lambda \in \mathbb{C}\ ;\ -\lambda \text{ is an eigenvalue of } \triangle\}$ where $\triangle$ is the Laplacian on Riemannian manifold $\mathcal{M}$ (§ 5.1 on page 51).

$\log(x)$ is always the natural logarithm of x.

$\mathcal{M}_{\mathcal{EAS}}[\mathbf{X}]$**:** The space of probability measures on a measurable space $\mathbf{X}$. If $\mathbf{X}$ is a topological space, we assume the Borel sigma algebra, and endow $\mathcal{M}_{\mathcal{EAS}}[\mathbf{X}]$ with the weak* topology of convergence it inherits as the dual of $\mathcal{C}(\mathbf{X})$.

$\mathcal{M}_{\mathcal{EAS}}[\mathbf{X};\ \mathbb{R}]$**:** The space of finite, signed measures on a measurable space $\mathbf{X}$.

$\mathcal{M}_{\mathcal{EAS}}[\mathbf{X};\ \mathbb{C}]$**:** The space of complex-valued measures on a measurable space $\mathbf{X}$.

μ**:** The **center parameter** of a stable probability distribution (equivalent to the mean if $\alpha > 1$).

$\mathcal{N}_{\mu}^{\alpha}$**:** The α-stable random noise (ie. random measure) on some measurable space $\mathbf{X}$, with intensity measure μ (see § D.2 on page 181).

$\curvearrowright\!\curvearrowleft[\mu;\sigma]$**:** The normal (or "Gaussian") distribution on $\mathbb{R}$ with mean μ and variance σ.

$\curvearrowright\!\curvearrowleft^{\alpha}[\mu,\beta,\sigma]$**:** The α-stable distribution on $\mathbb{R}$ with centre parameter μ, skewness β, and variation σ. (see § A.1 on page 129).

$\mathbf{pr}_{\mathbf{X}}$**:** If $\mathbf{X} \times \mathbf{Y}$ is a Cartesian product, then $\mathbf{pr}_{\mathbf{X}} : \mathbf{X} \times \mathbf{Y} \longrightarrow \mathbf{X}$ is the projection onto the first coordinate.

$\mathbf{pr}_{\mathbb{V}}$**:** If $\mathbb{H}$ is a Hilbert space, and $\mathbb{V} \subset \mathbb{H}$ a subspace, then $\mathbf{pr}_{\mathbb{V}} : \mathbb{H} \longrightarrow \mathbb{V}$ is the orthogonal projection.

$\mathbf{re}[z]$**:** The real part of complex number z.

ρ usually denotes a (stable) probability measure.

$\mathfrak{S}\mathsf{T}\mathcal{M}$: The $\mathcal{C}^\infty$-algebra of smooth tangent vector fields over manifold $\mathcal{M}$ (§ F.2.2 on page 211).

$\mathfrak{S}\bigwedge^N \mathsf{T}^*\mathcal{M}$: The vector space of differential N-forms over manifold $\mathcal{M}$.

$\mathbb{S}^{D-1}$: The unit sphere in $\mathbb{R}^D$.

σ usually denotes the **variation** of a univariate stable probability distribution.

$\mathbb{SO}^D[\mathbb{R}]$: The special orthogonal group on $\mathbb{R}^D$.

$\star$: The **Hodge star operator** (§ E.4 on page 202)

supp$[\mu]$: The **support** of the measure μ.

$\mathsf{T}_{\mathbf{x}}\mathcal{M}$: The **tangent space** of manifold $\mathcal{M}$ at $\mathbf{x}$. (see § F.2 on page 209).

$\mathsf{T}\mathcal{M}$: The **tangent bundle** of manifold $\mathcal{M}$.

$\mathbb{T}^D$: The D-dimensional torus, as a topological group. Often, we identify $\mathbb{T}^D$ with the unit cube $[0,1)^D$, via the obvious coordinate system.

$\mathbb{U}^D$: The **unitary group** on $\mathbb{C}^D$ (Definition 154 on page 229).

$\mathfrak{U}[\mathbb{G}]$: The set of all **irreducible unitary representations** of a group $\mathbb{G}$ (Definition 152 on page 227).

$\mathbb{V}_\lambda$: In Chapter 5, the eigenspace of the Laplacian corresponding to eigenvalue λ.

$\mathbb{V}^*$: The **dual** of vector space $\mathbb{V}$.

$\bigotimes^N \mathbb{V}$: The N-fold **tensor product** of a vector space.

$\bigwedge^N \mathbb{V}^*$: The N-fold **exterior product** of a dual vector space (see § E.2 on page 201).

$\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \ldots \wedge \mathbf{v}_N = \sum_{\sigma \in \boldsymbol{\sigma}(N)} \mathbf{sign}[\sigma] \cdot \mathbf{v}_{\sigma(1)} \otimes \ldots \otimes \mathbf{v}_{\sigma(N)}$, where $\boldsymbol{\sigma}(N)$ is the permutation group on $[1..N]$, and $\mathbf{v}_1, \ldots, \mathbf{v}_N$ are covectors (see § E.2 on page 201).

$\mathbf{wk}^*\text{-}\lim_{n\to\infty} \rho_n$**:** The weak*-limit of the sequence of measures ρ_n.

Ω_n**:** The canonical volume form on a Riemannian manifold (Definition 140 on page 221).

$\mathcal{Z}_e(\mathbb{V})$**:** If $\mathbb{V} \subset \mathcal{C}^\infty(\mathcal{M})$ is a $\mathbb{G}$-invariant linear subspace, then $\mathcal{Z}_e(\mathbb{V})$ is the subspace of **zonal** elements of $\mathbb{V}$, relative to $\mathbb{G}$ and $e \in \mathcal{M}$. (see Definition 27 on page 52).

ζ**:** Usually indicates a **zonal** function.

$x^{\langle\alpha\rangle} = \mathbf{sign}(x)|x|^\alpha$, if $\alpha \neq 1$, while $x^{\langle 1\rangle} = x \log |x|$.

$\|\Gamma\|_{var}$**:** The **total variation norm** of measure Γ.

$\|\Gamma\|_{[p]}$**:** The $[p]$**-norm** of nice measure Γ (Definition 11 on page 27).

$\widehat{g}$**:** If $g : \mathbb{R}^D \longrightarrow \mathbb{C}$, then $\widehat{g} : \mathbb{R}^D \longrightarrow \mathbb{C}$ is the (classical) **Fourier Transform** of g: for all $\boldsymbol{\xi} \in \mathbb{R}^D$, $\widehat{g}(\boldsymbol{\xi}) = \int_{\mathbb{R}^D} g(\mathbf{x}) \cdot \mathcal{E}_{\vec{\xi}}(\mathbf{x})\ d\mathcal{L}^{\mathbf{bsg}}[\mathbf{x}]$.

$\widehat{g}$**:** If $g : \mathbb{T}^D \longrightarrow \mathbb{C}$, then $\widehat{g} : \mathbb{Z}^D \longrightarrow \mathbb{C}$ is defined: $\forall \mathbf{n} \in \mathbb{Z}^D$, $\widehat{g}(\mathbf{n}) = \int_{[0,1]^D} g(\theta) \cdot \exp\left(2\pi\mathbf{i}\left\langle \mathbf{x}, \mathbf{n}\right\rangle\right)\ d\mathcal{L}^{\mathbf{bsg}}[\theta]$, where we identify $\mathbb{T}^D \cong [0,1)^D$ in the obvious way.

$\widehat{\rho}$**:** If $\rho \in \mathcal{M}_{\mathcal{EAS}}\left[\mathbb{R}^D\right]$, then $\widehat{\rho}$ is the (classical) **Fourier transform** $\widehat{\rho}(\boldsymbol{\xi}) = \int_{\mathbb{R}^D} \mathcal{E}_{\vec{\xi}}(\mathbf{x})\ d\rho[\mathbf{x}]$.

$\widehat{\rho}$**:** If $\rho \in \mathcal{M}_{\mathcal{EAS}}\left[\mathbb{T}^D\right]$, then $\widehat{g} : \mathbb{Z}^D \longrightarrow \mathbb{C}$ is defined: $\forall \mathbf{n} \in \mathbb{Z}^D$, $\widehat{g}(\mathbf{n}) = \int_{[0,1]^D} \exp\left(2\pi\mathbf{i}\left\langle \mathbf{x}, \mathbf{n}\right\rangle\right)\ d\rho[\theta]$, where we identify $\mathbb{T}^D \cong [0,1)^D$ in the obvious way.

$\check{\widehat{\Gamma}}_n$**:** The nth **spherical Fourier coefficient** of measure Γ (see Definition 46 on page 73 and Definition 52 on page 79).

$\langle \mathbf{x}, \mathbf{y}\rangle$**:** The inner product of vectors $\mathbf{x}$ and $\mathbf{y}$.

$\mathbf{X} \underset{distr}{\cong} \mathbf{Y}$ means that random variables $\mathbf{X}$ and $\mathbf{Y}$ have the same distribution.

$[1..N] = \{1, 2, \ldots, N\}$