ABSTRACT

Title of Dissertation: Uniqueness for continuous superresolution by means of Choquet theory and geometric measure theory

Ryan M. Cinoman
Doctor of Philosophy, 2021

Dissertation Directed by: Professor John Benedetto
Department of Mathematics

The problem of superresolution is to recover an element of a vector space from data much smaller than the dimension of the space, using a prior assumption of sparsity. A famous example is compressive sensing, where the elements are images with a large finite resolution. On the other hand, we focus on a continuous form of superresolution. Given a measure μ on a continuous domain such as the two dimensional torus, can we recover μ from knowledge of only a finite number of its Fourier coefficients using a total variation minimization method? We will see that the answer depends on certain properties of μ. Namely, a necessary condition is that μ be discrete.

We use methods from geometric analysis to investigate the continuous superresolution problem. Tools from measure theory relate properties of the support of a measure, such as Hausdorff dimension, to properties of its Fourier transform. We also use measure theory to investigate the possibility of alternatives to total variation that may allow us to recover surface measures defined on space curves.

There is a theorem of Choquet concerning representations of points in convex sets as sums of their extreme points. As it turns out, we can apply this to the superresolution problem because the extreme points of the set of measures with total variation 1 are precisely the set of delta measures. We consider superresolution as a convex optimization problem, where the goal is to find representations of the initial data as sums of delta measures. Choquet theory provides tools to investigate the previously unresolved problem of uniqueness. We use this to give a novel sufficient condition for a measure to be uniquely superresolved, given data on a known finite set of frequencies.

Uniqueness for continuous superresolution by means of Choquet theory and geometric measure theory

by

Ryan M. Cinoman

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2021

Advisory Committee:
Professor John J. Benedetto, Chair/Advisor
Professor Radu V. Balan
Professor Wojciech Czaja
Professor Matei Machedon
Dean's Representative:
Professor Alexander Barg

Acknowledgments

I would like to extend my sincere thanks to my advisor, Dr. John Benedetto. I deeply appreciate the opportunity he has given me to work on this exciting topic, as well as the invaluable advice and guidance he has given me over the years.

I'd also like to gratefully acknowledge ARO Grant W911NF-17-1-0014, which has supported me during my research.

Finally, I'd like to deeply thank my family and friends who have supported me throughout my research. I'd like to extend a special thanks to my parents, my boyfriend Lawrence Cho, and Jill Cassone. Without the unwavering support of each of them, this thesis would not have been possible.

Table of Contents

Chapter 1: Introduction

1.1 Compressive Sensing

A major motivation for studying superresolution is the success of the so-called discrete case, known under another name as compressive sensing. The field of compressed sensing is based on a famous set of papers published starting in 2004 written by Candès, Romberg and Tao [14, 15, 16, 17]. The motivating problem is to find sparse solutions to underdetermined systems of equations. The main results of the original papers were that: yes, it is possible to recover a k-sparse vector in $\mathbb{R}^N$ from only m measurements when $m \gtrsim k \ln(eN/k)$; the algorithm to do so is practically effecient; and random sensing matrices are satisfactory for this algorithm with high probability $\geq 1 - 2\exp(-Cm)$ provided the number of measurements is sufficient as before.

In more formal terms, what we want is a sensing matrix A which is "short and fat," i.e. $m \times N$ where $m \ll N$. We denote our sensed data $\mathbf{y} = A\mathbf{x}$, where $\mathbf{x}$ is some k-sparse (or approximately k-sparse) vector in $\mathbb{R}^N$, and we want an algorithm which recovers $\mathbf{x}^{\#} = \Delta(\mathbf{y})$, so that the distance from $\mathbf{x}^{\#}$ to $\mathbf{x}$ is small in some sense. The original work from Candès, Romberg and Tao showed that if $\mathbf{x}$ is indeed sparse, then we can get not only approximate, but exact recovery from the algorithm:

$$\mathbf{x}^{\#} = \operatorname{argmin}\|\mathbf{z}\|_1 \quad \text{such that } A\mathbf{z} = \mathbf{y}. \tag{1.1}$$

This can be formulated as a linear programming problem and hence has polynomial complexity, a vast improvement over the similar “ideal” optimization problem—sometimes referred to as ℓ_0 optimization,

$$\mathbf{x}^{\#} = \operatorname{argmin}|\operatorname{supp}\mathbf{z}| \quad \text{such that } A\mathbf{z} = \mathbf{y}, \tag{1.2}$$

which is essentially a combinatorial search.

This method is very powerful, and has implications for many applications, as nearly sparse signals and images are very common in natural data [27]. However there are a few obstacles which come up, which have been the topic of much study over the past decades.

In their original papers, Candès, Romberg and Tao defined a number of properties of sensing matrices which are relevant to the compressive sensing problem, including the Null Space Property (NSP), Uniform Uncertainty Principle (UUP), Exact Reconstruction Principle (ERP), and Restricted Isometry Property (RIP) [17]. Some of these have faded into mostly historical significance, but NSP and RIP, in particular the latter, continue to get attention today.

Definition 1. (Null Space Property) An $m \times N$ matrix A is said to have the NSP of order k if for any $\boldsymbol{\nu} \in \ker A \setminus \{0\}$, $S \subset \{1, \ldots, N\}$ with $|S| \leq k$, $\|\boldsymbol{\nu}_S\|_1 < \|\boldsymbol{\nu}_{S^c}\|_1$.

An equivalent charactarization is to say that the null space of A, considered as a hyperplane in $\mathbb{R}^N$, when placed on a k-sparse face of the ℓ_1 unit sphere, is tangent to it. Hence it is clear that whenever $A\mathbf{x} = \mathbf{y}$ has a k-sparse solution, the algorithm in (1.1) will recover it.

While the NSP provides a good geometric intuition, in practice it is difficult to work with. Also, although the NSP guarantees exact reconstruction if the signal is known to be sparse, the introduction of even a small amount of error can make $\left\|\mathbf{x} - \mathbf{x}^{\#}\right\|_2$ very large, which is troublesome for applications. The RIP is introduced to deal with both of these problems.

Definition 2. (Restricted Isometry Property) An $m \times N$ matrix is said to have the RIP of order k with constant $\delta \in (0,1)$ if for any k-sparse $\mathbf{x} \in \mathbb{R}^N$,

$$(1-\delta)\|\mathbf{x}\|_2^2 < \|A\mathbf{x}\|_2^2 < (1+\delta)\|\mathbf{x}\|_2^2 .$$

We say that δ_k is the restricted isometry constant of A if δ_k is the smallest $\delta > 0$ such that A satisfies RIP of order k.

RIP is more restrictive than NSP, but in return we get much more robustness. For this reason much research has been done on the RIP and various recovery algorithms that depend on it in the past 15 years.

Both NSP and RIP represent sufficient conditions for accurate and unique recovery in (1.70). The goal of this dissertation is to extend the results of compressive sensing to a continuous domain, the two dimensional torus $\mathbb{T}^2$. In section 1.9 I outline the space and algorithm of interest, and my main results, which are discussed in more detail in Chapter 4. As I will show, in continuous space the question of uniqueness of the reconstruction is much more of an issue than in the finite dimensional case. However I will prove both a necessary and a sufficient condition for unique reconstruction.

1.2 Prony's Method

There have been some limited attempts to extend compressive sensing into the continuous domain with good results, but limited generality. An example of a widely studied problem is the recovery of point masses, where we assume the signal to be recovered is assumed to be of the form

$$f(x) = \sum_{k=1}^{M} \delta_{\mathbf{x}_k}(x). \tag{1.3}$$

Perhaps the most archetypal form of this problem dates back long before compressive sensing. Prony's method was known in the 18th century, but its ideas still see some use today [26, 54].

Prony's method is an algorithm for resolving discrete signals as in (1.3) on $\mathbb{R}$. However it is typically formulated on the frequency spectrum as a sampling problem. The goal is to recover a signal $\widehat{f} \in L^\infty(\mathbb{R})$ which is a finite sum of sinusoids,

$$\widehat{f}(t) = \sum_{k=1}^{M} B_k e^{i\omega_k t} \tag{1.4}$$

from $N = 2M$ evenly spaced samples,

$$\widehat{f}(\Delta n) = \sum_{k=1}^{M} B_k e^{i\omega_k \Delta n}, \tag{1.5}$$

where $\Delta > 0$ is the fixed distance between samples.

The algorithm comes from recognizing two key relationships. First, since $\widehat{f}$ is the sum of M exponentials, it is the solution to a particular linear difference equation of M variables. That is to say we have

$$\widehat{f}(\Delta n) = \sum_{k=1}^{M} P_k \widehat{f}(\Delta(n-k)). \tag{1.6}$$

Then the frequencies ω_k are related to this equation by the roots of the characteristic polynomial,

$$z^M - P_1 z^{M-1} - \cdots - P_M = \prod_{k=1}^{M} (z - e^{i\omega_k}). \tag{1.7}$$

The process of Prony's method consists of first calculating the coefficients P_k using (1.6),

$$\begin{pmatrix} F_{M-1} & F_1 & \cdots & F_0 \\ F_M & F_{M-1} & & F_1 \\ \vdots & & \ddots & \vdots \\ F_{N-1} & \cdots & & F_{N-M-1} \end{pmatrix} \begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_k \end{pmatrix} = \begin{pmatrix} F_M \\ F_{M+1} \\ \vdots \\ F_N \end{pmatrix}, \tag{1.8}$$

where $F_k = \widehat{f}(\Delta k)$. Second, factor the polynomial in (1.7) to find the frequencies

ω_k. And finally the magnitudes B_k are found by another linear equation,

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ e^{\omega_1 \Delta} & e^{\omega_2 \Delta} & & e^{\omega_M \Delta} \\ \vdots & & \ddots & \vdots \\ e^{\omega_1 \Delta M} & \cdots & & e^{\omega_M \Delta M} \end{pmatrix} \begin{pmatrix} B_1 \\ B_2 \\ \vdots \\ B_k \end{pmatrix} = \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_M \end{pmatrix}. \tag{1.9}$$

Prony's method works on two conditions: first, that the number of samples is at least twice the number of sinusoids; and second, that the sampling rate is at least the Nyquist frequency. The second point is slightly more subtle, but it is because during the polynomial factorization step we only solve for e^{ω_k}, not the frequencies themselves, hence they are non-unique up to a factor of $2\pi\Delta$.

Prony's method is seen in a broad variety of applications. Identifying and demixing sinusoidal signals is useful in many signal processing applications, including electromagnetics and antenna engineering [54]. It is frequently seen in the related, slightly more modern algorithms called matrix pencils, for finding generalized eigenvalues of matrix operators, in a manner similar to finding the frequencies of the sinusoids above [54].

1.3 Continuous Superresolution

Use of the total variation norm for basis pursuit in the continuous domain is very recent, inspired by the massive success of compressive sensing in the discrete case. The first work was from de Castro and Gamboa [25], who used Beurling's theory of minimal extrapolations to generalize some of the concepts of the NSP and random sensing matrices from compressive sensing. Notably they showed that for a sensing apparatus defined from a family of functions $\mathcal{F} = (u_k)_{k=1}^K$, there must be a solution to the optimization problem

$$\mu^\# = \underset{\mu \in \mathcal{M}[0,1]}{\operatorname{argmin}} \|\mu\|_{TV} \quad \text{such that } \langle u_k, \mu \rangle = \mathbf{y}_k, \qquad 1 \leq k \leq K, \tag{1.10}$$

that is supported in a finite set, which are the roots to a "generalized polynomial" $1 - \sum_k c_k u_k$. Under certain conditions on the family $\mathcal{F}$, these solutions may also be unique.

The use of Beurling's minimal extrapolations were expounded in a different way by Benedetto and Li [6], who studied the case of $\mathcal{F}$ a family of complex exponentials. This is equivalent to having prior knowledge of the Fourier transform. In this case the zero sets of "generalized polynomials" are roots of trigonometric polynomials.

The contribution of Benedetto and Li was giving quite sharp results for one particular situation [6]. Let $\mu \in \mathcal{M}(\mathbb{T}^n)$ and our prior knowledge be $\widehat{f}(m)$ for $m \in \Lambda \subset \mathbb{Z}^n$, not necessarily a rectangular box. We consider the set of minimal extrapolations as follows. If

$$\epsilon = \epsilon(f) = \inf\{\|\nu\|_{TV} \mid \nu \in \mathcal{M}(\mathbb{T}^2),\ \forall m \in \Lambda\ \widehat{\nu}(m) = \widehat{f}(m)\}, \tag{1.11}$$

then the set of minimal extrapolations is the set of extensions of $\widehat{f}$ which achieve the upper bound:

$$E = \{\nu \in \mathcal{M}(\mathbb{T}^n) \mid \forall m \in \Lambda \quad \widehat{\nu}(m) = \widehat{f}(m) \quad \text{and} \quad \|\nu\|_{TV} \text{ is minimal}\}. \tag{1.12}$$

They divide into three possible situations, based on the observation that for all $m \in \Lambda$, $|\widehat{f}(m)| \leq \epsilon$. Define $\Gamma = \{m \in \Lambda \mid |\widehat{f}(m)| = \|\nu\|\}_{TV}$. The cases are whether $|\Gamma|$ is 0, 1 or greater than 1. In other words, how frequently does $|\widehat{f}|$ achieve its maximal value and does $\|\widehat{f}\|_\infty = \|\mu\|_{TV}$?

At first glance these distinctions may be arbitrary, but we will see in section 1.6, particularly in the work of Carathéodory on positive definite functions [19], that in the case of positive definite extensions, sequences which achieve their maxima represent points on the boundary of the convex set of positive definite functions. Similarly here, functions $f : \Lambda \to \mathbb{C}$ will fall on the boundary of the set of functions f with $\epsilon(f) = \epsilon$. This is not insignificant; we can relate this property back to the NSP from compressive sensing—in the context of compressive sensing, a signal is sparse if it is in the boundary of a face of the ℓ_1 ball, which is key to why compressive sensing works.

The result is if $|\Gamma| \neq 1$, then we have results as Castro and Gamboa, that solutions must fall within the roots of a trigonometric polynomial. If $|\Gamma| \geq 2$, we have even better, that this level set must take the form of a finite lattice, making the calculation of the results very straightforward, and unique for sufficiently large Λ.

1.4 Resolution of Point Sources and Minimum Separation

Other work on point superresolution was done by Fernandez-Granda, et al., who have developed a nice investigation on the relationship of the recovery problem and the uncertainty principle [12, 13, 30]. One disadvantage of moving to continuous space is that there is no longer a straightforward relationship between the sparsity and the fineness of our signal. In other words, if we only have prior knowledge up to a finite band limit, then we can not be expected to resolve signals which have details which are arbitrarily fine. It is easy to construct examples which are Fernandez-Granda approaches this by adding a constraint on the minimum separation of the signal. If $\mu \in \mathcal{M}(\mathbb{R})$ and $\operatorname{supp} \mu = S$, the minimum separation of μ is

$$\Delta(\mu) = \inf_{t \neq t' \in S} \left| t - t' \right| . \tag{1.13}$$

In [13], Candès and Fernandez-Granda show that minimum separation requirements are enough to find uniqueness. Remarkably, this doesn't require any modification of the search algorithm. Standard basis pursuit is enough to find these discrete signals, even though it has no "knowledge" of the minimum separation.

Their main theorem is as follows.

Theorem 1. Let $\mu \in \mathcal{M}(\mathbb{R})$ satisfy (1.13) with $\Delta(\mu) \geq D$. If the values of $\widehat{\mu}(k)$ are known for $|k| \leq f_c$, then if $f_c \geq 128$ and

$$\Delta(\mu) \geq \frac{2}{f_c}, \tag{1.14}$$

then μ is the unique solution to the basis pursuit algorithm,

$$\mu^{\#} = \underset{\mu \in \mathcal{M}[0,1]}{\operatorname{argmin}} \|\mu\|_{TV} \quad \text{such that } \widehat{\mu}(k) = \mathbf{y}_k \quad |k| \leq f_c. \tag{1.15}$$

The same result holds with a less restrictive constant for higher dimensions $\mathbb{R}^n$ as well, and it also holds up to stability in the face of noisy measurements. Later work from Fernandez-Granda has improved upon these constants, going as low as 1.26 in the 1-dimensional case [30].

1.5 Attempts at Superresolving Non-discrete Measures

There are only a few results that attempt to perform signal recovery on signals that are not discrete. They tend to lean strongly on the works prior described. A generalization from Unser, et al. [59] lets us perform superresolution on classes of splines, functions which behave like discrete functions when applied to a differential operator. He shows that splines, rather than deltas, are the archetypal solution for a slightly altered problem,

$$x^{\#} = \underset{x \in X}{\operatorname{argmin}} \|Lx\|_1 \quad \text{such that } A(x) = \mathbf{y}, \tag{1.16}$$

where L is a differential operator, $X = \{x \in BV(\mathbb{R}^n) \mid Lx \in \mathcal{M}(\mathbb{R}^n)\}$, and $A\colon X \to \mathcal{R}^m$ is a linear measurement function. The theory is very similar to the above work, but it demonstrates the flexibility of the theory nicely.

Another creative approach was by Ongie [48, 49], who looked to generalize

Prony's method to higher dimensions. Rather than starting with a mix of sinusoids, they recognize a key step of Prony's method is to factor the trigonometric polynomial in (1.7). Let

$$\phi(\mathbf{x}) = \sum_{k=1}^{N} e^{i\boldsymbol{\omega}_k \cdot \mathbf{x}}, \tag{1.17}$$

be a trigonometric polynomial in on $\mathbb{R}^2$. Let $f \in L^\infty(\mathbb{R}^2)$ be of the form

$$f = \sum_{k=1}^{M} \alpha_k \chi_{U_k}, \tag{1.18}$$

where U_k are disjoint open subsets of $\mathbb{R}^2$ and $\cup_k U_k = \{\mathbf{x} \in \mathbb{R}^2 \mid \phi(\mathbf{x}) \neq 0\}$. Then by recognizing that $\phi\, \mathrm{D} f \equiv 0$, it is possible to perform a similar calculation to the 1-dimensional case and recover the support of f by factoring a trigonometric polynomial. Bezout's theorem puts a bound on the number of distinct U_k, so a system of equations finds the weights α_k, and we have theoretically guaranteed recovery.

1.6 Positive Definite Functions

Carathéodory began studying positive definite functions in the early 20th century, at around the same time he was doing work on convex geometry. The context in which such functions were discovered was in complex function theory. The connection is not immediately apparent, but as we will see there are deep similarities.

There we say that a positive definite function on the unit disc is one of the form

$$f(z) = 1 + \sum_{k=1}^{\infty}(a_k + ib_k)z^k, \tag{1.19}$$

which is analytic and has positive real part. Remarkably, Carathéodory gave a characterization of these functions that as we will see is useful for the problem we are interested in today [19]. He said that such a function is positive definite if and only if for each $n \in \mathbb{N}$ the coordinates $(a_1, b_1, a_2, \cdots, a_n, b_n)$ fall in the convex hull generated by the curve $(cos(\alpha), sin(\alpha), cos(2\alpha)\cdots, cos(n\alpha), sin(n\alpha))$, for $0 \leq \alpha \leq 2\pi$. Call this set Q_n, and say that a point in Q_n represents a function f if f has those coefficients for $0 \leq k \leq n$.

The main line of his proof contains many of the same beats as the other results in this paper. He recognizes the convexity of the set of positive definite functions P. He identifies that the interior of P is nonempty, or in other words there are functions f for which an analytic neighborhood of f falls within P. Therefore Q_n cannot fall in a lower-dimensional hyperplane of $\mathbb{R}^{2n}$. And then he proves using complex analytic methods that the boundary of P are all rational functions, which are generated by those of the form

$$f(z) = \frac{1}{1 - \overline{\alpha}z} \quad |\overline{\alpha}| = 1, \tag{1.20}$$

which are represented by points on the curve $(cos(\alpha), sin(\alpha), cos(2\alpha), \cdots, cos(n\alpha), sin(n\alpha))$, $0 \leq \alpha \leq 2\pi$.

Toeplitz [58] later showed that an equivalent characterization of sequences $(a_1, b_1, \cdots)$ is that

$$\sum_{k,l=1}^{n} c_{k-l} z_k \overline{z}_l \geq 0 \quad n \in \mathbb{N},\ z_k \in \mathbb{C}, \tag{1.21}$$

where $c_0 = 2$, $c_k = a_k - ib_k$, $d_{-k} = \overline{d}_k$. This definition is the one that was used commonly later in the century through today, and as we will see it has inspired a host of problems of this form in more generality and under various conditions [56]. Indeed, as we will see it is quite relevant to the total variation optimization problem in different settings as well.

Definition 3. Given a group G and a symmetric subset $V \subset G$, with $0 \in V$, we say $f\colon V \to \mathbb{C}$ is positive definite if

$$\sum_{k,l=1}^{n} f(x_k x_l^{-1}) c_i \overline{c}_j \geq 0 \tag{1.22}$$

for all sequences $\Gamma = (x_j)_{j=1}^{n}$ such that $\Gamma - \Gamma \subset V$, and sequences $(c_j)_{j=1}^{n} \subset \mathbb{C}$.

For $G = \mathbb{Z}$ we get the definition above studied by Carathéodory and Toeplitz, and for $G = \mathbb{R}$ we get a situation that has been studied extensively, and which has particular interest in harmonic analysis because of Bochner's theorem, which relates them to the Fourier transform of positive measures.

Theorem 2. (Bochner) For any locally compact abelian group G with dual group $\widehat{G}$, a function $f\colon G \to \mathbb{C}$ is positive definite if and only if there exists a unique nonnegative measure $\mu \in \mathcal{M}(\widehat{G})$ such that

$$f(x) = \int_{\widehat{G}} e^{2\pi i x\xi} \, \mathrm{d}\mu(\xi). \tag{1.23}$$

The result is named after Salomon Bochner, who proved it first in 1932 for the case $G = \mathbb{R}$, and later generalized his result to higher dimensions [8, 9, 56]. Although, Bochner's result was actually inspired by Herglotz, who proved the theorem for the case $G = \mathbb{Z}$ in 1911. I will give details on both Herglotz and Bochner's Theorems in the following chapter, but first I will give a few other properties of positive definite functions.

Proposition 1. Let G be a locally compact abelian group, let $V = \Gamma - \Gamma \subset G$ and let $f : V \to \mathbb{C}$ be a positive definite function.

1. $f(0) \geq 0$
2. $f(-x) = \overline{f(x)}$
3. $\left|f(x)\right| \leq f(0)$

These are all easily verified from the definitions above, and are standard in the literature [56]. The following is a standard but slightly less trivial result which was first shown by Artjomenko [2].

Proposition 2. If f is positive definite on $\mathbb{R}$ and continuous at 0 then it is uniformly continuous.

Proof. Let $|x - y| < \delta$. Choosing $\Gamma = \{x, y, 0\}$, it follows from the definition of positive definiteness that for any choice of $\{c_x, c_y, c_0\} \subset \mathbb{C}$,

$$(|c_0|^2 + |c_x|^2 + |c_y|^2) f(0) + 2\operatorname{Re}\left[c_x \bar{c}_0 f(x) + c_y \bar{c}_0 f(y) + c_x \bar{c}_y f(x-y)\right] \geq 0. \quad (1.24)$$

With the choice $c_0 = 1$, $c_x = -c_y = |f(x) - f(y)| / (f(x) - f(y))$, and using the fact that $2f(0) \geq |f(x) - f(y)|$, we get

$$3f(0) - 2\operatorname{Re} f(x-y) \geq 2|f(x) - f(y)| \quad (1.25)$$

$$2(f(0) - \operatorname{Re} f(x-y)) \geq \frac{3}{2}|f(x) - f(y)|. \quad (1.26)$$

Hence we have a uniform bound on the continuity of f. $\square$

Other regularity results follow from positive definiteness as well. In a similar manner to the previous result, it is possible to bound the derivatives of a positive definite function.

Theorem 3. If for some $k \in \mathbb{N}$ a positive definite function f defined on $\mathbb{R}$ is $2k$-times differentiable at 0, then f is $2k$-times differentiable on $\mathbb{R}$ as well.

The following on analytic positive definite functions was proven by each of Lévy and Raikov independently [42, 51].

Theorem 4. If f is analytic and positive definite in a symmetric interval $(-a, a)$, then there exist positive constants $\alpha, \beta \in (0, \infty]$ such that f can be extended to a function holomorphic on the horizontal strip

$$\{z \in \mathbb{C} \mid -\alpha < \operatorname{Im} z < \beta\}. \quad (1.27)$$

And finally, more recently Sasvári has proven a similar result for measurability [55].

Theorem 5. If f is positive definite, and it is measurable on a symmetric interval $(-A, A)$, then it is measurable on all of $\mathbb{R}$.

1.7 Bochner's Theorem

I will not prove the theorem in full generality but I will provide details for the two most relevant cases, $G = \mathbb{Z}$ (this case is often referred to as Herglotz's theorem) and $G = \mathbb{R}$. The proofs given are not performed as in Bochner's and Herglotz's original papers but are updated using distribution theory. The proofs as given are based on a review from Maruyama [46]. Before I proceed to the proofs, I will make a short statement on distribution theory that is necessary for them.

We will use distribution theory somewhat loosely in the upcoming section for the purpose of being as general as possible while keeping the use of notation reasonable. We will refer to $\mathcal{S}(G)$ as a set of test functions appropriate for G. For example if $G = \mathbb{R}^n$ then the natural choice is the Schwartz functions. For $G = \mathbb{T}^n$ we take the periodic summations $\sum_n \phi(x + nT)$ of Schwartz functions. The particular choice for other groups G is not important, as the properties that are necessary will be clear in the proofs of the following results, and we will focus our results in this paper on the two cases above. Given our space of test functions $\mathcal{S}(G)$, $\mathcal{S}'(G)$ is the set of distributions. We can extend the Fourier transform to $\mathcal{S}'(G)$ in a standard way by Parseval's formula. For $T \in \mathcal{S}'(G)$, $\mathcal{F}T$ is defined on $\mathcal{F}(\mathcal{S}(G))$ by

$$\mathcal{F}T(\mathcal{F}\phi) = T(\phi) \quad \forall \phi \in \mathcal{S}(G). \tag{1.28}$$

Lemma 6. Let $T \in \mathcal{S}'(G)$ be a distribution. Assume for all $\phi \in \mathcal{S}(G)$ with $\phi \geq 0$, $T(\phi) \geq 0$. Then T is a positive measure on G.

Proof. Assume that T is a positive distribution, as given above. T is a measure if and only if it is a positive linear functional on $C_0(G)$. Because $\mathcal{S}(G)$ is dense in $C_0(G)$, and T is positive, all that is necessary is to show that it is continuous with respect to the supremum norm. Assume to the contrary, that there is a sequence $(\phi_k)_{k=0}^n$ of positive functions in $\mathcal{S}(G)$ that is uniformly bounded by a constant M, but for which $T(\phi_k) \to \infty$. Assume without loss of generality as well that each ϕ_k has compact support. We lose no generality from doing this because compactly supported functions are dense in both $\mathcal{S}(G)$ and $C_0(G)$. Define a sequence of smooth compactly supported bump functions $\eta_k \in \mathcal{S}(G)$ such that for all k, $0 \leq \eta_k \leq 1$, $\eta_k \equiv 1$ on the support of ϕ_k, and all derivatives of η_k are uniformly bounded (Note that this construction is valid for Schwartz functions).

Now see that $T(\eta_k \phi_k) = T(\phi_k) \to \infty$. But since $0 \leq \phi_k \leq M$,

$$\eta_k(M - \phi_k) \geq 0 \quad \forall k \in \mathbb{N}. \tag{1.29}$$

Because the derivatives of η_k are uniformly bounded we have that

$\sup|T(M\eta)| < \infty$, so we can conclude that

$$\lim_{k\to\infty} T(\eta_k(M-\phi_k)) = \lim_{k\to\infty} T(M\eta_k) - T(\eta_k\phi_k) = -\infty. \tag{1.30}$$

But this result conflicts with (1.29). Hence T is a bounded linear functional on $C_0(G)$, and therefore is identified with a measure $\mu \in \mathcal{M}(G)$ by the Riesz representation theorem. □

The following proofs are from [46].

1.7.1 $G = \mathbb{Z}$

Proof. (Herglotz) First, let $\mu \in \mathcal{M}(\mathbb{T})$ be positive and

$$f(n) = \int_0^1 e^{2\pi i n\xi}\,\mathrm{d}\mu(\xi) \quad n \in \mathbb{Z}. \tag{1.31}$$

Let $(c_k)_{k=1}^n \subset \mathbb{C}$ and let $1 \le k < l \le n$. For all $x \in \mathbb{T}$ we can write

$$\sum_{a,b=\{l,k\}} e^{2\pi i(a-b)x} c_a \overline{c}_b = e^{2\pi i(k-l)x} c_k \overline{c}_l + e^{2\pi i(l-k)x} c_k \overline{c}_l + |c_k|^2 + |c_l|^2 \tag{1.32}$$

$$= 2\,\mathrm{Re}(e^{2\pi i(k-l)x} c_k \overline{c}_l) + |c_k|^2 + |c_l|^2 \tag{1.33}$$

$$\ge |c_k|^2 + |c_l|^2 - 2|c_k||c_k| \tag{1.34}$$

$$\ge 0. \tag{1.35}$$

We can rewrite (1.22) by exchanging the order of integration:

$$\sum_{k,l=1}^{n} c_k \overline{c}_l \int_0^1 e^{2\pi i(k-l)\xi}\, \mathrm{d}\mu(\xi) = \int_0^1 \sum_{k=1}^{n} \sum_{l=k}^{n} \sum_{a,b=\{l,k\}} e^{2\pi i(a-b)\xi} c_a \overline{c}_b \, \mathrm{d}\mu(\xi), \tag{1.36}$$

which as shown in (1.32) is nonnegative.

We have shown that the Fourier transform of a positive measure is positive definite. Now we must show the converse. Let $f\colon \mathbb{Z} \to \mathbb{C}$ be positive definite. By one of the properties in proposition 1, $\left|f(n)\right| \leq f(0)$ for all $n \in \mathbb{Z}$. This is enough to demonstrate that $f = \mathcal{F}T$ for some distribution $T \in \mathcal{S}'(G)$. Note that because $e^{-2\pi i n x} \in \mathcal{S}(\mathbb{T})$, we can quickly calculate that $f(n) = T(e^{2\pi i n x})$. We want to show that T is positive.

Let q be a trigonometric polynomial of order N,

$$q(x) = \sum_{n=-N}^{N} c_k e^{2\pi i n x}. \tag{1.37}$$

We can write

$$\left|q(x)\right|^2 = \sum_{k,l=-N}^{N} c_k \overline{c}_l e^{2\pi i(k-l)x}, \tag{1.38}$$

and then we get the formula

$$T\left(\left|q(x)\right|^2\right) = \mathcal{F}T\left(\mathcal{F}\left(|q(x)|^2\right)\right) = \sum_{k,l=-N}^{N} f(l-k) c_k \overline{c}_l \geq 0. \tag{1.39}$$

As the trigonometric polynomials are dense in $\mathcal{S}(\mathbb{T})$, we can say then that for generic $q \in \mathcal{S}(\mathbb{T})$, $T(|q(x)|^2) \geq 0$. So for arbitrary $p \in \mathcal{S}(\mathbb{T})$ such that $p \geq 0$, choose

$q(x) = \sqrt{p(x)}$ and it immediately follows that $T(p) \geq 0$.

By lemma 6, we can conclude that T is a positive measure on $\mathbb{T}$, and the proof is complete. $\square$

1.7.2 $G = \mathbb{R}$

Theorem 7. A function $f\colon \widehat{\mathbb{R}} \to \mathbb{C}$ is the Fourier transform of a finite positive measure $\mu \in \mathcal{M}(\widehat{\mathbb{R}})$ if and only if it is positive definite and continuous at 0.

Proof. The proof begins similar to that of Herglotz' theorem. For $\mu \in \mathcal{M}(\mathbb{T})$ positive and

$$f(x) = \int_{\mathbb{R}} e^{2\pi i x\xi} \,\mathrm{d}\mu(\xi), \tag{1.40}$$

we must show that (1.22) holds for any choice of $(c_k)_{k=1}^n \subset \mathbb{C}$, and any $(x_k)_{k=1}^n \subset \mathbb{R}$. Write

$$\sum_{k,l=1}^{n} c_k \overline{c}_l \int_{\mathbb{R}} e^{-2\pi i(x_k - x_l)\xi} \,\mathrm{d}\mu(\xi) = \int_{\mathbb{R}} \sum_{k=1}^{n} \sum_{l=k}^{n} \sum_{a,b=\{l,k\}} e^{2\pi i(a-b)\xi} c_a \overline{c}_b \,\mathrm{d}\mu(\xi). \tag{1.41}$$

The rightmost sum is positive as shown before, so the function $\widehat{\mu}$ is positive definite. Note that we may justify changing the order of integration because μ is finite. By the same reasoning we may differentiate $\widehat{\mu}$ under the integral sign to show that it is differentiable and therefore continuous.

Now assume that $f\colon \widehat{\mathbb{R}} \to \mathbb{C}$ is continuous and positive definite. We have that $\left|f(x)\right| \leq f(0) < \infty$ because it is positive definite. So $f \in \mathcal{S}'(\widehat{\mathbb{R}})$ and we can identify a distribution T such that $\check{f} = T \in \mathcal{S}'(\mathbb{R})$. We must show that T is positive.

Let $\phi \in \mathcal{S}(\widehat{\mathbb{R}})$ be a test function. Then it follows from (1.22) that

$$\iint_{\widehat{\mathbb{R}}^2} f(x-y)\phi(x)\overline{\phi}(y)\,\mathrm{d}x\,\mathrm{d}y \geq 0. \tag{1.42}$$

If this is not clear, then consider approximating the integral by a Riemann sum, which will be of the form in (1.22) irrespective of the choice of partition or ϕ. Through a change of variable we can write

$$\iint_{\widehat{\mathbb{R}}^2} f(y)\phi(x)\overline{\phi}(x-y)\,\mathrm{d}x\,\mathrm{d}y \geq 0. \tag{1.43}$$

Let $\Phi(x) = \int \phi(x)\overline{\phi}(x-y)\,\mathrm{d}y$, and note that $\langle f, \Phi\rangle \geq 0$. Calculate the following for $\check{\Phi}$:

$$\begin{aligned}
\check{\Phi}(x) &= \int_{\widehat{\mathbb{R}}} \Phi(\xi)e^{2\pi i x\xi}\,\mathrm{d}\xi \\
&= \iint_{\widehat{\mathbb{R}}^2} \phi(\zeta)\overline{\phi}(\zeta-\xi)\,\mathrm{d}\zeta e^{2\pi i x\xi}\,\mathrm{d}\xi \\
&= \iint_{\widehat{\mathbb{R}}^2} \phi(\zeta)\overline{\phi}(\xi)e^{2\pi i x\zeta}e^{-2\pi i x\xi}\,\mathrm{d}\zeta\,\mathrm{d}\xi \\
&= \int_{\widehat{\mathbb{R}}} \phi(\zeta)e^{2\pi i x\zeta}\,\mathrm{d}\zeta \int_{\widehat{\mathbb{R}}} \overline{\phi}(\xi)e^{-2\pi i x\xi}\,\mathrm{d}\xi \\
&= \left|\check{\phi}(x)\right|^2.
\end{aligned}$$

We can conclude that if $p(x) \in \mathcal{S}(\mathbb{R})$ with $p(x) \geq 0$, then by choosing $\check{\phi}(x) = \sqrt{p(x)} \in \mathcal{S}(\mathbb{R})$, $T(p) = T(|\check{\phi}|^2) = \langle f, \Phi\rangle \geq 0$. Hence T is a positive functional, and by lemma 6 there exists a unique finite positive measure $\mu \in \mathcal{M}(\mathbb{R})$ such that

$$T(\phi) = \int_{\mathbb{R}} \phi \, \mathrm{d}\mu \quad \forall \phi \in \mathcal{S}(\mathbb{R}). \tag{1.44}$$

□

1.7.3 Additional Examples and Applications of Bochner's Theorem

Notably, there are not many examples of functions which verify Bochner's theorem through an easy computation. One of the few is $f(x) = e^{2\pi i \alpha x} = \mathcal{F}\delta_\alpha$, for $\alpha \in \widehat{G}$ For $(c_k)_{k=1}^n \subset \mathbb{C}$ and $(x_k)_{k=1}^n \subset G$,

$$\sum_{k,l=1}^{n} c_k \overline{c}_l e^{2\pi i \alpha (x_k - x_l)} = \sum_{k=1}^{n} \sum_{l=k}^{n} \sum_{a,b=\{l,k\}} c_a \overline{c}_b e^{2\pi i \alpha (a-b)} \geq 0. \tag{1.45}$$

This is suggestive of Carathéodory's work on holomorphic functions on the disc. Recall that his characterization of positive definite functions identified that a certain set of rational functions form the boundary for the convex set of positive definite functions [19]. Then Toeplitz made the connection between holomorphic functions

$$f(z) = \sum_{k=0}^{\infty} c_k z^k, \tag{1.46}$$

and positive definite sequences $(c_k)_{k=0}^{\infty} \subset \mathbb{C}$. Formally we can write that on the boundary of the unit circle, $f(e^{-2\pi i\theta}) = \sum c_k e^{-2\pi i k\theta} = \mathcal{F}(c_k)(\theta)$. And see that if

$c_k = e^{2\pi i\alpha k}$ as in (1.45), then $f(z)$ is a rational function of the form

$$f(z) = \frac{1}{1 - \alpha z}, \tag{1.47}$$

which as we've seen generate the set of positive definite functions as their convex hull.

The theoretical interest in Bochner's theorem is obvious, as it gives a characterization of an important class of Fourier transforms. But it also has yielded more applications through its life. A classical example due to Khintchin comes from probability theory of continuous random processes [37].

Let $(\Omega, \varepsilon, \mathcal{P})$ be a probability space, and $X_t(\omega)\colon \mathbb{R} \times \Omega \to \mathbb{C}$ be a stochastic process. That is to say X is measurable in $\varepsilon \times \mathcal{L}$ and for all $t \in \mathbb{R}$, X_t is a random variable on Ω. X is said to be second order stationary if second moments exist, and both the expectation $\mathbb{E}(X_t)$ and covariance $\mathrm{Cov}(X_t, X_{t+\tau})$ do not depend on t. Define the covariance function

$$C(\tau) = \mathrm{Cov}(X_t, X_{t+\tau}). \tag{1.48}$$

As it turns out, $C(\tau)$ is a positive definite function. In fact, since $\mathrm{Cov}(\cdot, \cdot)$ forms an inner product, we can calculate for any pair (s,t), and complex numbers (c_s, c_t),

$$
\begin{aligned}
&\sum_{a,b=\{s,t\}} c_a \overline{c}_b \operatorname{Cov}(X_a, X_b) \\
&\qquad = |c_s|^2 \operatorname{Cov}(X_a, X_a) + 2 \operatorname{Re} c_s \overline{c}_t \operatorname{Cov}(X_a, X_b) + |c_t|^2 \operatorname{Cov}(X_b, X_b) \\
&\qquad \geq 0.
\end{aligned}
$$

So for any collection $(x_k)_{k=1}^n \subset \mathbb{R}$, $(c_k)_{k=1}^n \subset \mathbb{C}$, we have

$$
\begin{aligned}
\sum_{k,l=1}^{n} c_k \overline{c}_l C(x_k - x_l) &= \sum_{k=1}^{n} \sum_{l=k}^{n} \sum_{a,b=\{k,l\}} c_a \overline{c}_b C(x_a - x_b) \\
&= \sum_{k=1}^{n} \sum_{l=k}^{n} \sum_{a,b=\{k,l\}} c_a \overline{c}_b \operatorname{Cov}(X_{x_b}, X_{x_a}) \\
&\geq 0.
\end{aligned}
$$

By application of Bochner's theorem, we have that for any such X, there exists a measure $\nu \in \mathcal{M}(\mathbb{R})$ such that

$$
C(\tau) = \int_{\mathbb{R}} e^{-2\pi i t\tau} \, \mathrm{d}\nu(t). \tag{1.49}
$$

1.8 Positive Definite Extensions

What we call the positive definite extension problem was first posed explicitly by Krein in 1940 [38]. For a locally compact abelian group G and symmetric subset V, as in definition 3, denote the set of positive definite functions on V by $P(V)$.

The question is, can we extend every function $f \in P(V)$ to a function in $P(G)$?

The case $G = \mathbb{Z}$ has the characterization from Carathéodory [19], which as we've seen predates Krein's question by thirty years. For $V_n = \{k \in \mathbb{Z} \,|\, |k| \leq n\}$, he gave a characterization of those sequences in $P(V_n)$ which have extenstions to $P(\mathbb{Z})$. Recall that the point $(a_1, b_1, \cdots, a_n, b_n)$ is said to be the geometric representative of a function $f(z) = 1 + \sum_k a_k cos(2\pi k) + b_k sin(2\pi k)$ if and only if it falls within the convex hull Q_n of the curve defined by

$$(\cos(2\pi\alpha), \sin(2\pi\alpha), \cdots, \cos(2n\pi\alpha), \sin(2n\pi\alpha)), \quad 0 \leq \alpha < 2\pi. \tag{1.50}$$

Krein's problem is related by a simple change of variables. In fact in a later paper Carathéodory showed that the two sets coincide [20].

Theorem 8. (Carathéodory) Every positive definite function on V_n can be extended to a positive definite function on $\mathbb{Z}$.

Proof. A function f defined on V_n is positive definite if and only if the matrix $A = (\mathbf{a}_{ij})_{i,j=0}^{n}$, given by $\mathbf{a}_{ij} = f(i-j)$, is positive definite, i.e. for all $\mathbf{c} \in \mathbb{C}^{n+1}$, $\mathbf{c}^{\intercal} A\mathbf{c} \geq 0$. Consider for some $z \in \mathbb{C}$ the matrix formed by extending A by one row and column in the following way:

$$A_z = \begin{pmatrix} f(0) & f(1) & \cdots & f(n) & z \\ \overline{f}(1) & f(0) & \cdots & f(n-1) & f(n) \\ \vdots & & \ddots & & \vdots \\ \overline{f}(n) & \overline{f}(n-1) & \cdots & f(0) & f(1) \\ \overline{z} & \overline{f}(n) & \cdots & \overline{f}(1) & f(0) \end{pmatrix} \tag{1.51}$$

A_z is a Hermitian matrix so we can say a few things. First, that it has a unitary decomposition, and therefore $n+1$ distinct real eigenvalues $(\lambda_i)_{i=0}^{n+1}$ (including multiplicity) and orthogonal eigenvectors $(\mathbf{v}^i)_{i=0}^{n+1}$.

We know by assumption that for $\mathbf{c} = (\mathbf{c}_i)_{i=0}^{n+1} \in \mathbb{C}^{n+2}$, if $\mathbf{c}_{n+1} = 0$, then $\mathbf{c}^\intercal A_z \mathbf{c} \geq 0$. Consider the $n+1$-dimensional subspace $K = \{\mathbf{c} \in \mathbb{C}^{n+2} \mid \sum_i \mathbf{c}_i \mathbf{v}^i_{n+1} = 0\}$. For all $\mathbf{c} \in K$, we have

$$\left(\sum_{i=0}^{n+1} \mathbf{c}_i \mathbf{v}^i\right)^\intercal A_z \left(\sum_{i=0}^{n+1} \mathbf{c}_i \mathbf{v}^i\right) = \sum_{i=0}^{n+1} |\mathbf{c}_i|^2 \lambda_i \geq 0. \tag{1.52}$$

Because this is true for $\mathbf{c}$ in an $(n+1)$-dimensional space, we can conclude that at least $n+1$ of the eigenvalues λ_i are nonnegative. Since A_z is positive definite if and only if each of its eigenvalues are nonnegative, it follows that since at most one λ_i is nonpositive, A_z is positive definite if and only if $|A_z| = \prod \lambda_i \geq 0$.

The problem then reduces to solving a polynomial inequality $|A_z| \geq 0$, which can be computed to have a solution of the form

$$|z - z_0| \leq r \tag{1.53}$$

for some $z_0 \in \mathbb{C}$, $r \geq 0$.

We can continue this process for all n to inductively show that an extension exists to $P(\mathbb{R})$. □

In Krein's paper, he proved the logical next step for the problem by extending such a result to $G = \mathbb{R}$ [38].

Theorem 9. (Krein) Any continuous positive definite function f on an interval $(-A, A)$ can be extended to a positive definite function on $\mathbb{R}$.

I'll give a sketch of the proof courtesy of Sasvári [56]. Let L_A be the set of functions with the form

$$\phi(t) = \sum_{k=1}^{n} c_k e^{2\pi i x_k t}, \tag{1.54}$$

where $c_k \in \mathbb{C}$ and $-A < x_k < A$ for all k. Then define the functional $\Phi\colon L_A \to \mathbb{C}$ by

$$\Phi(\phi) = \sum_{k=1}^{n} c_k f(x_k). \tag{1.55}$$

It follows that Φ is a nonnegative functional, and can therefore be extended continuously to a nonnegative functional on $\mathcal{L}^\infty(\mathbb{R})$. Then by setting $F(x) = \Phi(e^{2\pi i x t})$, it can be shown that $f(x) = F(x)$ on $(-A, A)$, and that

$$\sum_{k,l=1}^{n} c_k \bar{c}_l F(x_k - x_l) = \Phi\left(\left|\sum_{k=1}^{n} c_k e^{2\pi i x_k t}\right|^2\right) \geq 0, \tag{1.56}$$

for any collections $c_k \in \mathbb{C}$, $x_k \in \mathbb{R}$.

1.8.1 Extensions in Higher Dimensions

When we go to higher dimensions the situation quickly becomes much more complicated. Consider the simple situation where $G = \mathbb{Z}^n$, $n \geq 2$. Ten years later it was proven by Calderón and Pepinsky that a similar extension is not possible.

Theorem 10. For $n > 1$, there exist positive definite functions defined on $\{(k_1, \cdots, k_n) \in \mathbb{Z}^n \mid |k_i| \leq M\}$ such that they cannot be extended to positive definite functions on $\mathbb{Z}^n$.

And another decade later it was Rudin who expounded and extended their result to $\mathbb{R}^n$ [52].

Theorem 11. (Rudin) For $n > 1$, there exist continuous positive definite functions $f\colon I^n \to \mathbb{C}$, where $I = (-a, a)$ is a symmetric interval, such that f cannot be extended to a positive definite function on $\mathbb{R}^n$.

The proof is based on a relationship between positive definite functions and sums of squares of polynomials, a problem that had been studied by Hilbert in the 19$^{\text{th}}$ century. Hilbert discovered that there exist nonnegative polynomials of three variables that are not sums of squares, and later examples for $n = 2$ were found as well [33].

Theorem 12. (Hilbert) For a given n and $d \in \mathbb{N}$, every nonnegative polynomial of degree d in n variables can be expressed as a finite sum of squared polynomials if and only if one of the following are true:

- $n = 1$
- $d = 2$
- $n = 2$ and $d = 4$.

Given a finite set $\Gamma \subset \mathbb{Z}^n$, define a subset of the trigonometric polynomials which are supported on $V = \Gamma - \Gamma$. X_Γ will be the set of all trigonometric polynomials of the form

$$f(\boldsymbol{\gamma}) = \sum_{\mathbf{x} \in V} c(\mathbf{x}) e^{-2\pi i \boldsymbol{\gamma} \cdot \mathbf{x}}. \tag{1.57}$$

We will call these Γ-polynomials. It is obvious that the coefficcients $c(\mathbf{x})$ can be retrieved from the Fourier transform of f. Define P_Γ as the subset of X_Γ of nonnegative trigonometric polynomials. Then let Q_Γ be the set of sums of squares—that is functions of the form

$$f(\boldsymbol{\gamma}) = \sum_{j=1}^{J} \left| g_j(\boldsymbol{\gamma}) \right|^2, \quad g_j \in X_\Gamma. \tag{1.58}$$

Immediately we have that $Q_\Gamma \subset P_\Gamma \subset X_\Gamma$.

The goal is to establish a correspondence between Q_Γ, P_Γ and the positive definiteness of $c(\boldsymbol{\gamma})$. Then Hilbert's result will tell us when positive definite extensions are possible. With that in mind refer to $\mathrm{PD}(V)$ and $\mathrm{PD}(\mathbb{Z}^n)$ as the set of positive definite functions on V and $\mathbb{Z}^n$ respectively. Given a function ϕ defined on V such that $\phi(-\mathbf{x}) = \overline{\phi}(\mathbf{x})$, we can define a real functional on X_Γ by

$$L_\phi(f) = \sum_{\mathbf{x}\in V} \phi(\mathbf{x})\widehat{f}(\mathbf{x}). \tag{1.59}$$

In fact, since X_Γ is finite dimensional, it is easy to see that every linear functional on X_Γ is of this form for some ϕ. The following two results establish the connection to positive definite sequences.

Proposition 3. A function ϕ on V is in $\mathrm{PD}(\Gamma)$ if and only if $L_\phi(f) \geq 0$ for every $f \in Q_\Gamma$.

Proof. Let

$$g(\boldsymbol{\gamma}) = \sum_{\mathbf{x}\in\Gamma} c(\mathbf{x})e^{-2\pi i\boldsymbol{\gamma}\cdot\mathbf{x}}. \tag{1.60}$$

Then we have

$$\begin{aligned}
\left|g(\boldsymbol{\gamma})\right|^2 &= \sum_{\mathbf{x},\mathbf{y}\in\Gamma} c(\mathbf{x})\overline{c}(\mathbf{y})e^{-2\pi i\boldsymbol{\gamma}\cdot(\mathbf{x}-\mathbf{y})} \\
L_\phi(|g|^2) &= \sum_{\mathbf{x},\mathbf{y}\in\Gamma} c(\mathbf{x})\overline{c}(\mathbf{y})\phi(\mathbf{x}-\mathbf{y}).
\end{aligned}$$

So $L_\phi(|g|^2)$ is positive regardless of the choice of g if and only if ϕ is positive definite. By the linearity of L_ϕ, the same is true for any $f \in Q_\Gamma$. $\square$

Proposition 4. A function ϕ defined on V can be extended to a member of $\mathrm{PD}(\mathbb{Z}^n)$

if and only if $L_\phi(f) \geq$ for all $f \in P_\Gamma$.

Proof. If $\phi \in \mathrm{PD}(\mathbb{Z}^n)$ then by Bochner's theorem there is a nonnegative measure $\mu \in \mathcal{M}(\mathbb{T}^n)$ such that

$$\phi(\mathbf{x}) = \int_{\mathbb{T}^n} e^{2\pi i \boldsymbol{\gamma}\cdot\mathbf{x}}\, \mathrm{d}\mu(\boldsymbol{\gamma}). \tag{1.61}$$

If $f \in X_\Gamma$ then by Parseval's formula we have

$$L_\phi(f) = \sum_{\mathbf{x}\in V} \widehat{f}(\mathbf{x})\phi(\mathbf{x}) = \int_{\mathbb{T}^n} f\, \mathrm{d}\mu. \tag{1.62}$$

Then $L_\phi(f) \geq 0$ if $f \in P_\Gamma$.

For the other implication, suppose $L_\phi(f) \geq 0$ for all $f \in P_\Gamma$. Without loss of generality assume $|f| \leq 1$. Because $L_\phi(1) = \phi(0)$ we have that $L_\phi(f) = L_\phi(1-f) - \phi(0) \geq -\phi(0)$ and $L_\phi(f) = \phi(0) - L_\phi(1-f) \leq \phi(0)$, so $\left|L_\phi(f)\right| \leq \phi(0)$. If $\phi(0) = 0$ then $L_\phi = 0$ on X_Γ and $\phi = 0$ on V, and the result is true.

Otherwise without loss of generality assume that $\phi(0) = 1$. Then L_ϕ is a bounded linear functional with $\left\|L_\phi\right\| = 1$, and by the Hahn-Banach theorem it extends to a linear functional of norm 1 on $C(\mathbb{T}^n)$, which is identified with a measure $\mu \in \mathcal{M}(\mathbb{T}^n)$. We have that for $f \in X_\Gamma$,

$$L_\phi(f) = \int_{\mathbb{T}^n} f(-\boldsymbol{\gamma})\, \mathrm{d}\mu(\boldsymbol{\gamma}). \tag{1.63}$$

Because $1 = L_\phi(1) = \mu(\mathbb{T}^n) \leq \|\mu\|_{TV} = 1$, we can conclude that $\mu \geq 0$. We have that

$$\phi(\mathbf{x}) = L_\phi(e^{2\pi i \boldsymbol{\gamma}\cdot\mathbf{x}}) = \int_{\mathbb{T}^n} e^{2\pi i \boldsymbol{\gamma}\cdot\mathbf{x}}\,\mathrm{d}\mu(\boldsymbol{\gamma}), \tag{1.64}$$

so $\widehat{\mu}$ is an extension of ϕ to $\mathbb{Z}^n$. By Bochner's theorem it is positive definite, so the result is proved. □

Then we complete this process by coming to the following result.

Theorem 13. (Rudin) Given a finite set $\Gamma \subset \mathbb{Z}^n$, $P_\Gamma = Q_\Gamma$ if and only if every $\phi \in \mathrm{PD}(\Gamma)$ can be extended to a function in $\mathrm{PD}(\mathbb{Z}^n)$.

Proof. The proof will be in two parts. First we must show that Q_Γ is closed in P_Γ, and then we go on to prove the result.

Say that $\dim X_\Gamma = |V| = d$. Let $r > d$, and $f = \sum_{i=1}^r |g_i|^2$, where $g_i \in X_\Gamma$. There is a nontrivial set of $(\lambda)_{i=1}^r \subset \mathbb{R}$ such that

$$\sum_{i=1}^r \lambda_i |g_i|^2 = 0. \tag{1.65}$$

Assume without loss of generality that $\lambda_i \leq \lambda_j$ whenever $i \leq j$. By solving (1.65) for $|g_r|^2$, we get that

$$f = \sum_{i=1}^{r-1} \left(1 - \frac{\lambda_i}{\lambda_r}\right) |g_i|^2 . \tag{1.66}$$

So f is a sum of $r-1$ squares. By repeating this process iteratively, we can conclude that each $f \in Q_\Gamma$ is a sum of at most d squares.

Now find a sequence of $(f_n)_{n=1}^\infty \subset Q_\Gamma$ such that $f_n \to f \in X_\Gamma$ uniformly. For each f_n there are d Γ-polynomials $(g_{jn})_{j=1}^{d}$ such that

$$f_n = \sum_{j=1}^{d} \left|g_{jn}\right|^2. \tag{1.67}$$

Each f_n is uniformly bounded, hence g_{jn} are also. We can conclude that there is a subsequence $(n_i)_{i=1}^{\infty}$ such that for each $1 \leq j \leq d$ and for all $\mathbf{x} \in \Gamma$, there exist $c_j(\mathbf{x})$ such that

$$\lim_{i\to\infty} \widehat{g}_{jn_i}(\mathbf{x}) = c_j(\mathbf{x}). \tag{1.68}$$

Then it is clear that for

$$g_j(\boldsymbol{\gamma}) = \sum_{\mathbf{x}\in\Gamma} c_j(\mathbf{x}) e^{-2\pi i \boldsymbol{\gamma}\cdot\mathbf{x}}, \tag{1.69}$$

$f = \sum_j \left|g_j\right|^2$. We conclude that $f \in Q_\Gamma$, thus Q_Γ is a closed subset of P_Γ.

Now we will proceed to the second part of the proof. Assume that $P_\Gamma = Q_\Gamma$. Let $\phi \in \mathrm{PD}(\Gamma)$. We know from proposition 4 that $L_\phi(f) \geq$ for all $f \in Q_\Gamma$. Since $Q_\Gamma = P_\Gamma$, we have also $L_\phi(f) \geq 0$ for $f \in P_\Gamma$, so by proposition 4, ϕ can be extended to a function in $\mathrm{PD}(\mathbb{Z}^n)$.

For the other direction, assume that $P_\Gamma \neq Q_\Gamma$. Then there exists some $f_0 \in P_\Gamma$ such that $f_0 \notin Q_\Gamma$. Because Q_Γ is a closed convex set, the Hahn-Banach theorem guarantees the existence of a hyperplane in X_Γ that separates Q_Γ from f_0. In other words there is a bounded linear functional L such that $L(Q_\Gamma) \geq 0$ and $L(f_0) < 0$. Because X_Γ is finite dimensional, there exists a function ϕ on V such that $L = L_\phi$. By propositions 3 and 4, we can conclude that $\phi \in \mathrm{PD}(\Gamma)$ but can't be extended to a function in $\mathrm{PD}(\mathbb{Z}^n)$. The proof is complete. $\square$

Rudin contributed once more in 1970 to extend the one-dimensional theorem to one for radially symmetric functions [53].

Theorem 14. (Rudin) If f is positive definite on $B(0,r) \subset \mathbb{R}^n$ and radially symmetric, that is $f(x) = \psi(|x|)$, then f can be extended to a positive definite radially symmetric function on $\mathbb{R}^n$.

1.9 My Contributions

The remainder of this dissertation will be focused on the following algorithm:

$$\mu^{\#} = \underset{\mu \in \mathcal{M}(\mathbb{T}^2)}{\operatorname{argmin}} \|\mu\|_{TV} \quad \text{such that } \widehat{\mu}(m,n) = \mathbf{y}_{mn}, \tag{1.70}$$

$\widehat{\mu}$ is the Fourier transform and $\mathbf{y}_{mn}$ is the a priori known values of the Fourier transform, for $-N \leq m, n \leq N$. In Chapters 2 and 3, I will introduce concepts in measure theory and Choquet theory respectively to address the motivating question: which measures can be recovered uniquely through (1.70)?

We have seen some results that give unique recovery for certain kinds of discrete measures, so in Chapter 2 we create tools to try and recover measures which are not discrete, focusing on measures defined on smooth manifolds. There are difficult limitations to this task, though we generate ideas that may be fruitful with future research. But the difficulties provide motivation not only to focus in on the relationship between delta measures and (1.70), but if possible to try and represent all solutions in terms of discrete measures.

Chapter 3 discusses a field of convex geometry called Choquet theory, which is concerned precisely with representation problems of this type. We will see that delta measures form the so-called extreme points of the total variation norm, and as such are well suited to be solutions to the algorithm in (1.70).

In Chapter 4 we use results from Choquet theory to state our main results. Our Theorem 30 says that if we restrict to positive solutions, any initial data $\mathbf{y}$ admits at least one solution $\mu^{\#}$ which is discrete, and thus our only hope for unique recovery can come from discrete measures. Theorem 31 generalizes this result to signed and complex measures. Then our Theorem 33 gives a novel sufficient condition on the initial data $\mathbf{y}$ and integers N such that (1.70) yields a unique result.

Chapter 2: Measure Theory and Surface Measures

2.1 Definitions and Notation

One thing we are interested in is whether it is possible to find application to a broader category of measures than discrete ones. For example, is there a circumstance in which we can use total variation methods to recover a surface measure? Or perhaps a singular measure like the Cantor measure? As it turns out, there are some fundamental difficulties inherent in the geometry of algorithm (1.70) which make these problems difficult. In order to explore these questions more fully will require a deeper understanding of measure theory.

First some preliminaries on geometric measure theory. The field was developed in the mid 20^{th} century primarily to work on problems related to energy-minimizing surfaces. Early pioneers include Wendell Fleming and Herbert Federer, the latter of whose textbook remains a fundamental source on the field today [29]. As the theory developed it found application to many more problems in both analysis and geometry and has become a staple of the mathematical landscape. Fundamental concepts include rectifiable sets, which are generalizations of sets witch admit tangent spaces, and integral currents, which generalize the ideas of manifolds [47].

Fourier analysis has played an ever increasing role in the research of modern geometric analysts, such as Kenig and Toro, Hofmann, Mitrea and Taylor, Hofmann, Martell and Uriarte-Tuero [47]. Two excellent sources on the development of Fourier analysis in geometric measure theory are by Kahane and Salem [35], and a more recent book from Mattila [47]. While the specifics of these works aren't of interest to us for the most part, the machinery involved is, particularly results which might shed insight on the behavior of the Fourier transform. We hope to use those results in conjunction with our work on positive definite measures.

As going forward I will begin to look at more complicated measures I will make some definitions in order to be precise with terminology. For the n-dimensional Lebesgue measure in $\mathbb{R}^n$ (or $\mathbb{T}^n$, when applicable), I will write $\mathcal{L}^n$. I refer sometimes to a surface measure, which I will denote σ_K, where K is the image of a rectifiable curve if $n = 2$, or some $(n-1)$-dimensional surface for general n. Frequent choices are $K = S^{n-1}$ for the surface measure on the unit sphere, or $K = \{(x_1, \cdots, x_n) \mid x_n = 0\}$. The surface measure may be defined in a few equivalent ways. Here I will say that it may be defined by an isometry from a rectifiable curve to $(\mathbb{R}, \mathcal{L}^1)$.

Definition 4. Let $\gamma\colon [0,1] \to \mathbb{R}^2$ be a unit speed nonintersecting curve with image K. $\sigma_K(A) = \mathcal{L}^1(\gamma^{-1}(A))$.

This construction is also called a push-forward of $\mathcal{L}^1$ under γ. It turns out that this definition is equivalent to the 1-dimensional Hausdorff measure restricted to K also, which I will define.

In addition, in the cases where $K = S^{n-1}$ or $K = \{(x_1, \cdots, x_n) \mid x_n = 0\}$, K is a topological group under rotations and horizontal translations respectively, and therefore has a unique Haar measure. Because the Lebesgue measure—and therefore the Hausdorff measure as well—is invariant under each of those transformations, it can be seen easily that the unique Haar measure is equivalent to the Hausdorff measure definition up to a constant.

Denote by $\mathcal{M}(X)$ the set of Radon measures. We say a measure μ on X is Radon if it is Borel-regular and locally finite. μ is Borel regular if for each $A \subset X$, there is a Borel set $B \subset A$ such that $\mu(B) = \mu(A)$. If in addition $|\mu(X)| < \infty$, then we say that μ is bounded and we denote the set of bounded Radon measures $\mathcal{M}_b(X)$. In general for this chapter I refer to positive real-valued measures, but for later purposes I use this notation for complex-valued measures as well as specified. In addition, in this chapter it becomes necessary at times to distinguish when a measure is bounded or unbounded. In all other chapters we will only be concerned with bounded measures, so I will use the notation $\mathcal{M}(X)$ to mean the set of bounded measures.

Definition 5. The Hausdorff measure $\mathcal{H}^s$ in $\mathbb{R}^n$ is defined by

$$\mathcal{H}^s(S) = \lim_{\delta \to 0} \mathcal{H}^s_\delta(S), \tag{2.1}$$

where for $\delta > 0$,

$$\mathcal{H}^s_\delta(S) = \inf \left\{ \sum_j \alpha(s) 2^{-s} d(E_j)^s \;\middle|\; S \subset \cup_j E_j,\, d(E_j) < \delta \right\}. \tag{2.2}$$

$d(A)$ is the diameter of A, and $\alpha(s)$ is a positive constant, which may be scaled so that when s is an integer it agrees with the volume of the s-dimensional unit ball. If they agree, then in $\mathbb{R}^2$, $\mathcal{L}^n = \mathcal{H}^n$.

Likewise, the Hausdorff dimension of a set $S \subset \mathbb{R}^n$ is

$$\dim S = \inf\{s \mid \mathcal{H}^s(S) = 0\} = \sup\{s \mid \mathcal{H}^s = \infty\}. \tag{2.3}$$

2.2 Energy Integrals

Understanding Hausdorff measures is important, but what we are really interested in are more general measures, defined on lower-dimensional sets. We would like a natural way to relate a measure supported on a set to the dimension of its support. As it turns out, one of the first central results in geometric measure theory does just that. The goal of the following lemma is to aid in computing lower bounds for Hausdorff dimensions. From the definition of $\mathcal{H}^s$, it seems that computing lower bounds might be much more difficult that upper bounds, because we must prove bounds for arbitrary coverings. Frostman's lemma gives us a way to pseudo-approximate from below, by showing the existence of measures that "behave well" on balls that approximate the set from below.

Theorem 15. (Frostman's Lemma) For a Borel set $S \subset \mathbb{R}^n$, $\mathcal{H}^s(S) > 0$ if and only if there is a measure $\mu \in \mathcal{M}(\mathbb{R}^n)$ with $\mathrm{supp}(\mu) \subset S$ such that

$$\mu(B(\mathbf{x}, r)) \leq r^s \quad \forall \mathbf{x} \in \mathbb{R}^n,\ r > 0. \tag{2.4}$$

I will not prove the theorem here, but I will point out that one direction is immediate. If we have such a μ satisfying the condition, then given any covering of S with balls (B_j),

$$\sum_j d(B_j)^s \geq \sum_j \mu(B_j) \geq \mu(S) > 0. \tag{2.5}$$

An immediate consequence of Frostman's lemma is that

$$\dim S = \sup\{s \mid \exists \mu \in \mathcal{M}(S)\, \text{such that (2.4) holds}\}, \tag{2.6}$$

which is a demonstration of the lemma's usefulness in computing Hausdorff dimensions.

We now have a way to relate geometry of sets and measures. A natural question is whether we can go in the reverse direction, and find the "dimension" of a measure in some sense? The answer is that it is much more difficult, but there are some things we can do. Begin by defining the s-energy of a measure, which will quantify the criterion given in (2.4).

$$I_s(\mu) = \iint |\mathbf{x} - \mathbf{y}|^{-s}\, \mathrm{d}\mu(\mathbf{x})\, \mathrm{d}\mu(\mathbf{y}). \tag{2.7}$$

We can relate this to equation (2.4) through a change of variables.

$$\int |\mathbf{x} - \mathbf{y}|^{-s}\, \mathrm{d}\mu(\mathbf{y}) = s \int_0^\infty \frac{\mu(B(\mathbf{x}, r))}{r^{s+1}}\, \mathrm{d}r \tag{2.8}$$

If μ satisfies (2.4), then we have that if $0 < t < s$, and if μ is compactly supported with diameter $\leq M$,

$$I_t(\mu) = t \iint_0^M \frac{\mu(B(\mathbf{x}, r))}{r^{t+1}} \,\mathrm{d}r \,\mathrm{d}\mu(\mathbf{x}) \leq \int_0^M r^{s-t-1} \,\mathrm{d}r < \infty. \tag{2.9}$$

In addition, if $I_s(\mu) < \infty$, then we can construct a measure that satisfies (2.4). Since $\int |\mathbf{x}-\mathbf{y}|^{-s} \,\mathrm{d}\mu(\mathbf{x}) < \infty$ for almost all $\mathbf{y}$, we can choose $M > 0$ and construct $A = \{\mathbf{x} \mid \int |\mathbf{x}-\mathbf{y}|^{-s} \,\mathrm{d}\mu(\mathbf{y}) < M\}$ such that $\mu(A) > 0$. Then for all $\mathbf{x}$, $r > 0$, $\mu{\restriction}_A(B(\mathbf{x}, r)) \leq 2^s M r^s$. So $\mu{\restriction}_A$ satisfies (2.4).

We can conclude with a new full characterization of the Hausdorff dimension in terms of the energy of supported measures.

$$\dim S = \sup\{s \mid \exists \mu \in \mathcal{M}(S) \text{ such that } I_s(\mu) < \infty\}. \tag{2.10}$$

For example, if $\mu{=}\mathcal{L}^1{\restriction}_{[0,1]}$, $I_s(\mu) = \iint_0^1 |\mathbf{x}-\mathbf{y}|^{-s} \,\mathrm{d}\mathbf{x} \,\mathrm{d}\mathbf{y} < \infty$ if and only if $s < 1$. We also get a similar result if instead of $[0, 1]$ we take $\mathcal{L}^1{\restriction}_K$ where $K \subset \mathbb{R}^n$ is the image of a rectifiable curve.

Analogously in higher dimensions if $A \subset \mathbb{R}^n$ has $\mathcal{L}^n(A) > 0$, then $I_s(\mathcal{L}^n{\restriction}_A) < \infty$ if and only if $s < n$.

2.3 Cantor Measures

As an aside I'd also like to examine a less trivial example, which has non-integer dimension. For this section I'm going to define and prove a few results on Cantor sets and measures. I will define the Cantor set in the standard way.

Let $I \subset \mathbb{R}$ be a closed interval of length one. We form the Cantor set iteratively by deleting the middle third from each continuous interval at each step of the process. Let $I_{1,1} = I$, then $I_{2,1}$ and $I_{2,2}$ are the two closed intervals that are produced by removing the middle third of I. Likewise $I_{2,1}$ is split into $I_{3,1}$ and $I_{3,2}$, and so on in this fashion. The Cantor set itself is then defined

$$C = \bigcap_{k=1}^{\infty} \bigcup_{i=1}^{2^k} I_{k,i}. \tag{2.11}$$

The Cantor measure is typically defined using methods from real analysis. One can define a Cantor function which is constant on each interval in the complement of C, and then define the Cantor measure as the distributional derivative of this function. The method of construction is not important, but what is important is to note the following property of the Cantor measure, which I will heretofore refer to as μ: for any $I_{k,i}$, $\mu(I_{k,i}) = 2^{-k}\mu(C)$. In fact this property alone, along with the criterion that $0 < \mu(C) < \infty$, is enough to characterize μ up to a constant.

Let $s = \log(2)/\log(3)$. I will take note of two identities.

$$2\left(\frac{1}{3}\right)^s = 1, \tag{2.12}$$

$$d(I_{k,i})^s = \left(\frac{1}{3}\right)^{sk} = 2^{-k} = \mu(I_{k,i}). \tag{2.13}$$

Then I will show that $0 < \mathcal{H}^s(C) \leq 1$. For any k, we have

$$\sum_{i=1}^{2^k} d(I_{k,i})^s = \sum_{i=1}^{2^k} \mu(I_{k,i}) = 2^k \mu(I_{k,i}) = 1. \tag{2.14}$$

Since $\cup_{i=1}^{2^k} I_{k,i}$ is a cover of C, $\mathcal{H}^s(C) \leq 1$. For the other inequality, we wish to apply Frostman's lemma to μ. Let J be an interval which intersects C but does not contain C. Let $I_{l,j}$ be (one of) the largest intervals completely contained in J. Then $J \cap C$ is contained in the union of at most four intervals $I_{l,j_1}, \cdots, I_{l,j_4}$. Hence

$$\mu(J) \leq 4\mu(I_{l,j}) = 4d(I_{l,j})^s \leq 4d(J)^s. \tag{2.15}$$

We conclude from Frostman's lemma that $\mathcal{H}^s(C) > 0$.

In fact, because of how μ was defined, an immediate consequence of this fact is that $\mathcal{H}^s\lceil_C$ is equivalent to μ up to a constant. Lastly, we can combine this result with equation (2.9) to conclude that $I_t(\mu) < \infty$ if and only if $t < s$.

2.4 Fourier Analysis Methods

Next I will move into some tools for Fourier analysis by returning to the energy integral (2.7). I can write

$$I_s(\mu) = \int k_s * \mu(\mathbf{x}) \, \mathrm{d}\mu(\mathbf{x}), \tag{2.16}$$

where k_s is the Riesz kernel $|\mathbf{x}|^{-s}$. I can use the fact that for $0 < s < n$, $\widehat{k}_s = \gamma(n,s)k_{n-s}$ and apply Parseval's theorem to get a new formula for the s-energy:

$$I_s(\mu) = \gamma(n,s) \int |\mathbf{x}|^{s-n} |\widehat{\mu}(\mathbf{x})|^2 \, \mathrm{d}\mathbf{x}. \tag{2.17}$$

Now the flavor for the Fourier analysis of geometric measures becomes clear: the dimension of the support of a measure must be related to the rate of decay of its Fourier transform. At the very least we have a quick bound. If

$$\widehat{\mu}(\mathbf{x}) \leq -s/2 \tag{2.18}$$

then $I_s(\mu) < \infty$ and therefore $\dim\{\operatorname{supp}\mu\} \geq s$.

However the reverse picture is not so clear. Not only do there exist measures that don't obey this rule, in fact there are sets which don't admit any measures which achieve (2.6) for s equal to their Hausdorff dimension. To express this gap we have two new definitions. Define the Fourier dimension of a set S by

$$\dim_F S = \sup\{s \leq n \mid \exists \mu \in \mathcal{M}(S) \text{ such that } \widehat{\mu}(\mathbf{x}) \leq |\mathbf{x}|^{-s/2}\}. \tag{2.19}$$

As we showed above we have the bound $\dim_F S \leq \dim S$. But equality is not achieved in general. Those sets that do achieve equality are called Salem sets.

Formally we can calculate the form of $\widehat{k}_s$ from the two facts that it is radial and it has the property that $k_s(r\mathbf{x}) = r^{-s}(\mathbf{x})$. That implies that $\widehat{k}_s$ must also be translation invariant and it has the property that

$$\widehat{k}_s(r\boldsymbol{\xi}) = \int_{\mathbb{R}^n} e^{-2\pi i r\mathbf{x}\cdot\boldsymbol{\xi}} k_s(\mathbf{x})\,\mathrm{d}\mathbf{x} = r^{-n}\int_{\mathbb{R}^n} e^{-2\pi i\mathbf{x}\cdot\boldsymbol{\xi}} k_s(\mathbf{x}/r)\,\mathrm{d}\mathbf{x} = r^{s-n}\widehat{k}_s(\boldsymbol{\xi}). \tag{2.20}$$

It follows that $\widehat{k}_s = \gamma(n,s)k_{n-s}$ for some constant γ depending only on s and n.

It is important to note that $k_s \notin L^p$ for any $1 \leq p \leq \infty$, but can be written as a sum of L^p functions, and therefore is a tempered distribution. The above may be formalized in a more careful manner using this fact. I would next like to give a derivation of the Fourier form of the engergy integral I_s in (2.17).

Theorem 16. If $\mu \in \mathcal{M}(\mathbb{R}^n)$ is a measure and $0 < s < n$, then

$$I_s(\mu) = \int_{\mathbb{R}^n} k_s * \mu(\mathbf{x}) \, \mathrm{d}\mu(\mathbf{x}) = \gamma(n, s) \int_{\mathbb{R}^n} |\widehat{\mu}(\boldsymbol{\xi})|^2 |\boldsymbol{\xi}|^{s-n} \, \mathrm{d}\boldsymbol{\xi}. \tag{2.21}$$

Proof. If we assume that Parseval's theorem holds, then the computation is straightforward.

$$\int_{\mathbb{R}^n} k_s * \mu(\mathbf{x}) \, \mathrm{d}\mu(\mathbf{x}) = \int_{\mathbb{R}^n} (\widehat{k_s * \mu}) \overline{\widehat{\mu}} = \int_{\mathbb{R}^n} \widehat{k}_s \widehat{\mu} \overline{\widehat{\mu}} = \gamma(n, s) \int |\mathbf{x}|^{s-n} |\widehat{\mu}|^2 \, \mathrm{d}\mathbf{x}. \tag{2.22}$$

However, that is not a trivial assumption in this case, because neither k_s nor μ may be integrable functions, nor can they be assumed to have compact support, so we will have to make the calculation more directly.

Let $\mathcal{S}(\mathbb{R}^n)$ be the space of Schwartz functions, and $\mathcal{S}'(\mathbb{R}^n)$ the tempered distributions. It is straightforward to show that for any $0 < s < n$, $k_s \in \mathcal{S}'(\mathbb{R}^n)$, so the convolution is at least well-behaved for $\mu = \phi \in \mathcal{S}(\mathbb{R}^n)$.

$$I_s(\phi) = \int_{\mathbb{R}^n} |\widehat{\phi}(\mathbf{x})|^2 |\mathbf{x}|^{s-n} \, \mathrm{d}\mathbf{x} \quad \forall \phi \in \mathcal{S}(\mathbb{R}^n). \tag{2.23}$$

What we wish to do then is approximate μ with test functions and show that the equality holds true in the limit.

Let η_ϵ be an approximate identity. That is $\eta \in \mathcal{S}(\mathbb{R}^n)$, with $0 \leq \eta \leq 1$, $\int \eta = 1$. Define $\eta_\epsilon(\mathbf{x}) = \epsilon^{-n}\eta(\mathbf{x}/\epsilon)$. Then $\mu_\epsilon = \mu * \eta_\epsilon \xrightarrow[\epsilon\to 0]{} \mu$ in $\mathcal{S}'(\mathbb{R}^n)$. By applying Parseval's formula to $\phi = \mu_\epsilon$, we get

$$I_s(\mu_\epsilon) = \int k_s * (\mu * \eta_\epsilon)(\mathbf{x})(\mu * \eta_\epsilon)(\mathbf{x})\,\mathrm{d}\mathbf{x} \tag{2.24}$$

$$= \gamma(n,s)\int |\mathbf{x}|^{s-n}|\widehat{\mu}(\mathbf{x})|^2|\widehat{\eta}(\epsilon\mathbf{x})|^2\,\mathrm{d}\mathbf{x}. \tag{2.25}$$

The term in (2.25) converges monotonely to $\gamma(n,s)\int |\mathbf{x}|^{s-n}|\widehat{\mu}(\mathbf{x})|^2$, so we only need to examine the term in (2.24). Write

$$\begin{aligned} I_s(\mu_\epsilon) &= \iint \left(\int k_s(\mathbf{x}-\mathbf{y})\eta_\epsilon(\mathbf{y}-\mathbf{z})\,\mathrm{d}\mu(\mathbf{z})\right)\mathrm{d}\mathbf{y}\left(\int \eta_\epsilon(\mathbf{x}-\mathbf{w})\,\mathrm{d}\mu(\mathbf{w})\right)\mathrm{d}\mathbf{x} \\ &= \iiiint |\mathbf{x}-\mathbf{y}|^{-s}\eta_\epsilon(\mathbf{y}-\mathbf{z})\eta_\epsilon(\mathbf{x}-\mathbf{w})\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y}\,\mathrm{d}\mu(\mathbf{z})\,\mathrm{d}\mu(\mathbf{w}) \end{aligned}$$

To show that we may exchange the order of integration, consider the integral

$$\begin{aligned} &\iint |\mathbf{x}-\mathbf{y}|^{-s}\eta_\epsilon(\mathbf{y}-\mathbf{z})\eta_\epsilon(\mathbf{x}-\mathbf{w})\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y} \\ &\quad = \iint |\epsilon(\mathbf{u}-\mathbf{v})+\mathbf{z}-\mathbf{w}|^{-s}\eta(\mathbf{u})\eta(\mathbf{v})\,\mathrm{d}\mathbf{u}\,\mathrm{d}\mathbf{v} \\ &\quad \leq |\mathbf{z}-\mathbf{w}|^{-s} + \epsilon\iint |\mathbf{u}-\mathbf{v}|^{-s}\eta(\mathbf{u})\eta(\mathbf{v})\,\mathrm{d}\mathbf{u}\,\mathrm{d}\mathbf{v}. \end{aligned}$$

For appropriately small ϵ, we can bound the final term by $C(s)|\mathbf{z}-\mathbf{w}|$, and if $I(\mu) < \infty$, we may conclude that it is absolutely integral with respect to $\mu(\mathbf{z})\times\mu(\mathbf{w})$.

Therefore we may swap the order of integration and pass through the limit to see

$$\begin{aligned} I(\mu_\epsilon) &= \iiiint |\epsilon(\mathbf{u}-\mathbf{v})+\mathbf{z}-\mathbf{w}|^{-s}\eta(\mathbf{u})\eta(\mathbf{v})\,\mathrm{d}\mathbf{u}\,\mathrm{d}\mathbf{v}\,\mathrm{d}\mu(\mathbf{z})\,\mathrm{d}\mu(\mathbf{w}) \\ &\xrightarrow[\epsilon\to 0]{} \iint |\mathbf{z}-\mathbf{w}|^{-s}\,\mathrm{d}\mu(\mathbf{z})\,\mathrm{d}\mu(\mathbf{w}) = I(\mu). \end{aligned}$$

Alternatively, if $I(\mu)=\infty$, then the result follows from Fatou's lemma, since $\infty = I(\mu) = I(\lim \mu_\epsilon) \leq \lim I(\mu_\epsilon)$. $\square$

2.5 Fourier Dimension and Salem Sets

Recall that the Fourier dimension $\dim_F S$ of a set is given by the fastest decay rate of $\widehat{\mu}$, for all μ supported in S, in a way that follows naturally from (2.17), and that a Salem set is a set whose Fourier dimension matches its Hausdorff dimension. Unfortunately we do not have to look far to find sets which are not Salem. For example, say S is supported in a lower-dimensional hyperplane H. Let $\boldsymbol{\xi} \in \mathbb{R}^n$ be orthogonal to that hyperplane, and let $\mu \in \mathcal{M}(S)$. Then we have for the Fourier transform of μ,

$$\widehat{\mu}(\boldsymbol{\xi}) = \int_S e^{-2\pi i \mathbf{x}\cdot\boldsymbol{\xi}}\,\mathrm{d}\mu(\mathbf{x}) = \int_H \mathrm{d}\mu. \tag{2.26}$$

$\widehat{\mu}$ is $O(1)$ as $|\boldsymbol{\xi}| \to \infty$, so the Fourier dimension of S is 0, although the Hausdorff dimension of S may be anywhere from 0 to $n-1$.

As we will derive in section 2.6, the surface measure σ on the unit sphere does achieve the optimal decay rate $\widehat{\sigma}(\mathbf{x}) \lesssim |\mathbf{x}|^{1-n}$, so S^{n-1} is a Salem set. As it turns out, the key factor here is the curvature of S^{n-1}, which plays a role in the analysis of oscillatory integrals. In a certain sense $\mathcal{H}^{n-1}\lceil_{S^{n-1}}$ has the optimal decay rate of any similarly defined set precisely because its curvature is uniform.

Another example is the Cantor set, which we will see is not Salem.

Theorem 17. Let C be the Cantor set. For any $\mu \in \mathcal{M}(C_d)$,

$$\limsup_{|x|\to\infty} |\widehat{\mu}(x)| > 0. \tag{2.27}$$

Proof. Assume for contradiction that such a measure $\mu \in \mathcal{M}(C)$ exists, such that $\lim \widehat{\mu}(x) = 0$. Without loss of generality we may actually assume less; that the Fourier series coefficients $\widehat{\mu}(k)$, $k \in \mathbb{Z}$ tend to 0. Let $\phi \in \mathcal{S}(\mathbb{R})$ be a Schwartz function such that $\text{supp}(\phi) \subset (1/3, 2/3)$ and $\int \phi = 1$. For $j \in \mathbb{N}$, define $\phi_j(x) = \phi([3^j x])$, where $[x]$ represents the fractional part of x.

First, I claim that for all $x \in C$, $j \in \mathbb{N}$, $[3^j x] \in C$. This is clear by recognizing that $x \in [0, 1]$ is in the Cantor set if and only if its base-3 representation contains no 1's. Multiplication by 3 in base-3 is represented by shifting each digit to the left, so for any $j \in \mathbb{N}$, the base-3 representation of $[3^j x]$ also contains no 1's, and is still in C.

Using this fact, we know that for all $j \in \mathbb{N}$, $\text{supp}\, \phi_j \cap C = \varnothing$. We can represent

$$\phi_j(x) = \sum_{k=-\infty}^{\infty} \widehat{\phi} e^{2\pi i x 3^j k}. \tag{2.28}$$

Therefore $\widehat{\phi}_j(3^j k) = \widehat{\phi}(k)$, and all other Fourier coefficients are 0. By Parseval's theorem, we have

$$
\begin{aligned}
0 = \int \phi_j \,\mathrm{d}\mu = \sum_{k=-\infty}^{\infty} \overline{\widehat{\phi}_j(k)}\widehat{\mu}(k) = \sum_{k=-\infty}^{\infty} \overline{\widehat{\phi}(k)}\widehat{\mu}(3^j k) \\
= \overline{\widehat{\mu}(0)}\widehat{\mu}(0) + \sum_{|k|>0} \overline{\widehat{\phi}(k)}\widehat{\mu}(3^j k)
\end{aligned}
$$

Because the sum

$$
\sum_{|k|>0} \overline{\widehat{\phi}(k)}\widehat{\mu}(3^j k) \tag{2.29}
$$

converges absolutely, and $\lim_j \sup_{|k|>0} \widehat{\mu}(3^j k) = 0$, the sum tends to 0 as $j \to \infty$. But recognize that $\overline{\widehat{\mu}(0)}\widehat{\mu}(0) = \mu(C)$, and hence we have shown that $\mu(C) = 0$, which is a contradiction. □

I have demonstrated the proof only for the standard 1/3-Cantor set, but the proof is easily generalizable to any number of other Cantor-type sets, including those formed by removing intervals of ratio other than 1/3.

While we are on the subject I will take the opportunity to do one more calculation, for the Fourier transform $\widehat{\mu}$. For this, return to the representation of C as the numbers which contain no 1's in their base-3 representation. With that in mind, we can construct C as a closed union $C = \overline{\cup_{k=1}^{\infty} E_k}$, where

$$
E_k = \left\{ \sum_{l=1}^{k} c_l 3^{-l} \;\middle|\; c_l = 0 \text{ or } 2,\ \forall l \right\}. \tag{2.30}
$$

Then define the measure

$$
\nu_k = 2^{-k} \sum_{e \in E_k} \delta_e. \tag{2.31}
$$

It is clear that ν_k converges weakly, so by the same symmetry arguments we used in defining μ, we can show that $\nu_k \to \mu$ weakly.

Then on the frequency spectrum, we can write

$$\widehat{\nu}_k(n) = 2^{-k} \sum_{e \in E_k} e^{-2\pi i e n} = 2^{-k} \sum_{e \in E_k} e^{-2\pi i n \sum_{l=1}^{k} e_l 3^{-l}}, \tag{2.32}$$

where e_l is the l^{th} decimal point of e in base-3. A basic calculation shows then that

$$\widehat{\nu}_k(n) = \prod_{j=1}^{k} \frac{1 + e^{-4\pi i 3^{-j} n}}{2} = \prod_{j=1}^{k} e^{-2\pi i 3^{-j} n} cos(2\pi 3^{-j} n). \tag{2.33}$$

Noting that $\sum_{j=1}^{k} 3^{-j} = (1 - 3^{-k})/2$,

$$\widehat{\nu}_k(n) = e^{-\pi i (1-3^{-k}) n} \prod_{j=1}^{\infty} \cos(2\pi 3^{-j} n). \tag{2.34}$$

Let $k \to \infty$, and we finally have

$$\widehat{\mu}(n) = e^{-\pi i n} \prod_{j=1}^{\infty} \cos(2\pi 3^{-j} n). \tag{2.35}$$

It is easy to verify that $\widehat{\mu}$ does not tend to 0 as $n \to \infty$. For $k \in \mathbb{N}$, $\widehat{\mu}(3^k) = -\prod_{j=1}^{\infty} \cos(2\pi 3^{k-j})$ is constant with respect to k and nonzero.

Given the above examples, it seems perhaps difficult to construct Salem sets. Indeed there are few explicit constructions, especially with Hausdorff dimension not an integer. However probabilistic methods are more fruitful. Particularly there are good results for sets that are produced from Brownian motion [34].

Theorem 18. If $\omega\colon [0,\infty) \to \mathbb{R}^n$ is a random Brownian motion, and $S \subset [0,\infty)$ is a Borel set, then the following is true almost surely.

1. If $n \geq 2$ or $\dim S \leq n/2$, then $\dim \omega(S) = 2\dim S$.
2. If $n = 1$ and $\dim S > 1/2$, then $\mathcal{L}^1(\omega(S)) > 0$.
3. $\omega(A)$ is a Salem set.

The only known explicit examples of Salem sets of non-integer dimension are of a form discovered by Kaufman, which takes advantage of famous results on Diophantine approximation [36]. Let $\alpha > 0$ and let E be the set of $x \in \mathbb{R}$ such that for infinitely many rationals p/q,

$$|x - p/q| \leq q^{-2-\alpha}. \tag{2.36}$$

E is a Salem set with dimension $2/(2+\alpha)$. Recent results have generalized this to $\mathbb{R}^n$, so constructive Salem sets are now known to exist in every dimension [32].

2.6 Fourier Decay of Surface Measures

Here I go through an important derivation of the Fourier transform of the surface measure $\sigma \in \mathcal{M}(\mathbb{R}^n)$, defined by $\sigma = \mathcal{H}^{n-1}\restriction_{S^{n-1}}$. First consider a generic radial function. It is a basic fact from calculus that if $f \in L^1(\mathbb{R}^n)$, then

$$\int_{\mathbb{R}^n} f(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int_{S^{n-1}} \int_0^\infty f(r\mathbf{x}) r^{n-1}\,\mathrm{d}r\,\mathrm{d}\sigma(\mathbf{x}). \tag{2.37}$$

We will need another related formula for rewriting the surface integral. Let $\mathbf{x}_0 \in S^{n-1}$ be fixed and define $S_\theta = \{\mathbf{x} \in S^{n-1} \mid \mathbf{x} \cdot \mathbf{x}_0 = \cos(\theta)\}$. S_θ is an $(n-2)$-dimensional sphere lying in S^{n-1}, with radius $\sin(\theta)$. Call the surface measure on S_θ by $\mathcal{H}^{n-2}\lceil_{S_\theta} = \sigma_\theta^{n-2}$, and note that $\sigma_\theta^{n-2}(S_\theta) = b(n)\sin^{n-2}\theta$, where $b(n)$ is the surface area of the unit $(n-2)$-ball. Then we can write

$$\int_{S^{n-1}} f \,\mathrm{d}\sigma = \int_0^\pi \int_{S_\theta} f(\mathbf{x}) \,\mathrm{d}\sigma_\theta^{n-2}(\mathbf{x}) \,\mathrm{d}\theta. \tag{2.38}$$

Then for a radial function $f(\mathbf{x}) = \psi(|\mathbf{x}|)$, we can compute for the Fourier transform using these two identities:

$$\widehat{f}(r\mathbf{x}_0) = \int_{\mathbb{R}^n} f(\mathbf{y}) e^{-2\pi i \mathbf{y}\cdot r\mathbf{x}_0} \,\mathrm{d}\mathbf{y} = \int_0^\infty \psi(s) s^{n-1} \int_{S^{n-1}} e^{-2\pi i \mathbf{y}\cdot r\mathbf{x}_0} \,\mathrm{d}\sigma(\mathbf{y}) \,\mathrm{d}s.$$

We can take advantage of the fact that for $\mathbf{y} \in S_\theta$, $e^{-2\pi i \mathbf{y}\cdot r\mathbf{x}_0} = e^{-2\pi i r\cos\theta}$ is constant with respect to $\mathbf{y}$.

$$\widehat{f}(r\mathbf{x}_0) = \int_0^\infty \psi(s) s^{n-1} \int_0^\pi e^{-2\pi i r\cos\theta} b(n) \sin^{n-2}\theta \,\mathrm{d}\theta \,\mathrm{d}s. \tag{2.39}$$

With the change of variable $t = -\cos\theta$, this becomes

$$\int_0^\infty \psi(s) s^{n-1} \int_{-1}^1 e^{2\pi i s r t} (1-t^2)^{(n-3)/2} \,\mathrm{d}t \,\mathrm{d}s, \tag{2.40}$$

which is in the form of a Bessel function. Using the definition of Bessel functions

$J_m \colon [0, \infty) \to \mathbb{R}$,

$$J_m(u) = \frac{(u/2)^m}{\Gamma(m+1/2)\Gamma(1/2)} \int_{-1}^{1} e^{iut}(1-t^2)^{m-1/2}\,\mathrm{d}t, \tag{2.41}$$

we can substitute and obtain the formula for $\widehat{f}$,

$$\begin{aligned}\widehat{f}(r\mathbf{x}) &= \int_0^\infty \psi(s)s^{n-1}c(n)(rs)^{-(n-2)/2}J_{(n-2)/2}(2\pi rs)\,\mathrm{d}s \\ &= c(n)r^{-(n-2)/2}\int_0^\infty \psi(s)J_{(n-2)/2}(2\pi rs)s^{n/2}\,\mathrm{d}s.\end{aligned}$$

We can now use this formula to derive some decay estimates. Using the following two classical properties of Bessel functions,

$$J_m(t) \leq C(m)t^{-1/2} \quad t > 0, \tag{2.42}$$

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\left(t^m J_m(t)\right) = t^m J_{m-1}(t), \tag{2.43}$$

we can quickly calculate the Fourier transform of the unit ball.

$$\begin{aligned}\widehat{\chi}_{B(0,1)}(\mathbf{x}) &= \frac{c(n)}{2\pi|\mathbf{x}|}|\mathbf{x}|^{(2-n)/2}J_{n/2}(2\pi|\mathbf{x}|s)s^{n/2}\Big|_{s=1} \\ &= C(n)|\mathbf{x}|^{-n/2}J_{n/2}(2\pi|\mathbf{x}|).\end{aligned}$$

Using (2.42), this satisfies the decay $\widehat{\chi}_{B(0,1)}(\mathbf{x}) \leq C(n)|\mathbf{x}|^{-(n+1)/2}$. Then finding a bound for the surface measure σ on S^{n-1} is as simple as recognizing that we can define it as a weak limit of the functions

$$\delta^{-1}\chi_{B(0,1+\delta)\setminus B(0,1)}, \quad \delta \to 0. \tag{2.44}$$

Then we get the formula

$$\begin{aligned}\widehat{\sigma}(\mathbf{x}) &= \lim_{\delta\to 0} c(n)|\mathbf{x}|^{-(n-2)/2}\delta^{-1}\int_1^{1+\delta} J_{(n-2)/2}(2\pi r s)s^{n/2}\,\mathrm{d}s \\ &= c(n)|\mathbf{x}|^{-(n-2)/2}J_{(n-2)/2}(2\pi|\mathbf{x}|).\end{aligned}$$

This has a decay rate of $\widehat{\sigma}(\mathbf{x}) \lesssim |\mathbf{x}|^{(1-n)/2}$.

2.7 Discussion

Our goal is to find options for recovering non-discrete measures. We know that certain surface measures cannot be recovered by means of total variation minimization [6]. I consider two main candidates for accomplishing this: we may either minimize for something other than total variation or limit the optimization to a subspace of $\mathcal{M}(\mathbb{T}^2)$.

If we wish to recover surface measures a natural question is: what is a possible subspace $S \subset \mathcal{M}(\mathbb{T}^2)$ which contains some or all 1-dimensional surface measures but no measures of dimension 0? We can rule out S being the set of all surface measures of rectifiable curves, because the weak-* closure of this set is $\mathcal{M}(\mathbb{T}^2)$—for example, a sequential limit of surface measures on concentric circles with decreasing radius will converge to a point mass. Therefore nothing is gained by limiting optimization to S; we must choose a subset which is closed.

Limiting the decay rate is also tempting. Take $S = \{\mu \in \mathcal{M}(\mathbb{T}^2 \mid \widehat{\mu}(\mathbf{x}) \leq C|\mathbf{x}|^{-1/2}\}$, for some aptly chosen $C > 0$. We know from our derivation in the prior section that S will contain at least some surface measures, and cannot include any measures with support smaller than 1 dimension. The problem is with the choice of C. Any solution to (1.70) will necessarily fall on the boundary of S, so the choice of C has an effect on the outcome of the algorithm. This may be a good option. If, for example, we know that the desired solution is a circle with prescribed radius, we may find the correct value of C by direct calculation by means of the formula in the previous section. However this method requires a large amount of prior knowledge which may not be realistic.

One more option to explore is using the energy integrals I_s as an alternative or addendum to the total variation norm. Consider an algorithm,

$$\sigma^{\#} = \underset{\mu \in \mathcal{M}(\mathbb{T}^2)}{\operatorname{argmin}} \|\mu\|_{TV} + \lambda I_s(\mu) \quad \text{such that } F(\mu) = \mathbf{y}, \tag{2.45}$$

for some appropriately chosen s and $\lambda > 0$. This algorithm would exclude any discrete measures as solutions, but beside that fact the results are somewhat unclear. One problem is that for μ a surface measure, $I_1(\mu) = \infty$, so we must choose $s < 1$, which will also allow potentially for solutions with dimension between 0 and 1. Otherwise this territory is fairly unexplored to the author's knowledge, and I would recommend it as a topic of future research.

The relationship between total variation, Hausdorff dimension and the Fourier transform is rich and complex. There are relationships which may be exploited, but dealing with them takes caution. The moral of this chapter is that it is difficult to divorce the total variation norm on a continuous space from the discrete measures. Particularly to note is the fact stated previously that a discrete measure may be expressed as a limit of surface measures or continuous measures (approximate identity), but the converse is not true. In the next chapter I will explore the question of why the discrete measures are unique in this regard, and I get a result that confirms our expectations that unique recovery of non-discrete measures is impossible with total variation minimization.

Chapter 3: Choquet Theory

3.1 Choquet's Theorem

Choquet theory is concerned with convex sets and their extreme points. Extreme points can provide a characterization of convex sets, in particular if they are compact. The content of this section is focused on results that provide representation theorems of convex sets, and points within convex sets, centered on their extreme points. For a motivating example I begin with a classical theorem of Carathéodory [19].

Theorem 19. Let a point $\mathbf{x}$ fall in the convex hull of a set $E \subset \mathbb{R}^n$. $\mathbf{x}$ can be written as a convex sum of at most $n+1$ points in E.

An easy corollary is that if E consists of $n+1$ affinely independent points (the points $\mathbf{x}_k - \mathbf{x}_1$ are linearly independent for $k > 1$), then the convex sum is unique. In this case the convex hull of E is a simplex, and E are its vertices. From this it is easy to see that the bound $n+1$ is the best that can be done. If E is affinely independent and $\mathbf{x}$ lies in the interior of E, then $\mathbf{x}$ is affinely independent from any strict subset of E, therefore it can only be written as a convex combination of all $n+1$ elements of E.

Definition 6. A point x in a convex set S in a locally convex topological vector space is an extreme point if for any $a, b \in S$, $a \neq b$, and for any $t \in (0, 1)$, $x \neq ta + (1-t)b$. Denote the set of extreme points of S by $\mathrm{ex}(S)$.

Definition 7. Let E be a subset of a locally convex topological vector space. The convex hull of E is the smallest convex set containing E. The closed convex hull of E is the smallest closed set containing E.

Extreme points are in a sense generators of compact convex sets. Because they cannot be written as convex combinations of other elements, they carry essential information about the set. The Krein-Milman theorem tells us that they are sufficient to characterize their underlying sets as well [39].

Theorem 20. (Krein-Milman) If X is a locally convex topological vector space and S is a compact convex subset of X, then S is the closed convex hull of its extreme points.

Results of this sort in finite dimensions date back to the early 20$^{\text{th}}$ century, first attributed to Minkowski and Carathéodory [50]. Carathéodory's theorem 19 suggests the idea of simplices as a basic archetype of a convex set. As we will see in section 3.3, Choquet theory develops tools to generalize the idea of the simplex to infinite dimensional space.

The insight of Choquet to convex geometry is generalizing the idea of a convex combination to infinite-dimensional domains, by considering them as integrals rather than sums. Now instead of being represented by the nonnegative weights of a finite sum, elements of a convex set can be represented as an integral over a nonnegative measure supported on a generating set.

Definition 8. Let X be a locally convex topological vector space and S be a nonempty compact subset. If μ is a probability measure on S, we say that a point $x \in X$ is represented by μ if

$$f(x) = \int_S f(t)\,\mathrm{d}\mu(t) \qquad \text{for all affine functions } f \text{ on } X. \tag{3.1}$$

We say that $f\colon X \to \mathbb{R}$ is affine if for all $x, y \in X$, $\lambda \in \mathbb{R}$, $f(\lambda x + (1-\lambda)y) = \lambda f(x) + (1-\lambda)f(y)$. Note that any point x is trivially represented by the delta measure $\delta_{\mathbf{x}}$.

For example, in the finite dimensional case we might take $\mathbf{x} \in \mathbb{R}^N$, and standard basis $E = (\mathbf{e}_n)_{n=1}^N$. Let $\mathbf{x} = \sum_n \xi_n \mathbf{e}_n$, where $\xi_n > 0$ and $\sum_n \xi_n = 1$. This means that $\mathbf{x}$ is a convex combination of basis vectors. Alternatively, we can let $\mu = \sum_n \xi_n \delta_{\mathbf{e}_n}$ be a measure on $\mathbb{R}^N$. For all $\mathbf{y}^\intercal \in (\mathbb{R}^N)'$,

$$\int_E \mathbf{y}^\intercal \mathbf{z}\,\mathrm{d}\mu(\mathbf{z}) = \sum_{n=1}^N \xi_n \mathbf{y}^\intercal \mathbf{e}_n = \langle \mathbf{y}, \mathbf{x} \rangle\,. \tag{3.2}$$

Hence $\mathbf{x}$ is represented by the probability measure μ in a way that is analogous to writing it as a convex combination. Note also that the restriction that μ be positive and norm one is equivalent in this case to $\xi_n > 0$ and $\sum_n \xi_n = 1$.

For an example that is less trivial, I want an infinite dimensional space. Let $X = \mathcal{M}_b(\mathbb{T}^2)$, the set of bounded complex measures on the Torus, which is relevant since I will return to it when I discuss superresolution. Let $S = \{\delta_{\mathbf{x}} \mid \mathbf{x} \in \mathbb{T}^2\}$, and σ be the 1-dimensional surface measure on the set $\{(x_1, x_2) \mid x_2 = 0\}$. It is easy to see that σ cannot be written as a finite convex combination of delta measures, but I wish to show that it can be represented by a probability measure.

Proposition 5. Let $\Sigma \in \mathcal{M}_b(S)$ be defined by $\Sigma(A) = \sigma(\{\mathbf{x} \in \mathbb{T}^2 \mid \delta_{\mathbf{x}} \in A\})$. Then Σ represents σ.

Proof. First observe that Σ is well defined, and that it is a positive measure, with $\|\Sigma\| = 1$. Let $f \in X'$ be a continuous linear functional on $\mathcal{M}_b(\mathbb{T}^2)$. We wish to show that $f(\sigma) = \int_S f \, \mathrm{d}\Sigma$. Note that because of how Σ is defined, (S, Σ) and $(\mathbb{T}^2, \sigma)$ are isomorphic as measure spaces under the isomorphism $\delta_{\mathbf{x}} \mapsto \mathbf{x}$. We can conclude that

$$\int_S f(\mu) \, \mathrm{d}\Sigma(\mu) = \int_{\mathbb{T}^2} f(\delta_{\mathbf{x}}) \, \mathrm{d}\sigma(\mathbf{x}). \tag{3.3}$$

Now consider the sequence of measures

$$\mu_N = \frac{1}{N} \sum_{n=1}^{N} \delta_{\left(\frac{n}{N}, 0\right)}. \tag{3.4}$$

μ_N converges to σ weakly. We can conclude that

$$\begin{aligned} \int_{\mathbb{T}^2} f(\delta_{\mathbf{x}}) \, \mathrm{d}\sigma(\mathbf{x}) &= \lim_{N \to \infty} \int_{\mathbb{T}^2} f(\delta_{\mathbf{x}}) \, \mathrm{d}\mu_N(\mathbf{x}) \\ &= \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} f\left(\delta_{\left(\frac{n}{N}, 0\right)}\right) \\ &= \lim_{N \to \infty} f(\mu_N) \\ &= f(\sigma). \end{aligned} \tag{3.5}$$

Therefore Σ represents σ. $\square$

One thing to note is that the final inequality requires that f is continuous with respect to the weak topology on $\mathcal{M}_b(\mathbb{T}^2)$. This is significant, as it is a requirement for Choquet's representation theorem for S to be compact, which it is in this case only under the weak topology.

The ultimate goal of this machinery is to say that we can actually represent all points in a convex set by measures on its extreme points, which was achieved by Choquet in 1956 [22].

Theorem 21. (Choquet-Bishop-De Leeuw) Let S be a compact convex subset of a locally convex space X. Let $x \in S$. Then there exists a probability measure $\mu \in \mathcal{M}_b(X)$ such that μ vanishes on every Baire subset of S that doesn't contain an extreme point of S, and that for any bounded linear functional f on X,

$$f(x) = \int_S f \, \mathrm{d}\mu \tag{3.6}$$

Choquet initially proved the theorem under the slightly stricter case where X is metrizable. Bishop and De Leeuw found the above strengthening three years later.

3.2 Application to Completely Monotone Functions

Choquet's theorem has found application to a variety of subjects, providing elegant proofs of some well-known representation theorems [3, 24, 28, 50]. I will go through an example here to provide context for the theory and demonstrate its wide application, and to illustrate the recurring themes of convex geometry that I wish to exploit.

A function $f\colon (0,\infty) \to \mathbb{R}$ is said to be completely monotonic if it is smooth and satisfies

$$(-1)^n f^{(n)} \geq 0 \quad \forall n \in \mathbb{N}. \tag{3.7}$$

For example, $e^{-\alpha x}$ and $x^{-\alpha}$ are completely monotonic. There is a theorem of Bernstein that characterizes completely monotone functions in terms of exponentials of the form $e^{-\alpha x}$ [7]. This result is particularly interesting in the context of the Laplace transform. In a certain sense it can be viewed as a version of Bochner's Theorem for the Laplace transform. For more details see Widder's comprehensive exposition of the Laplace transform [60]. Phelps gives a proof of his theorem for bounded completely monotonic functions, using methods from Choquet theory [50].

Theorem 22. (Bernstein) Let f be a completely monotone function. Then there exists a unique nonnegative measure $\mu \in \mathcal{M}[0,\infty]$ such that $\mu([0,\infty]) = f(0^+)$ and

$$f(x) = \int_0^\infty e^{-\alpha x}\,\mathrm{d}\mu(\alpha). \tag{3.8}$$

The rest of this section will be dedicated to a proof of Bernstein's theorem. The idea is clear from the representation formula. We wish to find an adequate topology on the set of completely monotone functions and a subset which is compact and convex, so that the extreme points are precisely the functions $e^{-\alpha x}$, $\alpha \in [0,\infty]$.

Let X be the convex cone of bounded completely monotonic functions. Note that because $f \in X$ is necessarily strictly nonnegative and nonincreasing, f is bounded if and only if $f(0^+) < \infty$. Define $S \subset X$ as the set of all such functions for which $f(0^+) \leq 1$. Choose the topology such that $f_k \to f$ if and only if for all $n \in \mathbb{N}$, $\|f_k^{(n)} - f\|_\infty \to 0$. By definition, this is a locally convex topology. We wish to show that in this case, S is compact.

We can construct the chosen topology of uniform convergence with a countable family of seminorms,

$$p_{m,n}(f) = \sup\{|f^{(k)}(x)| \mid 1/m \leq x \leq m,\, 0 \leq k \leq n\}. \tag{3.9}$$

By a standard construction, X is metrizable, with the metric

$$d(f,g) = \sum_{m,n=1}^{\infty} \frac{p_{m,n}(f-g)}{1 + p_{m,n}(f-g)}. \tag{3.10}$$

Under this definition, a set is bounded if and only if it is bounded in each seminorm. Therefore, given a bounded sequence of functions, the sequence is uniformly equicontinuous on each derivative and each compact set. An application of the Arzèla-Ascoli theorem, along with a diagonalization argument shows that the sequence has a convergent subsequence. Therefore every closed and bounded set is compact.

S is clearly closed, so it remains to show that S is bounded.

Lemma 23. Let $S_n = \{(-1)^n f^{(n)} \mid f \in K\}$. Let $a > 0$ and $n \geq 0$. Then S_n is uniformly bounded above on $[a, \infty)$ by $a^{-n}2^{(n+1)(n/2)}$.

Proof. The proof is by induction. For n=0, the lemma holds because for any $f \in K$, $f(0^+) \leq 1$ and f is strictly nonincreasing.

Assume for arbitrary $k \in \mathbb{N}$ that the lemma holds for S_k. Let $a > 0$. By the mean value theorem we have the existence of a point $c \in [a/2, a]$ such that

$$\frac{a}{2} f^{(k+1)}(c) = f^{(k)}(a) - f^{(k)}(a/2). \tag{3.11}$$

Then using monotonicity of functions in S_k and the induction hypothesis, we have

$$\begin{aligned}
(a/2)^{-k}2^{(k+1)(k/2)} &\geq (-1)^k f^{(k)}(a/2) && (3.12)\\
&\geq (-1)^{k+1}\frac{a}{2} f^{(k+1)}(c) && (3.13)\\
&\geq (-1)^{k+1}\frac{a}{2} f^{(k+1)}(a). && (3.14)
\end{aligned}$$

The inequality follows, and because of the monotonicity of $f \in S_k$, it will hold for $x \geq a$ as well. □

We have now shown that we have an adequate setting for Choquet's theorem to apply, and all that remains is to characterize the extreme points of S to obtain the desired representation theorem.

Lemma 24. The extreme points of S are precisely the functions of the form $f(x) = e^{-\alpha x}$, where $0 \leq \alpha \leq \infty$.

Proof. A straightforward computation shows that any such function falls in S. First

I will show that any extreme point must take this form.

Let f be an extreme point of S. Let $x_0 > 0$ and define $u(x) = f(x + x_0) - f(x)f(x_0)$. I claim that $f \pm u \in S$. Note first that

$$
\begin{aligned}
(f+u)(0^+) &= f(0^+) + f(x_0) - f(0^+)f(x_0) \\
&= f(0^+)(1 - f(x_0)) + f(x_0) \\
&\leq 1, \\
(f-u)(0^+) &= f(0^+) - f(x_0) + f(x_0)f(0^+) \\
&= f(x_0)f(0^+) + (1 - f(x_0)) \\
&\leq 1.
\end{aligned}
$$

And for $k \in \mathbb{N}$,

$$
\begin{aligned}
|u^{(k)}(x)| &= |f^{(k)}(x + x_0) - f(x_0)f^{(k)}(x)| \\
&\leq \max\{|f^{(k)}(x + x_0)|, |f(x_0)f^{(k)}(x)|\} \\
&\leq (-1)^k f^{(k)}(x).
\end{aligned}
$$

Therefore $(-1)^k(f \pm u)^{(k)}(x) \geq 0$, and $f \pm u \in S$.

Because we made the assumption that f is an extreme point of S, $u \equiv 0$. Hence for any $x, x_0 > 0$, $f(x + x_0) = f(x)f(x_0)$. Because f is continuous, it must be of the form $e^{-\alpha x}$. Because of the restriction that $f' \leq 0$, $\alpha \geq 0$. Note that I am allowing $\alpha = \infty$ for the case $f \equiv 1$.

Finally, what remains is to show that every function $e^{-\alpha x}$ is an extreme point. We can solve this by taking note of the mapping $f(x) \mapsto f(rx)$ for $r > 0$. This mapping is a bijection of S onto itself and preserves convex combinations, so it takes extreme points to extreme points. Since the Krein-Milman theorem implies that there is at least one nonconstant exponential function $e^{-\alpha x}$, which is an extreme point of S, we can generate any such exponential function $e^{-\alpha r x}$ by the above mapping to show that it is an extreme point. The lemma is proved. $\square$

It follows from Choquet's theorem that for any $f \in S$ there exists a probability measure $\mu \in \mathcal{M}(\mathrm{ex}(S))$ such that $f(x) = \int_{\mathrm{ex}(S)} \mathrm{d}\mu$. Now, because the correspondence $T\colon [0,1] \to \mathrm{ex}(S)$ given by $\alpha \mapsto e^{-\alpha x}$ is actually an homeomorphism, we have a natural correspondence from $\mathcal{M}[0,\infty]$ to $\mathcal{M}(\mathrm{ex}(S))$. We can identify $\mu \in \mathcal{M}(\mathrm{ex}(S))$ with $\widetilde{T}(\mu) \in \mathcal{M}[0,\infty]$ by $\widetilde{T}(\mu)(A) = \mu(\{e^{-\alpha x} \mid \alpha \in A\})$ and it turns out that

$$f(x) = \int_0^\infty e^{-\alpha x}\,\mathrm{d}\widetilde{T}(\mu)(\alpha), \tag{3.15}$$

as intended.

To show uniqueness, one needs only to use the Stone-Weierstrass theorem. Suppose that $\int_0^\infty e^{-\alpha x}\,\mathrm{d}\mu(\alpha) = \int_0^\infty e^{-\alpha x}\,\mathrm{d}\nu(\alpha)$ for all $x > 0$. If A is the subspace of $C[0,\infty]$ generated by functions of the form $\alpha \mapsto e^{-\alpha x}$, then $\mu = \nu$ as functionals on A. Since A separates points of $[0,\infty]$, it is dense, and therefore $\mu \equiv \nu$.

3.3 Uniqueness and the Choquet Simplex

While Choquet's theorem is a great tool for proving existence, it doesn't say anything about uniqueness. In fact we don't need to go further than the finite dimensional case to see examples of the failure of convex representations to be unique. For example, the unit square in the plane has extreme points $(\pm 1, \pm 1)$, and the origin is a convex combination of either of the two diagonal vertices. In fact any point in the interior has at least two convex representations. The unit ball by contrast is in a sense "maximally nonunique," in that any point in the unit ball is either an extreme point or has uncountably many representations as convex combinations of extreme points.

The prototypical example of a set where Choquet-type representations are unique is a simplex.

Proposition 6. Each point in a finite-dimensional simplex has a unique representation as a convex sum of its vertices.

Proof. Proceed by induction. For $k = 0$ is trivial, as the 0-dimensional simplex is a single point. Assume the hypothesis is true for k. Refer to the k-dimensional simplex as Δ^k, with vertices $\{\mathbf{x}_1, \cdots, \mathbf{x}_k\}$. Δ^{k+1} is formed by adding a new vertex $\mathbf{x}_{k+1}$, linearly independent from Δ^k, and taking the convex hull of the result. Let $\mathbf{x} \in \Delta^{k+1}$, without loss of generality $\mathbf{x} \neq \mathbf{x}_{k+1}$. Since $\mathbf{x}$ is in the convex hull of the vertices $\{\mathbf{x}_1, \cdots, \mathbf{x}_{k+1}\}$, we can write $\mathbf{x} = \sum_{m=1}^{k+1} c_m \mathbf{x}_m$. Consider the line $\ell = \{t\mathbf{x} + (1-t)\mathbf{x}_m \mid t \in \mathbb{R}\}$. ℓ passes through both the points $\mathbf{x}_{k+1}$ and $\mathbf{x}$, therefore it passes through the face of Δ^{k+1} opposite $\mathbf{x}_{k+1}$, which is identified with Δ^k. The point $\mathbf{x}_0 \in \ell \cap \Delta^k$ has for some $t \in \mathbb{R}$,

$$\mathbf{x}_0 = t\left(\sum_{m=1}^{k+1} c_m \mathbf{x}_m\right) + (1-t)\mathbf{x}_{k+1} = (1 + c_{k+1}) \sum_{m=1}^{k} c_m \mathbf{x}_m, \tag{3.16}$$

which is in Δ^k. By the induction hypthesis, $(c_1, \cdots, c_k)$ are unique, therefore $\mathbf{x}_{k+1}$ is also unique. $\square$

As an obvious consequence of this, the vertices of a simplex are also shown to be its extreme points. Carathéodory's theorem can be seen as an immediate consequence of this result as well. Any n-dimensional polyhedron may be decomposed into simplices. If the polyhedron has more than $n+1$ vertices, then the decomposition is not unique. For any point in the interior of the polyhedron, it has at least one different Choquet representation for each choice of simplicial decomposition.

The generalization of this concept is the Choquet simplex, a construction that generalizes the notion of simplices to infinite-dimensional vector spaces in a way that guarantees that points in Choquet simpleces have unique representations as measures on the extreme points. Here I will define the Choquet simplex and prove Choquet's uniqueness theorem for metric spaces.

The definition will require a few more constructions. Say that S is a compact convex set in a real locally convex space X. I will be working with the cone of S, so I will assume that S falls in an affine hyperplane in X that does not intersect the origin. Denote the cone $C = \{\lambda x \in X \mid x \in S, \lambda \geq 0\}$. If S falls in a hyperplane as described, then for all $x \in C$, there is a unique $\lambda \geq 0$ such that $\lambda x \in S$. We say that S is a base of the cone C.

The cone C defines a partial ordering on the space X. Let $x \preceq y$ if $y - x \in C$. Since S falls in a hyperplane that misses the origin, $C \cap -C = \{0\}$, so $x \preceq y$ and $y \preceq x$ implies $y = x$. On the other hand if $x \preceq y \preceq z$, then $y - x = \alpha u$, $z - y = \beta v$, where $\alpha, \beta \geq 0$ and $u, v \in S$. Then

$$\alpha u + \beta v = (\alpha + \beta) \left(\frac{\alpha}{\alpha + \beta} u + \frac{\beta}{\alpha + \beta} v \right), \tag{3.17}$$

which is in C, so $x \preceq z$. Note as well that for $x \in X$, the set $C_x = \{y \in X \mid x \preceq y\} = C + x$, so the partial order is also translation invariant. Finally, we say that for two elements $x, y \in X$ have an upper bound z if $x \preceq z$ and $y \preceq z$. The upper bound is a least upper bound if for every upper bound z_0, $z \preceq z_0$.

Finally, we define a convex set S that is the base of a cone C to be a Choquet simplex if and only if the space $C - C$ has the property that for each $x, y \in C - C$, there is a unique least upper bound, which we denote $x \vee y$. We say that $C - C$ is a vector lattice.

Definition 9. A partially ordered set is called a join- (meet-) semilattice if each pair of elements has a unique least upper (lower) bound. If a vector space is a semilattice which satisfies

1. $x \preceq y$ implies $x + z \preceq y + z$ for all z

2. $x \preceq y$ implies $\alpha x \preceq \alpha y$ for all $\alpha \geq 0$,

then it is a vector lattice.

Proposition 7. $C - C$ is a vector lattice if and only if C is a meet-semilattice. That is each pair of elements in C has a greatest lower bound.

Proof. For each $x, y \in C$, call their unique greatest lower bound $x \wedge y$. Let $x = x_1 - x_2$ and $y = y_1 - y_2$ be elements of $C - C$. Let $z = (x_1 + y_1) - (x_1 + y_2) \wedge (y_1 + x_2)$. Because $z - x = (x_2 + y_1) - (x_1 + y_2) \wedge (y_1 + x_2)$, we have $x \preceq z$, and similarly $y \preceq z$.

Now let $w = w_1 - w_2$ be an upper bound for x and y. It is clear from the definitions that $w_2 + x_1 + y_1 \preceq w_1 + x_2 + y_1$ and $w_2 + x_1 + y_1 \preceq w_1 + x_1 + y_2$. Therefore,

$$w - z = (w_1 + x_1 + y_2) \wedge (w_1 + y_1 + x_2) - (w_2 + x_1 + y + 1) \succeq 0. \tag{3.18}$$

For the reverse direction, given $C - C$ is a vector lattice, one can define a greatest lower bound by $x \wedge y = -(-x \vee -y)$. By restricting to $C - \{0\} = C$, it is demonstrated that C is a meet-semilattice. $\square$

Geometrically, this can be interpreted as saying that the intersection of two identical cones will be another congruent cone. For basic examples, consider the finite dimensional examples I discussed earlier. The intersection of two circular cones $C_0 = \{(x, y, z) \in \mathbb{R}^3 \mid y^2 + z^2 \leq x^2\}$ and $C_w = C_0 + w$ situated in the same direction will be a hyperbola-shaped crescent. Therefore 0 and w can't have a least upper bound because for any $v \in C_0 \cap C_w$, $C_v = C_0 + v \subset C_0 \cap C_w$ is a strict subset; it does not contain all upper bounds of 0 and w. Therefore the circle is not a Choquet simplex.

By contrast, the nonempty intersection of two triangular pyramids, oriented in the same direction, is another identical prism. Thus, the triangle, a simplex, indeed meets the criterion to be a Choquet simplex as well.

3.4 Technical Proofs

Now I will provide some more technical proofs in Choquet theory. To start with there are some tools we will need in the upcoming proofs. First, I define an equivalence relation on measures. Say that $\mu \sim \nu$ if for all $h \in A(X)$, $\int h \, \mathrm{d}\mu = \int h \, \mathrm{d}\nu$.

Definition 10. Let X be a convex subset of a topological vector space and $f \colon X \to \mathbb{R}$. f is called convex if for all $x, y \in X$, and for all $0 < \lambda < 1$,

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y). \tag{3.19}$$

If the inequality above is strict, we say f is strictly convex. If it is an equality, then f is said to be affine. $-f$ is concave.

Define the set of all affine functions on X by $A(X)$. Notice that $A(X)$ contains functions of the form $f(x) + r$, where $r \in \mathbb{R}$ and f is linear, so it is rich enough to separate points.

Definition 11. Let $f \colon X \to \mathbb{R}$ be bounded. Call the upper envelope of f the function given by

$$\overline{f}(x) = \inf\{h(x) \mid h \in A(X) \quad \text{and} \quad h \geq f\}. \tag{3.20}$$

These are few basic properties of $\overline{f}$, which will be useful:

Proposition 8.

1. $\overline{f}$ is concave, bounded and upper semicontinuous.

2. $f \leq \overline{f}$ for all f. $f = \overline{f}$ if and only if f is concave and upper semicontinuous.

3. If f, g are bounded functions, $\overline{f+g} \leq \overline{f} + \overline{g}$, and $|\overline{f} - \overline{g}| \leq \|f - g\|_{C_0}$. $\overline{f+g} = \overline{f} + g$ if $g \in A(X)$.

4. $\overline{rf} = r\overline{f}$, for $r > 0$.

The following proposition shows how the upper envelope function interacts with the extreme points of its underlying set.

Proposition 9. Let X be a convex subset of a topological vector space. If $f \in C(X)$, then for each $x \in X$,

$$\overline{f}(x) = \sup\left\{\int f \, \mathrm{d}\mu \mid \mu \sim \delta_x\right\}. \tag{3.21}$$

$\overline{f}(x) = f(x)$ if x is an extreme point of X.

Proof. Let $g = \sup\{\int f \, \mathrm{d}\mu \mid \mu \sim \delta_x\}$, and we wish to show that $g = \overline{f}$. For any μ, ν with μ representing x and ν representing y, and $0 < \alpha < 1$, $\pi = \alpha\mu + (1-\alpha)\nu$

represents $\alpha x + (1 - \alpha y)$. Then we have the inequality $\int f \, \mathrm{d}\pi \leq g(\alpha x + (1 - \alpha y))$. Taking the supremum over all μ and ν, we have that

$$g(\alpha x + (1 - \alpha y) \geq \alpha g(x) + (1 - \alpha) g(y), \tag{3.22}$$

Therefore g is concave.

To show that g is upper semicontinuous, let $\{x_\alpha\}$ be a net converging to a point $x \in X$, with $g(x_\alpha) \geq r$ for each α. Fix $\epsilon > 0$, and then for each α choose a measure $\mu_\alpha \sim \delta_{x_\alpha}$ such that $\mu_\alpha(f) > r - \epsilon$. By the weak-* compactness of the set of probability measures, there exists a convergent subnet $\{\mu_\beta\}$ that converges to a measure μ. For each $h \in A(X)$, $h(x_\beta) = \mu_\beta(h) \to \mu(h)$, so $\mu(h) = h(x)$, and μ represents x. Also because $\mu_\beta(f) > r - \epsilon$ for each β, we have $\mu(f) \geq r - \epsilon$, and therefore $g(x) \geq r - \epsilon$. Because ϵ was arbitrary, we can let it go to 0 and obtain $g(x) \geq r$, which shows that g is upper semicontinuous.

Now since g is concave and upper semicontinuous, it follows that $g \geq \overline{f}$. On the other hand, if $h \in A(X)$ and $h \geq f$, for any $x \in X$ and μ representing x, we have $h(x) = \int h \, \mathrm{d}\mu \geq \int f \, \mathrm{d}\mu$. From this we get $h(x) \geq g(x)$, and then taking the infimum over all such h, $\overline{f}(x) \geq g(x)$. □

3.4.1 Choquet's Theorem for Metrizable Spaces

Now we wish to move on to our first major proof. Recall Choquet's Theorem.

Theorem 25. (Choquet) Let S be a compact convex subset of a metric space X. Let $x \in S$. Then there exists a probability measure $\mu \in \mathcal{M}_b(X)$ such that μ is supported on $\mathrm{ex}(S)$, and that for any bounded linear functional f on X,

$$f(x) = \int_S f(t)\,\mathrm{d}\mu(t) \tag{3.23}$$

This was the first representation theorem proved by Choquet [22]. It was, of course, generalized not long after to the Choquet-Bishop-De Leeuw theorem, but the proof is much more technical and the added machinery for working in non-metrizable topological spaces is not necessary for our purposes. The proof is courtesy of Phelps [50].

Proof. (Choquet's theorem) Suppose X is a metrizable compact convex subset of a locally convex topological vector space E. Let $x_0 \in X$ be arbitrary. What we get from the metrizability of X is that $C(X)$, and therefore $A(X)$, is separable. Let $(h_k)_{k=1}^{\infty}$ be a sequence of functions in $A(X)$ such that $\|h_k\|_{C_0} = 1$ and $\{h_k \mid k \in \mathbb{N}\}$ is dense in the unit sphere of A. In other words, (h_k) separates points of X.

Then let $f = \sum_k 2^{-k} h_k^2$. Note two things: (h_k) is uniformly bounded, so the sum converges uniformly; and h_k^2 is strictly convex for all k, so f is also strictly convex. Then we can define a subspace $B = A(X) + rf$, $r \in \mathbb{R}$. For $h \in A(X)$ and $r \geq 0$, we have by Proposition 8 that $h + r\overline{f} = \overline{h + rf}$. If $r < 0$, then $h + rf$ is concave, and $h + r\overline{f} \leq h + rf = \overline{h + rf}$. In either case we have

$$h + r\overline{f} \leq \overline{h + rf}, \tag{3.24}$$

and the functional on B defined by $h + rf \mapsto h(x_0) + r\overline{f}(x_0)$ is dominated by the functional $g \mapsto \overline{g}(x_0)$, which is bounded on $C(X)$. Therefore we can apply the Hahn-Banach theorem to extend to a functional m on C(X) which satisfies $m(g) \leq \overline{g}(x_0)$ and $m(h + rf) = h(x_0) + r\overline{f}(x_0)$. Note that if g is negative then $m(g) \leq \overline{g}(x_0) \leq 0$, so m is positive, and therefore is identified with a measure on X. $m(1) = 1$, so m is a probability measure. Because for any $h \in A(X)$, $m(h) = h(x_0)$, m represents x_0.

Now for any $h \in A(X)$ such that $f \leq h$, and therefore $h \geq \overline{f}$. Hence $m(\overline{f}) \leq m(h) = h(x_0)$. By the definition of $\overline{f}$, $m(f) \geq m(\overline{f})$, and since $\overline{f} \geq f$, $m(f) = m(\overline{f})$. We can conclude that m is supported on the set $\{x \in X \mid f(x) = \overline{f}(x)\}$.

Let x, y, z be distinct points in X. If $x = y/2+z/2$, then by the strict convexity of f,

$$f(x) < \frac{1}{2}f(y) + \frac{1}{2}f(z) \leq \frac{1}{2}\overline{f}(y) + \frac{1}{2}\overline{f}(z) \leq \overline{f}(x). \tag{3.25}$$

Therefore x is not in the support of m. We can conclude that every m is supported on a subset of the extreme points of X. □

As it turns out, one additional fact that follows from this proof is that the constructed function f has the stronger property that $f(x) = \overline{f}(x)$ if and only if x is an extreme point of X, which gives us another way of characterizing extreme points.

3.4.2 The Choquet-Meyer Theorem

The Choquet-Meyer theorem says that Choquet simplices are precisely the convex sets that support uniqueness of Choquet representations [23].

Theorem 26. (Choquet-Meyer) Let X be a compact convex subset of a locally convex, metrizable topological vector space E. The following are equivalent.

1. X is a Choquet simplex.

2. For all $f \in C(\mathrm{ex}(X))$, if f is convex, then $\overline{f} \in A(X)$.

3. If $\mu \in \mathcal{M}(\mathrm{ex}(X))$ represents $x \in X$, and $f \in C(X)$ is convex, then $\overline{f}(x) = \mu(f)$.

4. For any convex $f, g \in C(X)$, $\overline{f+g} = \overline{f} + \overline{g}$.

5. For all $x \in X$, there is a unique measure $\mu \in \mathcal{M}(\mathrm{ex}(X))$ that represents x.

The process outlined in this section is again from Phelps, as was much of the material in this chapter [50]. We will need a few more materials before attacking the proof in its entirety.

Define a partial ordering on $\mathcal{M}(X)$ given by $\mu \prec \nu$ if for all convex functions h on X, $\int h \,\mathrm{d}\mu \leq \int h \,\mathrm{d}\nu$. This definition is meaningful. For example, we can show that the surface measure (normalized) on the unit sphere is greater than the Dirac delta at the origin. Let f be convex and continuous on the unit ball. Because for all $x \in S^{n-1}$, $f(0) < (f(x) + f(-x))/2$, we have

$$\int f(x) \,\mathrm{d}\delta(x) = f(0) = \int_{S^{n-1}} f(0) \,\mathrm{d}\sigma \tag{3.26}$$

$$< \int_{S^{n-1}} \frac{f(x) + f(-x)}{2} \,\mathrm{d}\sigma(x) = \int f(x) \,\mathrm{d}\sigma(x). \tag{3.27}$$

With that established, it follows from an application of Zorn's lemma that there are maximal measures in this ordering. We wish to show the following equivalence:

Proposition 10. A measure $\mu \in \mathcal{M}(X)$ is maximal in the ordering from $\prec$ if and only if it is supported on the extreme points of X.

Proof. Assume μ as above is maximal. Let $f \in C(X)$. Define a linear functional L on the subspace $\operatorname{span}\{f\}$ given by $rf \mapsto r\mu(\overline{f})$. This functional is dominated by the bounded sublinear (but not linear) functional $p\colon C(X) \to \mathbb{R}$ $p(g) = \mu(\overline{g})$. By the Hahn-Banach theorem we can extend L to a linear functional on $C(X)$ which is bounded by p. If $g \leq 0$, then $\overline{g} \leq 0$, so $L(g) \leq p(g) = \mu(\overline{g}) \leq 0$, so L is a positive functional, and can be represented by a measure $\nu \in \mathcal{M}(X)$. Now, for any convex function g, $-g = \overline{-g}$, so

$$\nu(-g) \leq p(-g) = \mu(-g), \tag{3.28}$$

and $\mu \prec \nu$. Since μ is maximal, we know that $\mu = \nu$, and

$$\mu(f) = \nu(f) = L(f) = \mu(\overline{f}). \tag{3.29}$$

Now assume that for any convex continuous function f, $\mu(f) = \mu(\overline{f})$. Let choose a maximal measure such that $\mu \prec \lambda$. We know λ exists again from Zorn's lemma. Then for any concave g, $\lambda(g) \leq \mu(g)$. We can write for convex f,

$$\lambda(\overline{f}) = \inf\{\lambda(g) \mid g \text{ is concave and } g \geq f\}$$
$$\leq \inf\{\mu(g) \mid g \text{ is concave and } g \geq f\} = \mu(\overline{f}).$$

Hence for any convex f, $\mu(f) = \lambda(f)$. Because $\{f - g \mid f, g \text{ are convex}\}$ is dense in $C(X)$, we can conclude that $\mu = \lambda$ and μ is maximal.

Finally, recall that in the proof of Choquet's theorem, we constructed a convex function $f \in C(X)$ that has the property that $\mathrm{ex}(X) = \{x \in X \mid \overline{f}(x) = f(x)\}$. Then we can conclude that a measure μ is supported on the extreme points of X if and only if $\mu(f) = \mu(\overline{f})$. The proof is complete. $\square$

Furthermore, the set of nonnegative measures on a vector space X forms a lattice under this ordering.

Proposition 11. Given two positive measures μ and ν in $\mathcal{M}(X)$, their greatest lower bound exists and is of the form

$$\mu \wedge \nu = \min\left(\tfrac{\mathrm{d}\mu}{\mathrm{d}\mu+\nu}, \tfrac{\mathrm{d}\nu}{\mathrm{d}\mu+\nu}\right)(\mu+\nu), \tag{3.30}$$

where $\frac{\mathrm{d}\cdot}{\mathrm{d}\cdot}$ denotes the Radon-Nikodym derivative.

Proof. Let $\sigma \prec \mu, \nu$. It's obvious that σ is absolutely continuous with respect to $\mu + \nu$, since it is absolutely continuous with respect to μ and ν individually. If $f \in C(X)$ is convex, then

$$\int f \frac{\mathrm{d}\sigma}{\mathrm{d}\mu+\nu}\,\mathrm{d}\mu+\nu = \int f\,\mathrm{d}\sigma \leq \int f\,\mathrm{d}\mu = \int f \frac{\mathrm{d}\mu}{\mathrm{d}\mu+\nu}\,\mathrm{d}\mu+\nu, \tag{3.31}$$

and likewise for ν. Hence, $\frac{\mathrm{d}\sigma}{\mathrm{d}\mu+\nu} \leq \min\left(\frac{\mathrm{d}\mu}{\mathrm{d}\mu+\nu}, \frac{\mathrm{d}\nu}{\mathrm{d}\mu+\nu}\right) = \frac{\mathrm{d}\mu\wedge\nu}{\mathrm{d}\mu+\nu}$, and $\sigma \prec \mu \wedge \nu$. $\square$

The last thing we will need is a decomposition lemma for vector lattices.

Lemma 27. Let V be a vector lattice. If we have two subsets of positive elements of V, $(x_i)_{i=1}^I$ and $(y_j)_{j=1}^J$, and if $\sum_{i=1}^I x_i = \sum_{j=1}^J$, then there exists $z_{ij} \geq 0$, $1 \leq i \leq I$,

$1 \leq j \leq J$, such that $x_i = \sum_{j=1}^{J} z_{ij}$ and $y_j = \sum_{i=1}^{I}$.

Proof. Consider the case where $I = J = 2$. We have $x_1 + x_2 = y_1 + y_2$ and let

$$z_{11} = x_1 \wedge y_1$$

$$z_{12} = x_1 - z_{11}$$

$$z_{21} = y_1 - z_{11}$$

$$z_{22} = x_2 - z_{21} = y_2 - z_{12}$$

The desired identities follow immediately, so what we must show is that the final equality holds, and that each z_{ij} are nonnegative. By the translation invariance of the vector lattice, we have

$$z_{12} \wedge z_{21} = (x_1 - z_{11}) \wedge (y_1 - z_{11}) = x_1 \wedge y_1 - z_{11} = 0, \tag{3.32}$$

so z_{11}, z_{12}, z_{21} are nonnegative. Furthermore,

$$z_{12} + x_2 = x_1 + x_2 + z_{11} = y_1 + y_2 + z_{11} = z_{21} + y_2. \tag{3.33}$$

It follows that $y_2 - z_{12} = x_2 - z_{21}$, so z_{22} is well defined. $z_{21} \leq z_{21} + y_2$, so $z_{21} = z_{21} \wedge (z_{12} + x_2) \leq z_{21} \wedge z_{12} + z_{21} \wedge x_2 = z_{21} \wedge x_2$. Hence $z_{21} \leq x_2$, and z_{22} is nonnegative.

Finally, we can generalize to higher I, J by induction. If we are given $x_1 + x_2 = y_1 + y_2 + y_3$, we can set $\widetilde{y}_2 = y_2 + y_3$ and then we have $x_1 + x_2 = y_1 + \widetilde{y}_2$, and we can generate $\widetilde{z}_{11}, \cdots, \widetilde{z}_{22}$ as before. Then we have $\widetilde{z}_{12} + \widetilde{z}_{22} = y_2 + y_3$. We can apply the above again to get $z_{12}, z_{13}, z_{22}, z_{23}$ such that $\sum_{j=2}^{3} z_{ij} = \widetilde{z}_{i2}, \sum_{i=1}^{2} z_{ij} = y_j$. Set $z_{11} = \widetilde{z}_{11}$ and $z_{21} = \widetilde{z}_{21}$ and the desired result follows. Proceed inductively in this manner for all I, J. □

Now we move on to the proof of the main theorem.

Proof. (Choquet-Meyer) ($1 \Rightarrow 2$) Let $f \in C(X)$ and X be a simplex. Let $x_1, x_2 \in X$, $\alpha_1, \alpha_2 > 0$ such that $\alpha_1 + \alpha_2 = 1$. Call $z = \alpha_1 x_1 \alpha_2 x_2$. We'd like to show that $\overline{f}$ is affine, so we need that $\overline{f}(z) = \alpha_1 \overline{f}(x_1) + \alpha_2 \overline{f}(x_2)$.

By Proposition 9, $\overline{f}(z) = \sup\{\mu(f) \mid \mu \sim \delta_z\}$. Suppose μ is a discrete measure and $\mu \sim \delta_z$. Then there exist a finite sequence of $\beta_j \geq 0$ and $y_j \in X$ such that $\sum \beta_j = 1$ and $\mu = \sum \beta_j \delta_{y_j}$. By using Lemma 27 on vector lattices, we can then get a sequence of z'_{ij} so that $\beta_j y_j = z'_{1j} + z'_{2j}$ and $\alpha_i x_i = \sum_j z'_{ij}$. If we write $z'_{ij} = \gamma_{ij} z_{ij}$, $\gamma_{ij} \geq 0$ and $z_{ij} \in X$, then we get that $x_i = \sum_j \alpha_i^{-1} \gamma_{ij} z_{ij}$, which is a convex combination of elements of X. It follows that it represents a discrete measure

$$\mu_i = \sum_{j=1}^{J} \alpha_i^{-1} \gamma_{ij} \delta_{z_{ij}} \qquad i = 1, 2, \tag{3.34}$$

with $\mu \sim \delta_{x_i}$. We can conclude that $\overline{f}(x_i) \geq \mu_i(f) = \sum_j \alpha_i^{-1} \gamma_{ij} f(z_{ij})$.

Finally, by the convexity of f, we have $f(y_j) \leq \beta_j^{-1} \gamma_{1j} f(z_{1j}) + \beta_j^{-1} \gamma_{2j} f(z_{2j})$, so we can conclude that

$$
\begin{aligned}
\mu(f) &= \sum_j \beta_j f(y_j) \\
&\leq \sum_{i,j} \gamma_{ij} f(z_{ij}) = \alpha_1 \mu_1(f) + \alpha_2 \mu_2(f) \\
&\leq \alpha_1 \overline{f}(x_1) + \alpha_2 \overline{f}(x_2).
\end{aligned}
$$

If we take the supremum over all possible μ, we get $\overline{f}(z) \leq \alpha_1 \overline{f}(x_1) + \alpha_2 \overline{f}(x_2)$. Note that we made the assumption that μ may be a discrete measure without justification. This is acceptable because we can approximate any $\mu \sim \delta_z$ with a discrete measure from below; this can be made formal using a partitioning argument relying on the compactness of X.

$(2 \Rightarrow 3)$ If $\mu \sim \delta_x$ is supported on $\mathrm{ex}(X)$, then $\mu(f) = \mu(\overline{f})$. We want to show that $\mu(\overline{f}) = \overline{f}(x)$. $\overline{f}$ is affine and upper semicontinuous. Because for $h \in A$, $\mu(h) = h(x)$, it is enough to show that $\overline{f}$ can be approximated from above by continuous affine functions. Let h_1 and h_2 be in A such that $h_i > \overline{f}$. We want to show that there is another $h \in A$, $h > f$ such that $h \leq h_1, h_2$. Define $J = \{(x, r) \in X \times (0, \infty) \mid r \leq \overline{f}(x)\}$ and $J_i = \{(x, r) \in X \times (0, \infty) \mid r = h(x)\}$. Notice that by the semicontinuity of $\overline{f}$, J is closed, and $J_1 \cup J_2$ is compact. Since the two sets are disjoint, by the Hahn-Banach separation theorem we can separate them with a hyperplane $L(x, r) = \alpha$. The function given by $h(x) = L(x, h(x))$ satisfies our requirements. Then the set $H = \{h \in A \mid h \geq \overline{f}\}$ is directed downwards. In addition, the closure of H is bounded below, so it has a unique minimal element f'. Because for all $z \in X$, there exists $h_n \in H$ such that $h_n(z) \to \overline{f}(z) \geq f'(z)$, we can conclude that $f' = \overline{f}$, and

$$\mu(\overline{f}) = \inf\{\mu(h) \mid h \in H\} = \inf\{h(x) \mid h \in H\} = \overline{f}(x). \tag{3.35}$$

($3 \Rightarrow 4$) Let $f, g \in C(X)$ be convex. Choose $\mu \sim \delta_x$ supported on $\mathrm{ex}(X)$.

$$(\overline{f+g})(x) = \mu(f+g) = \mu(f) + \mu(g) = \overline{f}(x) + \overline{g}(x). \tag{3.36}$$

($4 \Rightarrow 5$) Let $x \in X$. Consider the set $S \subset C(X)$ of convex functions, and define a linear functional on $S - S$ by $m(f-g) = \overline{f}(x) - \overline{g}(x)$. Because for any $h_1, h_2 \in C(X)$, $\overline{h}_1(x) + \overline{h}_2(x) = \overline{h_1 + h_2}(x)$, m is linear on $S - S$, and we have as a property of upper envelopes that $|m(f-g)| \leq \|f - g\|$, we can conclude that m is uniformly continuous on $S - S$, and therefore extends to a continuous function on $C(X)$ with norm at most 1. $m(1) = 1$, so m is a probability measure, and therefore is identified with a measure on X.

By Proposition 9, we have that for any $f \in C(X)$, $m(f) = \overline{f}(x) = \sup\{\mu(f) \mid \mu \sim \delta_x\}$, so $\mu \lesssim m$ for any $\mu \sim \delta_x$. Therefore m is the unique maximal measure which represents x. Equivalently, it is the unique measure supported on $\mathrm{ex}(X)$ which represents x.

(5 $\Rightarrow$ 1) Consider the cone P of nonnegative measures on X. As we have shown in Proposition 11, this cone forms a lattice. We would like to consider the subcone $Q \subset P$ of measures supported on $\text{ex}(X)$, and find that this is a lattice as well.

First, for any two $\lambda, \mu \in Q$, $(\lambda + \mu)(f) = (\lambda + \mu)(\overline{f})$, so $\lambda + \mu \in Q$. For $r \geq 0$, $r\mu \in Q$ also.

Let $x, y \in Q$, and let $x \wedge y$ be their greatest lower bound in P. Recall the formula for the greatest lower bound of two measures in (3.30). It is clear from this formula that $x \wedge y$ is supported on the union of the supports of x and y, which are contained in $\text{ex}(X)$.Therefore $x \wedge y$ is in Q. To show that $x \wedge y$ is still a greatest lower bound in the natural ordering on Q, take $w \in Q$ such that $x - w$ and $y - w$ are in Q. Because $x \wedge y$ is a greatest lower bound on P, we have that $0 \prec P \prec x \wedge y$. It follows that $x \wedge y - w \in P$ and $x \wedge y - w \prec z$. Because $x \wedge y \in Q$, $x \wedge y - w$ is absolutely continuous with respect to $x \wedge y$, and therefore it is also supported on $\text{ex}(X)$ and therefore $x \wedge y - w \in Q$, and w is less than $x \wedge y$ in the natural order on Q. $x \wedge y$ is a true greatest lower bound, and Q is a lattice. Moreover, the set $Q_1 = \{\mu \in Q \mid \mu(X) = 1\}$ is a base for Q, and Q_1 is a simplex.

The final step is to show that X is a simplex. Define the resultant map $r\colon Q_1 \to X$ by $r(\mu)$ is the unique point in X represented by μ. It is easy to see that the resultant is an affine function. Our assumption was that to each $x \in X$, there is a unique measure in Q_1 such that μ represents x. This implies that the resultant map is one-to-one. By Choquet's theorem it is also onto, and therefore it follows that X is a simplex isomorphic to Q_1. $\square$

Notice that throughout this section we used the assumption that X is a metric space. This assumption can be dropped, but not without a cost. The Choquet-Bishop-De Leeuw theorem shows that we still get representation, but we cannot say that measures are supported on the extreme points anymore. One problem is that in a metrizable space, the extreme points form a G_δ set, but in the absence of a metric they may not even be Borel. Bishop and De Leeuw gave examples of such convex sets [50].

A more subtle use of the fact that X is a metric space was in the proof of Proposition 10. Recall that in the proof of Choquet's theorem we constructed a convex continuous function f such that $\mathrm{ex}(X) = \{x \in X \mid \overline{f}(x) = f(x)\}$. This construction relied on the metrizability of X, and clearly is not possible without it, as it would imply that $\mathrm{ex}(X)$ is a G_δ set. Without this tool it isn't possible to prove Proposition 10 in its full generality. Rather we get the weaker fact that $\mu \in \mathcal{M}(X)$ is maximal if and only if $\mu(f) = \mu(\overline{f})$ for all $f \in C(X)$. Thus the generalized version of Choquet-Meyer is

Theorem 28. (Choquet-Meyer, non-metrizable) Let X be a compact convex subset of a locally convex, metrizable topological vector space E. X is a Choquet simplex if and only if for each $x \in X$ there is a unique maximal measure μ_x such that $\mu_x \sim \delta_x$.

This is not exactly comparable to the Choquet-Bishop-De Leeuw theorem for non-metrizable spaces. Recall that the Choquet-Bishop-De Leeuw theorem says that there exists a representing measure which vanishes on Baire sets which contain no extreme points. However, the Choquet-Meyer Theorem does not show uniqueness of measures which satisfy this condition. Mokobodzki found an example of a simplex X, for which $\mathrm{ex}(X)$ is Borel, but not Baire, and there exists a point $x \in X$ with two distinct representing measures μ and ν, with $\mu(X \setminus \mathrm{ex}(X)) = 0$ and $\nu(X \setminus \mathrm{ex}(X)) = 1$, but both vanish on every Baire subset of X which is disjoint from its extreme points [50].

We aren't too concerned with these anomalous spaces, but they are the subject of some continuing study today in functional analysis.

Chapter 4: My Results

4.1 Uniqueness of Positive Definite Extensions

Now I proceed to my contributions to the topic, including the proofs of my main results. I wish to apply Choquet theory to both the problems of positive definite extensions and continuous superresolution. Recall that the positive definite extension problem is a special case of the superresolution problem, when we restrict ourselves to only caring about probability measures. The first thing we must do is identify the space we are working with and its extreme points.

Proposition 12. Let P be the set of probability measures on the torus,

$$P = \{\mu \in \mathcal{M}(\mathbb{T}^2) \mid \forall f \geq 0\, \mu(f) \geq 0, \|\mu\|_{TV} = 1\}. \tag{4.1}$$

$\mu \in P$ is an extreme point if and only if $\mu = \delta_{\mathbf{x}}$ for some $\mathbf{x} \in \mathbb{T}^2$.

Proof. The extreme points of a closed convex set are those which are not a convex combination of two distinct elements. Note that for two probability measures ν and ξ, the support of $\lambda\nu + (1-\lambda)\xi$ is the union of the supports of ν and ξ. Let μ be a probability measure. If μ is supported on a single point, then it cannot be written as a convex combination of any distinct ν and ξ, because if they are distinct, their supports must contain points outside of the support of μ.

If, on the contrary, μ is supported on at least two points a and b, then we can divide $\mathbb{T}^2$ into two sets, A and B, such that $a \in A$ and $b \in B$, such that $\mu(A) > 0$ and $\mu(B) > 0$. Then we can write

$$\mu = \mu\restriction_A + \mu\restriction_B. \tag{4.2}$$

Because $\mu\restriction_A / \int \mathrm{d}\mu\restriction_A$ and $\mu\restriction_B / \int \mathrm{d}\mu\restriction_B$ are distinct probability measures, μ cannot be an extreme point. $\square$

The problem I'll focus on is to find whether the following algorithm produces a unique result:

$$\mu^{\#} = \operatorname*{argmin}_{\mu \in \mathcal{M}(\mathbb{T}^2)} \|\mu\|_{TV} \quad \text{such that } \widehat{\mu}(m,n) = \mathbf{y}_{mn} \qquad -N \leq m, n \leq N, \tag{4.3}$$

where $(\mathbf{y}_{mn}) \in \mathbb{C}^{(2N+1)\times(2N+1)}$ are the observed data.

Recall from Proposition 11 that the cone of nonnegative measures is a meet-semilattice. Because P is a base for that cone, we know that P is a Choquet simplex and has the property that each element has a unique representation as a probability measure supported on $\mathrm{ex}(P)$. But that clearly isn't sufficient for unique reconstruction from a finite number of frequency samples. We have many examples of non-unique reconstruction [6]. A simple example is the two probability measures

$$\mu_1 = \frac{1}{2}\left(\delta + \delta_{1/2}\right), \tag{4.4}$$

$$\mu_2 = \frac{1}{4}\left(\delta + \delta_{1/4} + \delta_{1/2} + \delta_{3/4}\right). \tag{4.5}$$

$\widehat{\mu}_1(n) = \widehat{\mu}_2(n)$ when $|n| < 2$, but not when $|n| \geq 2$.

Consider the map $F\colon P \to \mathbb{C}^{(2N+1)\times(2N+1)}$ given by

$$F(\mu)(m,n) = \widehat{\mu}(m,n) \quad -N \leq m,n \leq N. \tag{4.6}$$

$F(P)$ is a compact, convex set in $\mathbb{C}^{(2N+1)\times(2N+1)}$. F will preserve convex combinations in the sense that if μ is a probability measure that represents $x \in \mathcal{M}(\mathbb{T}^2)$, then the pushforward $F_*\mu$ represents $F(x)$ on $\mathbb{C}^{(2N+1)\times(2N+1)}$. Recall that μ is a linear bounded functional on the space $\mathcal{M}(\mathbb{T}^2)$ with the property that for all affine functions $p\colon \mathcal{M}(\mathbb{T}^2) \to \mathbb{C}$,

$$p(x) = \int_{\mathcal{M}(\mathbb{T}^2)} p(t)\,\mathrm{d}\mu(t). \tag{4.7}$$

Given an affine function q on $\mathbb{C}^{(2N+1)\times(2N+1)}$, we calculate

$$q(F(x)) = \int_{\mathcal{M}(\mathbb{T}^2)} q(F(t))\,\mathrm{d}\mu(t) = \int_{\mathbb{C}^{(2N+1)\times(2N+1)}} q(\mathbf{z})\,\mathrm{d}F_*\mu(\mathbf{z}), \tag{4.8}$$

and thus it is shown that $F(x)$ has a representation by $F_*\mu$, which is also a probability measure.

Notice that there is a characterization of the uniqueness problem in the following way: the measure $\mu^{\#}$ in (1.70) is unique if the point $(\mathbf{y}_{mn}) \in \mathbb{C}^{(2N+1)\times(2N+1)}$ has a unique representation $F_{*}\mu$, where μ is a measure supported on $\mathrm{ex}(\mathcal{M}(\mathbb{T}^2))$. We can conclude that, despite the fact we are attempting to reconstruct measures in an infinite dimensional space, the reconstruction is actually taking place in the finite dimensional space $\mathbb{C}^{(2N+1)\times(2N+1)}$. We wish to know whether we can formulate (1.70) in full without appealing to the infinite dimensional space $\mathcal{M}(\mathbb{T}^2)$. The following lemma helps with that goal.

Lemma 29. Let $P \subset \mathcal{M}(\mathbb{T}^2)$ be the set of probability measures as before, and F as above. Then,

$$\mathrm{ex}(F(P)) = F(\mathrm{ex}(P)). \tag{4.9}$$

Moreover, for each $\mathbf{z} \in \mathrm{ex}(F(P))$, there is a unique element $\mathbf{x} \in \mathbb{T}^2$ such that $\mathbf{z} = F(\delta_{\mathbf{x}})$. $F\restriction_{\mathrm{ex}(P)}$ is a homeomorphism.

Proof. First, P is a convex, compact (in the weak-* topology) set, so by the Krein-Milman theorem 20 it is the closed convex hull of $\mathrm{ex}(P)$. F is linear and continuous with respect to the weak-* topology, so $F(P)$ is not only compact and convex itself, but it is also the closed convex hull of $F(\mathrm{ex}(P))$. Therefore, by applying the Krein-Milman theorem to $F(P)$, we can conclude that $\mathrm{ex}(F(P)) \subset F(\mathrm{ex}(P))$.

We must show that each point of $F(\mathrm{ex}(P))$ cannot be a convex combination of two other points in $F(P)$. Notice first that each point $\mathbf{z} \in F(\mathrm{ex}(P))$ is of the form

$$\mathbf{z}_{mn} = e^{2\pi i \langle m,n\rangle \cdot \mathbf{x}}, \tag{4.10}$$

for some $x \in \mathbb{T}^2$. Hence each vector $\mathbf{z} \in F(\text{ex}(P))$ has $\|\mathbf{z}\|_2 = 2N+1$. Therefore $F(P) \subset \overline{B}(0, 2N+1)$. Likewise, for any two distinct points $\mathbf{z}, \mathbf{w} \in F(P)$, $t \in (0,1)$

$$\left\|t\mathbf{z} + (1-t)\mathbf{w}\right\|_2 < t\|\mathbf{z}\|_2 + (1-t)\|\mathbf{w}\|_2 \leq 2N+1. \tag{4.11}$$

Hence, $t\mathbf{z} + (1-t)\mathbf{w}$ is not in $F(\text{ex}(P))$, and every point in $F(\text{ex}(P))$ is an extreme point of $F(P)$.

Finally, if $\mathbf{z} \in \text{ex}(F(P))$, then it must be in $F(\text{ex}(P))$, so there is an element $\delta_{\mathbf{x}} \in \text{ex}(P)$ such that $\mathbf{z} = F(\delta_{\mathbf{x}})$. Note that for $\mathbf{x}, \mathbf{y} \in \mathbb{T}^2$, if for $m, n \in \{0,1\}$, $e^{2\pi i\langle m,n\rangle\cdot\mathbf{x}} = e^{2\pi i\langle m,n\rangle\cdot\mathbf{y}}$, then $\mathbf{x} = \mathbf{y}$. Therefore, given $N \geq 1$, $\mathbf{x}$ is unique.

Finally, we have shown that $F\restriction_{\text{ex}(P)}$ is a continuous bijection. Because F is an open map, so is its restriction. Therefore $F\restriction_{\text{ex}(P)}$ is a homeomorphism. $\square$

An immediate consequence of this lemma is that the pushforward map $F_*\colon \mathcal{M}(\text{ex}(P)) \to \mathcal{M}(\text{ex}(F(P)))$ is in fact an isomorphism of vector spaces. Therefore we conclude that the following two statements are equivalent:

- $\mathbf{z} \in \mathcal{M}(\mathbb{T}^2)$ has a unique representation by a measure $\mu \in \mathcal{M}(\text{ex}(P))$.
- $F(\mathbf{z})$ has a unique representation by a measure $\nu \in \mathcal{M}(\text{ex}(F(P)))$.

Of course, $F(P)$ is not a simplex, as it has infinitely many extreme points, so as we already knew from the examples in [6] and (4.4), unique reconstruction is not possible in general. But we are aware that in certain cases we can get unique construction. For example, the measure $1/2(\delta + \delta_{(1/2,0)})$ is the unique solution to

(1.70) when $N \geq 2$. This is because the line

$$tF(\delta) + (1-t)F(\delta_{(1/2,0)}) \quad t \in [0,1] \tag{4.12}$$

is a simplex which lies in the boundary of $F(P)$. We call this a face of $F(P)$, analogous to a face of a polyhedron. The idea of a face of a convex set came from Alfsen, who defined the face of a Choquet simplex [1]. This idea will be important for uniqueness.

Now I will prove my first main result on positive definite extensions.

Theorem 30. If a finite sequence $\mathbf{y} \in \mathbb{C}^{(2N+1)\times(2N+1)}$ has an extension to an infinite positive definite sequence $(\overline{\mathbf{y}}_{mn})_{m,n=-\infty}^{\infty}$, then at least one such positive definite extension must be a finite sum of the form

$$\overline{\mathbf{y}}_{mn} = \sum_{k=1}^{K} \lambda_k e^{-2\pi i \langle m,n\rangle \cdot \mathbf{x}_k} \quad \lambda_k > 0, \mathbf{x}_k \in \mathbb{T}^2, \tag{4.13}$$

where $K \leq 4N^2 + 4N + 2$.

Proof. Let $\mathbf{y} \in \mathbb{C}^{(2N+1)\times(2N+1)}$ be arbitrary, such that the given assumptions hold. Without loss of generality let $\mathbf{y}_{0,0} = 1$. By Bochner's theorem, there exists a $\mu \in P$ such that $\mathbf{y} = F(\mu)$. Because $\mathbb{C}^{(2N+1)\times(2N+1)}$ is finite dimensional, Carathéodory's theorem 19 implies that $\mathbf{y}$ has a representation as a convex combination of at most $(2N+1)^2 + 1 = 4N^2 + 4N + 2$ extreme points of $F(P)$. Because each extreme point is of the form $F(\delta_{\mathbf{x}})$ for some unique $\mathbf{x} \in \mathbb{T}^2$, we have

$$\mathbf{y} = \sum_{k=1}^{K} \lambda_k F(\delta_{\mathbf{x}_k}) = F\left(\sum_{k=1}^{K} \lambda_k \delta_{\mathbf{x}_k}\right) \quad \lambda_k > 0, \sum_{k=1}^{K} \lambda_k = 1. \tag{4.14}$$

The measure $\nu = \sum \lambda_k \delta_{\mathbf{x}_k}$ is a probability measure on $\mathbb{T}^2$ with $F(\nu) = \mathbf{y}$, so $\widehat{\nu}$ is a positive definite extension of $\mathbf{y}$. The proof is complete. □

4.2 Faces of Choquet Simplices

The first obstacle to generalizing Theorem 30 to signed measures is that there is no longer an obvious setting to apply Choquet theory. The set of complex measures $\{\mu \in \mathcal{M}(\mathbb{T}^2) \mid \|\mu\|_{TV} = 1\}$ is no longer even convex, let alone a Choquet simplex. We propose to solve this problem by introducing the idea of faces. Consider why compressive sensing works in the finite dimensional case: in high dimensions, the shape of the ℓ_1 ball is such that not only is it a polyhedron, but most of its tangent affine planes of sufficiently low dimension will hit one of the low-dimensional faces. The faces of the ℓ_1 ball are in fact simplices. I will refer to the ℓ_1 ball as a cross-polytope in n dimensions. The solid (hollow) cross-polytope is the set $\{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\|_1 \leq (=)\, 1\}$.

Notably, because its faces are simplices, a hollow cross-polytope has the property that each point is a unique convex combination of vertices. We would like to find an infinite-dimensional analogy for $\mathcal{M}(\mathbb{T}^2)$. We have by Proposition 7 that the probability measures form a Choquet simplex. It is easy to see that each set of the form $\{\sigma\mu \mid \mu \in \mathcal{M}(\mathbb{T}^2),\ \sigma \in L^\infty(\mathbb{T}^2),\ \mu \geq 0,\ |\sigma| \equiv 1\}$ is also an isomorphic simplex via the affine transformation $\mu \mapsto \sigma\mu$. The central question of this section is whether it is possible to define a generalized polytope, and in what generality can it be defined?

A similar topic was studied by Alfsen [1]. As it turns out we may define a face of a simplex in the following way.

Definition 12. Let X be a compact convex set in a topological vector space V. An affine space H of codimension 1 is said to be a supporting hyperplane for X if

$$H \cap X \neq \emptyset \quad \text{and} \quad X \setminus H \text{ is convex.} \tag{4.15}$$

If H is a supporting manifold, then $H \cap X$ is a face of X.

An immediate implication is that a face can consist of a single point $\{x\}$ if and only if x is an extreme point. Thus the extreme points are analogous to the vertices of a simplex. We can in fact strengthen that statement to the following.

Proposition 13. If x in a face $F = H \cap X$ has a representation $\mu \sim \delta_x$, then μ must be supported in H.

Proof. Assume x, F, H, μ as above, and that μ is not supported in H. It is a standard fact, and a corollary of the Hahn-Banach theorem, that there is a linear functional $f \colon V \to \mathbb{R}$ such that $f = 0$ on H and $f > 0$ on $X \setminus H$. Because μ is not supported in H, there must be some compact $S \subset X \setminus H$ such that $\mu(S) > 0$ and $f(S)$ is bounded below by $\epsilon > 0$. We can conclude that

$$\int_X f(t)\,\mathrm{d}\mu(t) \geq \epsilon\mu(S) > 0 = \int_X f(t)\,\mathrm{d}\delta_{\mathbf{x}}(t), \tag{4.16}$$

which is a contradiction. □

An immediate consequence is that extreme points of faces correspond well with extreme points of their underlying sets.

Corollary 1. Let C be a face of a compact convex subset X. Then C is also compact and convex, and $\mathrm{ex}(C) \subset \mathrm{ex}(X)$.

Proof. C is the intersection of a closed convex hyperplane H and a compact convex set X, so it is immediately compact and convex.

Let $x \in \mathrm{ex}(C)$. By the Krein-Milman theorem we have that $x \in C$, so Choquet's theorem implies that there exists a measure μ supported on $\mathrm{ex}(X)$ such that μ represents x. But by Proposition 13, μ is supported in $H \cap X = C$, and since x is an extreme point of C we can conclude that $\mu = \delta_x$. But because $\{x\} = \mathrm{supp}(\mu) \subset \mathrm{ex}(X)$, x is an extreme point of X. □

We now know that the face structure of a closed convex set has some useful properties with respect to its extreme points. One more important implication is a property of Choquet simplices which will be helpful in generalizing Theorem 30. Proposition 14 explores this, but first we need to return to some materials on cones and lattices.

Definition 13. A subcone $\widetilde{X}$ of a cone X is said to be hereditary if $0 \leq x \leq y$ and $y \in \widetilde{X}$ implies $x \in \widetilde{X}$.

Proposition 14. Each face of a Choquet simplex is itself a simplex. The extreme points of a face are each extreme points of X as well.

Proof. The proof will take place in two parts. Recall that associated to any Choquet simplex X is a cone $\widetilde{X}$. A compact convex set X is a Choquet simplex if and only if the associated cone $\widetilde{X}$ is a lattice under the associated order: $x \leq y$ if $y - x \in \widetilde{X}$. In part 1 of the proof, we wish to show that each face F of X forms a hereditary subcone of $\widetilde{X}$—that is given $y \in \widetilde{F}$ and $x \in \widetilde{X}$, if $x \leq y$ then $x \in \widetilde{F}$. In part 2 we will show that every hereditary subcone of a lattice is a lattice itself.

(Part 1) If F is a face of X, then there is a hyperplane $H \subset V$ such that $F = H \cap X$, and X falls on one side of H. Equivalently we can say that there exists a linear functional f such that $H = \{x \in V \mid f(x) = 1\}$, and $f(x) \leq 1$ for all $x \in X$. We can extend f to the cone $\widetilde{X}$ sitting in the space $V \times \mathbb{R}$ in a natural way by $f(\alpha x) = \alpha f(x)$. Then we can characterize $\widetilde{F} = \{\alpha x \in \widetilde{X} \mid f(\alpha x) = \alpha\}$.

Let $y \in F$ and $x \in X$ such that for some positive numbers α, β we have $\alpha y \geq \beta x$. We know immediately that $f(\beta x) \leq \beta$, so $f(\alpha y - \beta x) \geq f(\alpha y) - \beta = \alpha - \beta$. On the other hand, because $\alpha y - \beta x \in \widetilde{X}$, we have that

$$f(\alpha y - \beta x) = f\left((\alpha - \beta)\left(\frac{\alpha}{\alpha - \beta}y - \frac{\beta}{\alpha - \beta}x\right)\right) \leq (\alpha - \beta). \tag{4.17}$$

Therefore $f(\alpha y - \beta x) = \alpha - \beta$ and $\beta x \in \widetilde{F}$

(Part 2) Now I wish to show that every hereditary subcone of a lattice is a lattice itself. To that end, let $\leq_X$ be the ordering induced by X and $\leq_F$ be the one induced by F. Let $x, y \in \widetilde{F}$ and we wish to find a greatest lower bound for x and y in $\widetilde{F}$. Let z be the greatest lower bound in $\widetilde{X}$, and let $w \in \widetilde{F}$ such that $w \leq_F x$ and $w \leq_F y$. We know already that $w \in \widetilde{F}$. Since $F \subset X$, we have that $0 \leq_X w \leq_X z$. Then because $z - w \leq_X z \in \widetilde{F}$, so $z - w \in \widetilde{F}$ and $w \leq_F z$. Therefore z is a greatest lower bound and F is a lattice.

We can conclude that each face of a Choquet simplex X generates a hereditary subcone of $\widetilde{X}$, and is therefore a Choquet simplex itself. $\square$

To end this section, I will point out that the definition of a face of a simplex is easily generalizable. We will find the notion useful in exploring the shape of cross-polytopes (which are not Choquet simplices but share some useful properties) in $\mathcal{M}(\mathbb{T}^2)$.

Definition 14. Let C be a compact convex set in a topological vector space V. An affine space H of codimension 1 is said to be a supporting hyperplane for C if

$$H \cap C \neq \emptyset \quad \text{and} \quad C \setminus H \text{ is convex.} \tag{4.18}$$

If H is a supporting manifold, then $H \cap C$ is a face of C.

4.3 Uniqueness for Signed and Complex Measures

Let $B = \{\mu \in \mathcal{M}(\mathbb{T}^2) \mid \|\mu\|_{TV} \leq 1\}$. I observe that by the definition from Alfsen, P is actually a face of B. Take the hyperplane in $\mathcal{M}(\mathbb{T}^2)$ defined by

$$H = \left\{ \mu \in \mathcal{M}(\mathbb{T}^2) \mid \operatorname{Re} \int_{\mathbb{T}^2} \mathrm{d}\mu = 1 \right\}. \tag{4.19}$$

Then $P = H \cap B$ and for all $\mu \in B$, $\operatorname{Re} \int \mathrm{d}\mu \leq \|\mu\|_{TV} \leq 1$. Therefore H is a supporting hyperplane which carves out the face P. It is easy to imagine that one is able to prove similar results for other faces of B, which begs the question: what is a characterization of the faces of the cross-polytope B?

A hyperplane H in a vector space V can always be generated by a linear functional $f\colon V \to \mathbb{R}$, such that $H = f^{-1}(1)$. We can say that H is a supporting hyperplane for a compact convex set C if and only if f can be chosen such that for all $x \in C$, $f(x) \leq 1$. Otherwise the hyperplane would split C into two disjoint convex sets $f^{-1}((1, \infty)) \cap C$ and $f^{-1}((-\infty, 1)) \cap C$.

Let $\sigma\colon \mathbb{T}^2 \to \mathbb{C}$ be a continuous function with $\|\sigma\|_\infty = 1$. σ is identified with a complex valued functional on $\mathcal{M}(\mathbb{T}^2)$ given by $\mu \mapsto \sigma(\mu) = \int \sigma(t)\, \mathrm{d}\mu(t)$. In fact, recall that we've given $\mathcal{M}(\mathbb{T}^2)$ the weak-* topology, therefore by definition every element of the topological dual $\mathcal{M}(\mathbb{T}^2)^*$ is of this form for some σ [5]. Define the affine hyperplane generated by σ as

$$H_\sigma = \sigma^{-1}(1). \tag{4.20}$$

I make the following observations about H_σ.

Proposition 15. H_σ is a supporting hyperplane for B if and only if $\|\sigma\|_\infty = 1$.

Proof. If $\|\sigma\|_\infty = 1$ then there exists some $\mathbf{x} \in \mathbb{T}^2$ such that for all $\mathbf{y} \in \mathbb{T}^2$, $|\sigma(\mathbf{y})| \leq |\sigma(\mathbf{x})| = 1$. Then for all $\mu \in B$, $|\int \sigma(t)\,\mathrm{d}\mu| \leq 1$ and $\int \sigma(t)\,\mathrm{d}(\overline{\sigma}(\mathbf{x})\delta_{\mathbf{x}}(t)) = 1$. Thus $H_\sigma \cap B$ is nonempty and $B \setminus H_\sigma$ is convex.

Given H_σ is a supporting hyperplane of B, then for all $\mathbf{x} \in \mathbb{T}^2$, $|\sigma(\mathbf{x})| = |\sigma(\delta_{\mathbf{x}})| \leq 1$, so $\|\sigma\|_\infty \leq 1$. On the other hand, there must exist some $\mu \in B$ such that

$$1 = \int_{\mathbb{T}^2} \sigma(t)\,\mathrm{d}\mu(t) \leq \|\sigma\|_\infty |\mu|(\mathbb{T}^2) = \|\sigma\|_\infty . \tag{4.21}$$

□

Proposition 16. Let $\sigma \in C(\mathbb{T}^2)$ with $\|\sigma\|_\infty = 1$. The face $H_\sigma \cap B$ can be characterized as following. A measure $\mu \in B$ is in $H_\sigma \cap B$ if and only if $\sigma\mu$ is a positive measure supported on the set $\{\mathbf{x} \in \mathbb{T}^2 \mid |\sigma(\mathbf{x})| = 1\}$.

Proof. Given $\mu \in B$, μ is in H_σ if $\int \sigma(t)\,\mathrm{d}\mu(t) = 1$. Assume for contradiction that $\sigma\mu$ is not a positive measure. Then there exists a set $A \subset \mathbb{T}^2$ such that $\mathrm{Re}(\sigma\mu)(A) < |\sigma\mu|(A)$. Then

$$\begin{aligned}
|\sigma\mu|(\mathbb{T}^2) &= |\sigma\mu|(A) + |\sigma\mu|(A^c) && (4.22)\\
&> \mathrm{Re}(\sigma\mu)(A) + \mathrm{Re}(\sigma\mu)(A^c) && (4.23)\\
&= \mathrm{Re}(\sigma\mu)(\mathbb{T}^2) = 1. && (4.24)
\end{aligned}$$

But because $\mu \in B$, $\|\sigma\mu\|_{\mathcal{M}} \leq \|\sigma\|_\infty \|\mu\|_{TV} \leq 1$, which is a contradiction.

Similarly, if μ is not supported on the set $\{\mathbf{x} \in \mathbb{T}^2 \mid |\sigma(\mathbf{x})| = 1\}$, then there exists a constant $\epsilon < 1$ such that $A_\epsilon = \{x \in \mathbb{T}^2 \mid |\sigma(x)| < \epsilon\}$ has non-zero $|\mu|$-measure. Then

$$1 = \sigma(\mu) = \int_{A_\epsilon} \sigma(t)\,\mathrm{d}\mu(t) + \int_{A_\epsilon^c} \sigma(t)\,\mathrm{d}\mu(t) \tag{4.25}$$

$$\leq \epsilon|\mu|(A_\epsilon) + |\mu|(A_\epsilon^c) \tag{4.26}$$

$$< |\mu|(A_\epsilon) + |\mu|(A_\epsilon^c) \leq 1, \tag{4.27}$$

which is a contradiction. The proof is complete. □

We can conclude that faces of $\mathcal{M}(\mathbb{T}^2)$ are characterized by continuous functions with $\|\sigma\|_\infty = 1$. Now consider the set $F(B)$. We have the following result from Alfsen which links the face structure of B and $F(B)$ [1].

Proposition 17. Let $\phi\colon V \to W$ be a linear map between vector spaces and let $X \subset V$ be a convex set. $H \subset \phi(X)$ is a face of $\phi(X)$ if and only if $\phi^{-1}(H)$ is a face of X.

So because F is a linear map, we conclude that faces of $F(B)$ correspond to faces of B, but not every face of B necessarily corresponds to a face of $F(B)$. For example, the face of B defined by the function $\sigma = 1$ contains all measures of the form $\delta_\mathbf{x}$, $\mathbf{x} \in \mathbb{T}^2$, but the set $\{F(\delta_\mathbf{x}) \mid \mathbf{x} \in \mathbb{T}^2\}$ does not fall in a hyperplane of $\mathbb{C}^{(2N+1)\times(2N+1)}$, so it cannot map to a face of $F(B)$. In fact, in order for σ to represent a face of $F(B)$, it must factor through F, in the sense that there exists a function $\widetilde{\sigma}\colon \mathbb{C}^{(2N+1)\times(2N+1)} \to \mathbb{R}$ such that $\sigma(\mathbf{x}) = \widetilde{\sigma}(F(\mathbf{x}))$.

Now I will prove the main theorems of this section.

Theorem 31. For each sequence $\mathbf{y} \in \mathbb{C}^{(2N+1)\times(2N+1)}$, there exists an finite sum of point measures

$$\mu = \sum_{k=1}^{K} \alpha_k \delta_{\mathbf{x}_k} \quad \alpha_k \in \mathbb{C},\ \mathbf{x}_k \in \mathbb{T}^2, \tag{4.28}$$

such that for $-N \le m, n \le N$, $\widehat{\mu}(m,n) = \mathbf{y}(m,n)$. Furthermore, for all $\nu \in \mathcal{M}(\mathbb{T}^2)$, if $\widehat{\nu}(m,n) = \mathbf{y}(m,n)$ for $-N \le m,n \le N$, then $\|\nu\|_{TV} \ge \|\mu\|_{TV}$.

Proof. Let $0 \ne \mathbf{y} \in \mathbb{C}^{(2N+1)\times(2N+1)}$. Because $F(B)$ is closed, bounded and contains a neighborhood of the origin, we can guarantee that there exists an $\alpha > 0$ such that $\mathbf{y} \in \alpha F(B)$, and that α is the minimal such constant. Without loss of generality, assume that $\alpha = 1$. Then for all $\nu \in \mathcal{M}(\mathbb{T}^2)$, if $F(\nu) = y$, then $\|\nu\|_{TV} \ge 1$. It is sufficient to construct μ satisfying the desired property, such that $\|\mu\|_{TV} = 1$.

Because α is minimal and $F(B)$ is compact, $\mathbf{y}$ must fall in the boundary of $F(B)$. It is a fact that the boundary of a compact convex set is the union of its faces, therefore there exists a hyperplane H and a corresponding face of $F(B)$ such that $\mathbf{y} \in H \cap F(B)$. Recall, as in Propositions 15 and 16, that associated to H is a functional $\sigma \in (\mathbb{C}^{(2N+1)\times(2N+1)})^*$, such that for all $\mathbf{x} \in F(B)$, $\langle \sigma, \mathbf{x} \rangle \le 1$ and $H = \sigma^{-1}(1)$. Likewise, $F^{-1}(H)$ is a supporting hyperplane of B, which is generated by the functional $\sigma \circ F$. By Corollary 1, each extreme point of the face $H \cap F(B)$ is an extreme point of $F(B)$, which corresponds with a unique extreme point of B. Because G contains all its extreme points, we can then characterize it as a closed convex hull of delta measures.

$$\mathrm{ex}(G) = \{\overline{\widehat{\sigma}}(\mathbf{x})\delta_{\mathbf{x}} \in \mathcal{M}(\mathbb{T}^2) \mid |\widehat{\sigma}(\mathbf{x})| = 1\}. \tag{4.29}$$

Because G is isomorphic to a face of the set of positive definite measures, it is a Choquet simplex. We then proceed in an identical manner to the proof of Theorem 30. First I claim that for each $\mathbf{z} \in \mathrm{ex}(F(G))$, there is a unique $\mathbf{x} \in \mathcal{M}(\mathbb{T}^2)$ such that $\mathbf{z} = \overline{\widehat{\sigma}}(\mathbf{x})F(\delta_{\mathbf{x}})$. It follows from the Krein-Milman theorem that $\mathrm{ex}(F(G)) \subset F(\mathrm{ex}(G))$, so we must show that each point in $F(\mathrm{ex}(G))$ is an extreme point of $F(G)$. Let $\mathbf{x} \in \mathbb{T}^2$ such that $\mathbf{z} = \overline{\widehat{\sigma}}(\mathbf{x})F(\delta_{\mathbf{x}})$ is an extreme point of G. Then

$$\mathbf{z}_{mn} = \overline{\widehat{\sigma}}(\mathbf{x})e^{-2\pi i\langle m,n\rangle\cdot\mathbf{x}}, \tag{4.30}$$

and it is straightforward to see that $\|\mathbf{z}\|_2 = 2N + 1$. If we let $\mathbf{z} \neq \mathbf{w} \in F(G)$, $t \in (0, 1)$, then we can compute

$$\left\|t\mathbf{z} + (1-t)\mathbf{w}\right\|_2 < t\|\mathbf{z}\|_2 + (1-t)\|\mathbf{w}\|_2 \leq 2N + 1. \tag{4.31}$$

Therefore for any non-extreme point of $F(G)$, its norm must be strictly less than $2N + 1$, thus every point in $F(\mathrm{ex}(G))$ is extreme.

By Carathéodory's theorem, there exists a representation of $\mathbf{y}$ as a finite convex sum of extreme points of $F(G)$. As we have shown, to each extreme point $\mathbf{z} \in \mathrm{ex}(F(G))$, we can associate a unique $\mathbf{x} \in \mathbb{T}^2$ such that $\mathbf{z} = \overline{\widehat{\sigma}}(\mathbf{x})F(\delta_{\mathbf{x}})$, so we can write $\mathbf{y}$ as a convex sum with weights $\lambda_k > 0$:

$$\mathbf{y} = \sum_{k=1}^{K} \lambda_k \overline{\widehat{\sigma}}(\mathbf{x}_k) F(\delta_{\mathbf{x}_k}) = F\left(\sum_{k=1}^{K} \lambda_k \overline{\widehat{\sigma}}(\mathbf{x}_k) \delta_{\mathbf{x}_k}\right). \tag{4.32}$$

Finally, because each $\mathbf{x}_k$ is distinct, we can compute that

$$\left\|\sum_{k=1}^{K} \lambda_k \overline{\widehat{\sigma}}(\mathbf{x}_k) \delta_{\mathbf{x}_k}\right\|_{TV} = \sum_{k=1}^{K} \lambda_k |\widehat{\sigma}(\mathbf{x}_k)| = 1. \tag{4.33}$$

We have successfully constructed μ as desired. $\square$

Now we have shown that we cannot hope to uniquely recover non-discrete measures via the total variation minimization in (1.70). My final result is a characterization of exactly which measures allow for unique reconstruction. As we already know, any such measure must be supported on a finite set, but that is still not sufficient. Recall the work of Candès and Fernandez-Granda [12, 13, 30], which showed that a minimum separation requirement was sufficient. My characterization is based on a description of the faces of $F(B)$. Recall that faces of B are generated by continuous functions σ on $\mathbb{T}^2$, and that it corresponds to a face of $F(B)$ if the function σ factors through F. Since unique reconstruction depends on the face structure of $F(B)$, then we can characterize them by describing the admissible functions σ which generate faces of $F(B)$ which are also simplices.

For the proof of this theorem I first need the following lemma.

Lemma 32. Each face of B is a Choquet simplex.

Proof. Let $G = B \cap H_\sigma$ be a face of B. By Proposition 16, $\sigma \in C(\mathbb{T}^2)$ with $\|\sigma\|_\infty = 1$.

Let P be the set of probability measures on $\mathcal{M}(\mathbb{T}^2)$, and consider the linear operator $\sigma_*\colon G \to P$ given by

$$\sigma_*(\mu) = \sigma\mu. \tag{4.34}$$

Note that σ_* is invertible, because for any $\mu \in G$, $\overline{\sigma}\sigma\mu = \mu$. On the other hand for all $\nu \in P$, $\nu \in \sigma_*(G)$ if and only if $|\sigma|\nu = \nu$. Therefore we can conclude that σ_* is a linear isomorphism between G and $P \cap H_{|\sigma|}$. Because $P \cap H_{|\sigma|}$ is a face of a Choquet simplex, by Proposition 14 it is also a simplex, and likewise G is as well. □

Theorem 33. Let $\mu \in \mathcal{M}(\mathbb{T}^2)$. μ is the unique solution to the algorithm in (1.70) if there exists a trigonometric polynomial on $\mathbb{T}^2$,

$$\sigma = \sum_{m,n=-N}^{N} c_{mn} e^{-2\pi i \langle m,n\rangle \cdot \mathbf{x}} \qquad c_{mn} \in \mathbb{C}, \tag{4.35}$$

with the following properties:

1. $\|\sigma\|_\infty = 1$
2. $\sigma\mu = |\mu|$
3. The set $S = \{\overline{\sigma(\mathbf{x})}\delta_{\mathbf{x}} \in \mathbb{T}^2 \mid |\sigma(\mathbf{x})| = 1\}$ is finite, and $F(S)$ is affinely independent.

Proof. Assume without loss of generality that $\|\mu\|_{TV} = 1$. The proof will be in four parts.

(Part 1) First, it follows immediately from Proposition 16 that properties 1 and 2 are equivalent to the statement that μ is in a face G of B generated by σ.

(Part 2) Next I claim that any σ satisfying properties 1 and 2 is of the form (4.35) if and only if it factors through F, which likewise is true if and only if $F(G)$ is a face of $F(B)$.

Let $\sigma \in C(\mathbb{T}^2)$, with $\|\sigma\|_\infty = 1$. σ factors through F if and only if there exists a covector $\widehat{\sigma} \in (\mathbb{C}^{(2N+1)\times(2N+1)})^*$ such that for any $\mu \in \mathcal{M}(\mathbb{T}^2)$, $\sigma(\mu) = \widehat{\sigma}(F(\mu))$. This expands to

$$\int_{\mathbb{T}^2} \sigma(t)\, \mathrm{d}\overline{\mu(t)} = \sum_{m,n=-N}^{N} \widehat{\sigma}(m,n)\overline{\widehat{\mu}(m,n)}. \tag{4.36}$$

It is clear from this equality that $\widehat{\sigma}$ must be the Fourier transform of σ and by Parseval's theorem equality holds for all $\mu \in \mathcal{M}(\mathbb{T}^2)$ if and only if the Fourier transform of σ is supported on $\mathbb{C}^{(2N+1)\times(2N+1)}$, in which case it is a trigonemetric polynomial of the form in (4.35). $H_{\widehat{\sigma}}$ is clearly a supporting hyperplane of $F(B)$, and $F(G) = F(B) \cap H_{\widehat{\sigma}}$. The reverse direction is true trivially from Proposition 17. Thus the claim is proved.

(Part 3) $\mathbf{y}$ is a unique convex combination of extreme points of the face $F(G)$ if $F(S)$ is affinely independent.

$\overline{\sigma(\mathbf{x})}\delta_{\mathbf{x}} \in \mathrm{ex}(G)$ if and only if $|\sigma(\mathbf{x})| = 1$, so $S = \mathrm{ex}(G)$. Therefore if $F(S)$ is affinely independent, $F(G)$ is a simplex. Therefore by Carathéodory's theorem, we can write $\mathbf{y}$ as a unique convex combination of extreme points of G. Because every representation of $\mathbf{y}$ must be supported on G, this combination must be unique over all representations on $\mathrm{ex}(F(B))$.

(Part 4) My next claim is that μ must be the unique solution to (1.70) if $\mathbf{y} = F(\mu)$

can be written as a unique sum of extreme points of $F(G)$.

In the case that $\mathbf{y}$ is not in the boundary of $F(B)$, there exists some $0 < \alpha < 1$ such that $\mathbf{y} \in \alpha F(B)$. Then there exists some $\nu \in \alpha B$ such that $F(\nu) = \mathbf{y}$. $\|\nu\|_{TV} \leq \alpha < \|\mu\|_{TV}$, so μ is not a solution to (1.70). Likewise the pushforwards $F_*\mu$ and $F_*\nu$ are distinct measures supported on $\mathrm{ex}(F(B))$, which each represent $\mathbf{y}$. Hence the claim holds.

In the case that $\mathbf{y}$ is in the boundary of $F(B)$, there exists some supporting hyperplane $H_{\widehat{\sigma}} \subset \mathbb{C}^{(2N+1)\times(2N+1)}$ such that $\mathbf{y}$ is in the face $F(G) = F(B) \cap H_{\widehat{\sigma}}$. By Proposition 17, $G \cap B$ is also a face of B, which is generated by a supporting hyperplane H_σ. Let $\nu \in G \cap B$. By Proposition 32, $G \cap B$ is a Choquet simplex, so by the Choquet-Meyer theorem there is a unique probability measure $\widetilde{\nu} \in (\mathcal{M}(\mathbb{T}^2))^*$, supported on $\mathrm{ex}(G)$, which represents ν. By Lemma 29, $F\restriction_G$ is a homeomorphism, so F induces a vector space isomorphism F_* between $\mathcal{M}(\mathrm{ex}(G))$ and $\mathcal{M}(\mathrm{ex}(F(G)))$. For any functional A on $\mathbb{C}^{(2N+1)\times(2N+1)}$, see that

$$\int_{\mathbb{C}^{(2N+1)\times(2N+1)}} A(t)\,\mathrm{d}F_*\widetilde{\nu}(t) = \int_{\mathcal{M}(\mathbb{T}^2)} A(F(s))\,\mathrm{d}\widetilde{\nu}(s) = A(F(\nu)). \tag{4.37}$$

Because A was arbitrary, $F(\nu) = \mathbf{y}$ if and only if $F_*\widetilde{\nu}$ represents $\mathbf{y}$. Because F_* is an isomorphism of vector spaces, we can conclude that ν is the unique measure in B such that $F(\nu) = \mathbf{y}$ if and only if $F_*\widetilde{\nu}$ is the unique measure in $\mathcal{M}(\mathrm{ex}(G))$ which represents $\mathbf{y}$. In addition, by Theorem 31, if ν is unique then the measure $F_*\widetilde{\nu}$ must be a finite sum.

The proof is complete.

$\square$

A few notes on the preceding theorem. It may be possible to loosen the last requirement on σ slightly. Assume that $\mathrm{ex}(G)$ is finite. Using tools from algebraic geometry, we may put a bound on the number of extreme points of G. Let $\phi = (1-\sigma\overline{\sigma})$. The zero set of ϕ is identical to the set $\mathrm{ex}(F^{-1}(G))$. Under the assumption that this set is finite, Bézout's theorem gives us a bound. Because ϕ has degree at most $4N$, Bézout's theorem says that it can have at most $4N$ zeros. For a more in depth derivation of this result, see appendix A.3 of [49].

Now $F(G)$ is a simplex if and only if the set of extreme points of $F(G)$ is linearly independent in $\mathbb{C}^{(2N+1)\times(2N+1)}$. This may not seem like a simplification, but for any given polynomial σ we can guarantee this to be true for sufficiently large N. Consider that for $\delta_{\mathbf{x}} \in \mathrm{ex}(G)$, as $N \to \infty$, $\mathcal{F}^{-1}(F(\delta_{\mathbf{x}}))$ approaches the delta function $\delta_{\mathbf{x}}$, and it is easy to see that any finite set of Dirichlet kernels, $\mathcal{F}(F(\delta_{\mathbf{x}_k})$ for $\mathbf{x}_k \in \mathbb{T}^2$, will evenually be linearly independent for sufficiently large N. How large N must be relative to the degree of σ is unknown to the author's knowledge.

Finally, we conclude with one more useful result that is an immediate application of the preceding theorem.

Corollary 2. Any pair of positive delta measures can be uniquely recovered from (1.70), given $N \geq 2$.

Proof. Let $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2), \mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2) \in \mathbb{T}^2$. Define the polynomial $\rho(s,t) = (1 - cos(s) - cos(t))/2$. It is easy to see that $0 \leq \rho \leq 1$, and $\rho = 0$ if and only if

$s = t = 0$. Define

$$\sigma(s,t) = 1 - \rho(s - \mathbf{x}_1, t - \mathbf{x}_2)\rho(s - \mathbf{y}_1, t - \mathbf{y}_2). \tag{4.38}$$

$0 \leq \sigma \leq 1$, and $|\sigma(\mathbf{z})| = 1$ only when $\mathbf{z} = \mathbf{x}$ or $\mathbf{y}$. Therefore for any positive measure supported on $\{\mathbf{x}, \mathbf{y}\}$, $\mu\sigma = |\mu|$. Finally, note that as $\mathbf{x}$ and $\mathbf{y}$ are distinct, $\{F(\mathbf{x}), F(\mathbf{y})\}$ is trivially affinely independent. Therefore σ satisfies the properties for Theorem 33, and we can conclude that any positive measure supported on $\{\mathbf{x}, \mathbf{y}\}$ can be uniquely recovered by (1.70). $\square$

Bibliography

[1] Erik M. Alfsen, *On the geometry of Choquet simplexes*, Math. Scand. **15** (1965), no. 1, 97–110.

[2] A. P. Artemenko, *Hermitian positive functions and positive funcctionals*, Ph.D. thesis, Odessa State University, 1983.

[3] Joseph A. Ball and Moisés D. Guerra Huamán, *Convexity analysis and the matrix-valued Schur class over finitely connected planar domains*, J. Operator Theory **70** (2013), no. 2, 531–571.

[4] John J. Benedetto, *Harmonic Analysis and Applications*, CRC Press, 1996.

[5] John J. Benedetto and Wojciech Czaja, *Integration and Modern Analysis*, Birkhäuser Advanced Texts, Birkhäuser, Basel, 2009.

[6] John J. Benedetto and Weilin Li, *Super-resolution by means of Beurling minimal extrapolation*, 2016.

[7] Sergei N. Bernstein, *Sur les fonctions absolument monotones*, Acta Math. **52** (1928), 1–66.

[8] Salomon Bochner, *Vorlesungen über Fouriersche Integrale*, first ed., Akademische Verlagsgesellschaft, Leipzig, 1932.

[9] ______, *Monotone Funktionen, Stieltjessche Integrale und harmonische Analyse*, Math. Ann. **108** (1933), 378–410.

[10] Stephen Boyd and Lieven Vandenberghe, *Convex optimization*, first ed., pp. 215–232, Cambridge University Press, Cambridge, 2004.

[11] Alberto P. Calderón and Raymond Pepinsky, *On the phases of Fourier coefficients for positive real periodic functions*, Computing Methods and the Phase Problem in X-Ray Crystal Analysis (Raymond Pepinsky, ed.), The X-Ray Crystal Analysis Laboratory, Department of Physics, The Pennsylvania State College, 1952, pp. 339–348.

[12] Emmanuel Candès and Carlos Fernandez-Granda, *Super-resolution from noisy data*, J. Fourier Anal. Appl. **19** (2013), 1229–1254.

[13] ———, *Towards a mathematical theory of super-resolution*, Comm. Pure Appl. Math. **67** (2013), no. 6, 906–956.

[14] Emmanuel Candès, Justin Romberg, and Terence Tao, *Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory **52** (2006), no. 2, 489–509.

[15] ———, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math. **59** (2006), 1207–1223.

[16] Emmanuel Candès and Terence Tao, *Decoding by linear programming*, IEEE Trans. Inform. Theory **51** (2005), no. 12, 4203–4215.

[17] ———, *Near optimal signal recovery from random projections: universal encoding strategies?*, IEEE Trans. Inform. Theory **52** (2006), no. 12, 5406–5425.

[18] Emmanuel J. Candès, Justin Romberg, and Terence Tao, *Signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math **59** (2005), no. 8, 1207–1223.

[19] Constantin Carathéodory, *Über den Variabilitätsbereich der Koeffizienten von Potenzreihen, die gegebene Werte nicht annehmen*, Math. Ann. **64** (1907), 95–115.

[20] ———, *Über den Variabilitätsbereich der Fourierschen Konstanten von positiven harmonischen Funktionen*, Rend. Circ. Mat. Palermo **32** (1911), 193–217.

[21] Gustave Choquet, *Theory of capacities*, Ann. Inst. Fourier **5** (1955), 131–295.

[22] ———, *Existence et unicité des representations intégrales au moyen des points extrémaux dans cônes convexes*, Séminaire Bourbaki : années 1956/57 - 1957/58, exposés 137-168, Séminaire Bourbaki, no. 4, Société mathématique de France, 1958.

[23] Gustave Choquet and Paul-André Meyer, *Existence et unicité des representations intégrals dans les convexes compacts quelconques*, Ann. Inst. Fourier **13** (1963), 139–154.

[24] Krzysztof J. Ciosmak, *Optimal transport and Choquet theory*, 2020.

[25] Yohann de Castro and Fabrice Gamboa, *Exact reconstruction using Beurling minimal extrapolation*, J. Math. Anal. Appl. **395** (2012), no. 1, 336–354.

[26] Gaspard de Prony, *Essai éxperimantal et analytique: sur les lois de la dilatabilité de fluides élastique et sur celles de la force expansive de la vapeur de l'alkool, à différentes températures*, J. É. polytech. Math. **1** (1795), no. 22, 24–76.

[27] Ronald A. DeVore, *Nonlinear Approximation*, vol. 7, Cambridge University Press, 1998.

[28] David A. Edwards, *Some applications of Choquet theory to harmonic analysis*, J. Lond. Math. Soc. **59** (1999), no. 2, 657–668.

[29] Herbert Federer, *Geometric Measure Theory*, vol. 153, Springer, Berlin, 1969.

[30] Carlos Fernandez-Granda, *Super-resolution of point sources via convex programming*, Information and Inference **5** (2016), no. 3, 251–303.

[31] Robert Fraser and Kyle Hambrook, *Explicit Salem sets in* $\mathbb{R}^n$, 2019.

[32] ———, *Explicit salem sets in* $\mathbb{R}^n$, 2020.

[33] David Hilbert, *Über die Darstellung definiter Formen als Summe von Formenquadraten*, Mathem. Annalen **32** (1888), 342–350.

[34] Jean-Pierre Kahane, *Some Random Series of Functions*, Cambridge Studies in Advanced Mathematics, vol. 5, Cambridge University Press, Cambridge, 1993.

[35] Jean-Pierre Kahane and Raphael Salem, *Ensembles Parfaits et Séries Trigonométriques*, vol. 7, Cambridge University Press, 1964.

[36] R. Kaufman, *On the theorem of jarník and besicovitch*, Acta Arith. **39** (1981), no. 3, 265–267.

[37] Aleksandr Khinchin, *Korrelationstheorie der stationären stochastischen Prozesse*, Math. Ann. **109** (1934), 604–615.

[38] Mark Krein, *Sur le problème du prolongement des fonctions hermitiennes positives et continues*, Dokl. Akad. Nauk SSSR **26** (1940), 17–22.

[39] Mark Krein and David Milman, *On extreme points of regular convex sets*, Studia Math. **9** (1940), no. 1, 133–138.

[40] H. Landau, *Extrapolating a band-limited function from its samples taken in a finite interval*, IEEE Trans. Inf. Theory **32** (1986), 464–470.

[41] H. J. Landau, *Maximum entropy and the moment problem*, Bull. Amer. Mth. Soc. (N.S.) **16** (1987), no. 1, 47 – 77.

[42] Paul Lévy, *Théorie de l'Addition des Variables Aléatoires*, first ed., Gauthier-Villars, Paris, 1937.

[43] Xiaodong Li, Shuyang Ling, Thomas Strohmer, and Ke Wei, *Rapid, robust, and reliable blind deconvolution via nonconvex optimization*, Appl. Comput. Harmon. Anal. **47** (2018), no. 3, 893–934.

[44] Stéphane Mallat and Guoshen Yu, *Super-resolution with sparse mixing estimators*, IEEE Trans. Image Process. **19** (2010), no. 11, 2889–2900.

[45] Hassan Mansour, Dehong Liu, Ulugbek S. Kamilov, and Petros T. Boufounos, *Sparse blind deconvolution for distributed radar autofocus imaging*, 2018.

[46] Toru Maruyama, *Herglotz-Bochner representation theorem via theory of distributions*, J. Oper. Res. Soc. Japan **60** (2017), no. 2, 122–135.

[47] Pertti Mattila, *Fourier Analysis and Hausdorff Dimension*, first ed., Cambridge Studies in Advanced Mathematics, vol. 150, Cambridge University Press, Cambridge, 2015.

[48] Greg Ongie, Sampurna Biswas, and Mathews Jacob, *Convex recovery of continuous domain piecewise constant images from nonuniform Fourier samples*, IEEETrans. Sitnal Process. **66** (2018), no. 1, 236–250.

[49] Greg Ongie and Mathews Jacob, *Off-the-grid recovery of piecewise constant images from few Fourier samples*, SIAM Journal on Imaging Sciences **9** (2016), 1004–1041.

[50] Robert R. Phelps, *Lectures on Choquet's Theorem*, second ed., Lecture Notes in Mathematics, Springer, Berlin, 2001.

[51] Dmitrii A. Raikov, *On the decomposition of Gauss and Poisson laws*, Izv. Akad. Nauk. SSSR **2** (1937), 91–124 (russian).

[52] Walter Rudin, *The extension problem for positive-definite functions*, Illinois J. Math. **7** (1963), no. 3, 532–539.

[53] ______, *An extension theorem for positive-definite functions*, Duke Math. J. **37** (1970), no. 1, 49–53.

[54] Tapan K. Sarkar and Odilon Pereira, *Using the matrix pencil method to estimate the parameters of a sum of complex exponentials*, IEEE Antennas and Propagation Magazine **37** (1995), no. 1, 48–55.

[55] Zoltán Sasvári, *The extension problem for measurable positive definite functions*, Math. Zeitschrift **191** (1986), 475–478.

[56] ______, *The extension problem for positive definite functions. A short historical survey*, Operator theory and indefinite inner product spaces, Oper. Theory Adv. Appl., vol. 163, Birkhäuser, Basel, 2006, pp. 365–379.

[57] Thomas Strohmer and Ke Wei, *Painless breakups–efficient demixing of low rank matrices*, J. Fourier Anal. Appl. **25** (2019), 1–31.

[58] Otto Toeplitz, *Über die Fourier'sche Entwickelung positiver Funktionen*, Rend. Circ. Mat. Palermo **32** (1911), 191–192.

[59] Michael Unser, Julien Fageot, and John Paul Ward, *Splines are universal solutions of linear inverse problems with generalized TV regularization*, SIAM Review **59** (2017), no. 4, 769–793.

[60] David V. Widder, *The Laplace Transform*, Princeton Mathematical Series, vol. 6, Princeton University Press, Princeton, 1946.