SUPER-RESOLUTION BY MEANS OF BEURLING MINIMAL EXTRAPOLATION

JOHN J. BENEDETTO AND WEILIN LI

ABSTRACT. Let $M(\mathbb{T}^d)$ be the space of complex bounded Radon measures defined on the d-dimensional torus group $(\mathbb{R}/\mathbb{Z})^d = \mathbb{T}^d$, equipped with the total variation norm $\| \cdot \|$; and let $\widehat{\mu}$ denote the Fourier transform of $\mu \in M(\mathbb{T}^d)$. We address the super-resolution problem: For given spectral (Fourier transform) data defined on a finite set $\Lambda \subseteq \mathbb{Z}^d$, determine if there is a unique $\mu \in M(\mathbb{T}^d)$ for which $\widehat{\mu}$ equals this data on Λ. Without additional assumptions on μ and Λ, our main theorem shows that the solutions to the super-resolution problem, which we call *minimal extrapolations*, depend crucially on the set $\Gamma \subseteq \Lambda$, defined in terms of μ and Λ. For example, when $\#\Gamma = 0$, the minimal extrapolations are singular measures supported in the zero set of an analytic function, and when $\#\Gamma \geq 2$, the minimal extrapolations are singular measures supported in the intersection of $\binom{\#\Gamma}{2}$ hyperplanes. This theorem has implications to the possibility *and* impossibility of uniquely recovering μ from Λ. We illustrate how to apply our theory to both directions, by computing pertinent analytical examples. These examples are of interest in both super-resolution and deterministic compressed sensing. By theory and example, we show that the case $\#\Gamma = 1$ is different from other cases and is deeply connected with the existence of positive minimal extrapolations. By applying our theorem and examples, we study whether the minimal extrapolations of μ and $T\mu$ are related, for different types of linear operators $T\colon M(\mathbb{T}^d) \to M(\mathbb{T}^d)$. Additionally, our concept of an *admissibility range* fundamentally connects Beurling's theory of minimal extrapolation [Beu89a, Beu89b] with Candès and Fernandez-Granda's work on super-resolution [CFG13, CFG14].

1. INTRODUCTION

1.1. **Motivation.** The term *super-resolution* varies depending on the field, and, consequently, there are various types of super-resolution problems. In some situations [PPK03], super-resolution refers to the process of up-sampling an image onto a finer grid, which is a *spatial interpolation* procedure. In other situations [Lin12], super-resolution refers to the process of recovering the object's high frequency information from its low frequency information, which is a *spectral extrapolation* procedure. In both situations, the super-resolution problem is ill-posed because the missing information can be arbitrary. However, it is possible to provide meaningful super-resolution algorithms by using *prior knowledge* of the data. We develop a mathematical theory of the spectral extrapolation version, which we simply refer to as the *super-resolution problem*, see Problem (SR) for a precise statement.

To be concrete, our exposition focuses on imaging applications, but super-resolution ideas are of interest in other fields, e.g., [Rie99, KLM04]. In such applications, an image is obtained by convolving the object with the *point spread function* of an optical lens. Or alternatively, the Fourier transform of the object is multiplied by a *modulation transfer function*. The resulting image's resolution is inherently limited by the Abbe diffraction limit, which depends on the illumination light's wavelength and on the diameter of the optical lens. Thus, the optical lens acts as a low-pass filter, see [Lin12]. The purpose of super-resolution is to use prior knowledge about the object to obtain an accurate image whose resolution is higher than what can be measured by the optical lens.

We mention two specific imaging problems, and they motivate the theory that we develop.

(a) In astronomy [PK05], each star can be modeled as a complex number times a Dirac δ-measure, and the Fourier transform of each star encodes important information about that star. However,

---

2010 *Mathematics Subject Classification.* 43A25, 46E27, 46N10, 94A08, 94A12.

an image of two stars that are close in distance resembles an image of a single star. In this context, the super-resolution problem is to determine the number of stars and their locations, using the prior information that the actual object is a linear combination of Dirac δ-measures.

(b) In medical imaging [Gre09], machines capture the structure of the patient's body tissues, in order to detect for anomalies in the patient. Their shapes and locations are the most important features, so each tissue can be modeled as the characteristic function of a closed set, or as a surface measure supported on the boundary of a set. The super-resolution problem is to capture the fine structures of the tissues, given that the actual object is a linear combination of singular measures.

1.2. **Problem statement.** Motivated by these applications, we model objects as measures. There are two natural models.

(a) In this paper, we shall address the periodic case. We model objects as elements of $M(\mathbb{T}^d)$, the space of complex bounded Radon measures on $\mathbb{T}^d = (\mathbb{R}/\mathbb{Z})^d$, the d-dimensional torus group. $M(\mathbb{T}^d)$ equipped with the total variation norm $\|\cdot\|$ is a Banach algebra with unit, the Dirac δ-measure, where multiplication is defined by convolution $*$. See [BC09] for further details. Then, an image of $\mu \in M(\mathbb{T}^d)$, produced by the optical lens $\psi\colon \mathbb{T}^d \to \mathbb{C}$, is the function $F = F_{\mu,\psi}\colon \mathbb{T}^d \to \mathbb{C}$, given by

$$F(x) = (\mu * \psi)(x) = \int_{\mathbb{T}^d} \psi(x - y)\, d\mu(y).$$

(b) In the companion paper [BL16a], we shall address the super-resolution problem under the model that objects are elements of $M_b(\mathbb{R}^d)$, the space of bounded Radon measures on Euclidean space $\mathbb{R}^d$, see [BC09]. Then, an image of $\mu \in M_b(\mathbb{R}^d)$ produced by the optical lens $\psi\colon \mathbb{R}^d \to \mathbb{C}$, is the function $G = G_{\mu,\psi}\colon \mathbb{R}^d \to \mathbb{C}$, given by

$$G(x) = (\mu * \psi)(x) = \int_{\mathbb{R}^d} \psi(x - y)\, d\mu(y).$$

From this point onwards, we assume that the image is produced according to the first model. The *Fourier transform* of $\mu \in M(\mathbb{T}^d)$ is the function $\widehat{\mu}\colon \mathbb{Z}^d \to \mathbb{C}$, whose m-th Fourier coefficient is defined to be

$$\widehat{\mu}(m) = \int_{\mathbb{T}^d} e^{-2\pi i m \cdot x}\, d\mu(x).$$

For simplicity, we assume that $\widehat{\psi} = \mathbb{1}_\Lambda$, the characteristic function of some finite set $\Lambda \subseteq \mathbb{Z}^d$. Then, we have

$$F = (\mu * \psi) = (\widehat{\mu}\,|_\Lambda)^\vee.$$

Thus, we interpret $\widehat{\mu}\,|_\Lambda$ as the given low frequency information of μ, and $(\widehat{\mu}\,|_\Lambda)^\vee$ as the given low resolution image, even though in applications, we do not know the desired object μ.

We consider the super-resolution problem: For a given $\mu \in M(\mathbb{T}^d)$ and finite subset $\Lambda \subseteq \mathbb{Z}^d$, find

(SR) $$\underset{\nu \in M(\mathbb{T}^d)}{\arg\min} \|\nu\|, \quad \text{subject to } \widehat{\nu} = \widehat{\mu} \text{ on } \Lambda.$$

We interpret ν as a "simple" or "least complicated" high resolution extrapolation of the low resolution image. This version of the spectral extrapolation super-resolution problem was first studied in [CFG14].

Problem (SR) is a *convex minimization problem*. In practice, the primary objective is to recover the unknown μ from its partial frequency information, $\widehat{\mu}\,|_\Lambda$. However, if μ is not the unique solution to the minimization problem, then the output of a numerical algorithm is not guaranteed to approximate μ. Thus, it is important to determine sufficient conditions such that μ is the unique solution. For this reason, we say that *super-resolution reconstruction of μ from Λ is possible*, or *it*

is possible to super-resolve μ from Λ, if and only if μ is the unique solution to Problem (SR). Of course, it could be theoretically be possible to reconstruct μ by other means.

1.3. **Our approach.** Compressed sensing results typically state that, under certain conditions, it is possible to recover, deterministically or probabilistically, a sparse vector from incomplete linear measurements by solving a ℓ^1 minimization problem, e.g., [CRT06a, CRT06b, Don06]. Compressed sensing and super-resolution are connected because a vector can be thought of as a linear combination of equally spaced Dirac δ-measures, and the total variation norm is the natural generalization of the ℓ^1 norm, see Section 2.2. As a result of this connection, some recent literature has focused on determining sufficient conditions such that super-resolution of a discrete $\mu \in M(\mathbb{T}^d)$ from a finite set $\Lambda \subseteq \mathbb{Z}^d$ is possible, e.g., [DCG12, CFG14, CFG13, TBSR13, ADCG15, TBR15, FG15].

While compressed sensing suggests that it is possible to recover discrete measures by a total variation minimization process, Beurling's theorem on minimal extrapolation, e.g., Theorem 2.1, suggests that it is also possible to recover singular measures by the same total variation minimization process. Inspired by Beurling, we are interested in solving Problem (SR) for an *arbitrary* $\mu \in M(\mathbb{T}^d)$ and an *arbitrary* finite subset $\Lambda \subseteq \mathbb{Z}^d$.

Theorem 3.3 is our main result. It is an adaptation to the torus, and a generalization to higher dimensions, of Beurling's theorem on minimal extrapolation. As we shall see, applying Beurling's ideas to our situation is not immediate because solutions to Problem (SR) greatly depend on the dimension, and there are significant differences between working with $\mathbb{T}^d$ versus $\mathbb{R}^d$.

Theorem 3.3 has applications to the possibility *and* impossibility of super-resolution reconstruction.

(a) On the *possibility* of super-resolution reconstruction, we have at least two distinct advantages compared to previous results. First, we do not place additional restrictions on μ, so our theory includes discrete measures that do not have restrictions on their supports, and singular continuous measures that are of interest in applications, e.g., [Gre09]. We include some examples on the super-resolution of singular continuous measures, but since this paper focuses on building the preliminary theory, we shall carefully address singular continuous measures in the sequel [BL16b]. Second, we only assume $\Lambda \subseteq \mathbb{Z}^d$ is finite; in particular, Λ does not have to be of the form $\{-\lambda, -\lambda + 1, \ldots, \lambda\}^d$, and so, our results apply to the non-uniform sampling regime.

(b) One of our main strengths is being able to determine the *impossibility* of super-resolution reconstruction, which has not received much attention. To see why this problem is important, assume the object is $\mu = \sum_{k=1}^{K} a_k \delta_{x_k} \in M(\mathbb{T})$ and $\Lambda = \{-\lambda, -\lambda + 1, \ldots, \lambda\} \subseteq \mathbb{Z}$. Suppose we obtain a numerical solution ν to Problem (SR). How can we tell if $\mu \approx \nu$? Since we do not know μ, but only the low resolution version, $(\widehat{\mu}\,|_\Lambda)^\vee$, we are not certain that μ satisfies a separation condition as is required in [Don92, CFG14, CFG13, TBSR13, ADCG15, TBR15, FG15], or that the support of μ satisfies some other additional property, e.g., [DCG12]. It would be unfortunate if ν is also discrete, say $\nu = \sum_{j=1}^{J} b_j \delta_{y_j} \in M(\mathbb{T})$, and we erroneously mistook ν for μ!

We shall demonstrate how to apply Theorem 3.3 to compute pertinent *analytical* examples as opposed to giving only simulations, see Section 4.

(a) Several of our examples can be interpreted in the context of deterministic compressed sensing. In Section 2.2, we shall discuss the connection between super-resolution and deterministic compressed sensing.

(b) Since the super-resolution problem is nonlinear, it is not immediately clear whether solutions of Problem (SR) for μ are related to solutions of Problem (SR) for $T\mu$, where $T \colon M(\mathbb{T}^d) \to M(\mathbb{T}^d)$ is a linear operator. In Proposition 3.20, we shall demonstrate that, for several kinds of T, in general, their solutions are not related.

(c) Numerical simulations have suggested that, in general, a minimal separation condition is necessary to super-resolve two discrete measures, e.g., see [Don92, CFG14] and references therein. In Example 4.6, we shall prove that a minimal separation condition is, in fact, necessary in order to super-resolve the sum of two Dirac δ-measures.

Moreover, our results unify the classical, e.g., [Beu89a, Beu89b], and the modern, e.g., [CFG14, FG15], analysis of Problem (SR), through a concept which we shall refer to as the *admissibility range for ε*, see Proposition 3.15. We shall demonstrate that our results are strongest for the lowest endpoint of the admissibility range, whereas Candès and Fernandez-Granda's results only apply to the upper endpoint of the admissibility range, see Remark 3.17.

1.4. **Notation and terminology.** Since referring to Problem (SR) can be ambiguous when simultaneously working with several different measures, we introduce more precise notation and terminology that we use throughout the paper. For $\mu \in M(\mathbb{T}^d)$ and finite set $\Lambda \subseteq \mathbb{Z}^d$, we define

$$(1.1) \qquad \varepsilon = \varepsilon(\mu, \Lambda) = \inf\{\|\nu\| : \nu \in M(\mathbb{T}^d) \text{ and } \widehat{\mu} = \widehat{\nu} \text{ on } \Lambda\},$$

$$(1.2) \qquad \mathcal{E} = \mathcal{E}(\mu, \Lambda) = \{\nu \in M(\mathbb{T}^d) : \widehat{\mu} = \widehat{\nu} \text{ on } \Lambda \text{ and } \|\nu\| = \varepsilon(\mu, \Lambda)\},$$

$$(1.3) \qquad \Gamma = \Gamma(\mu, \Lambda) = \{m \in \Lambda : |\widehat{\mu}(m)| = \varepsilon(\mu, \Lambda)\}.$$

We shall see in Proposition 3.1 that $\mathcal{E} \neq \emptyset$. We shall prove in Proposition 3.15 that $\|\widehat{\mu}\|_{\ell^\infty(\Lambda)} \leq \varepsilon$, and so Γ *is the set of multi-integers that belongs to Λ, for which $|\widehat{\mu}|$ attains its maximum possible value.*

We say ν is a *extrapolation of μ from Λ* provided that $\widehat{\mu} = \widehat{\nu}$ on Λ. If ν is an extrapolation of μ from Λ and $\|\nu\| = \varepsilon$, then we say ν is a *minimal extrapolation of μ from Λ*. Hence, $\mathcal{E}$ is the *set of all minimal extrapolations of μ from Λ*, or equivalently, $\mathcal{E}$ is the set of solutions to Problem (SR). When there is no ambiguity, we shall say that "ν is an extrapolation (respectively, a minimal extrapolation)" as a shortened version of "ν is an extrapolation (respectively, a minimal extrapolation) of μ from Λ".

1.5. **Outline.** Section 2.1 compares our results with that of the literature and Section 2.2 explains the connection between super-resolution and deterministic compressed sensing. Section 3 contains the basis of our mathematical theory. Propositions 3.1 and 3.2 are basic results, and we use them to prove our main theorem, Theorem 3.3. Proposition 3.8 provides a uniqueness criterion for discrete minimal extrapolations, and it works well in conjunction with the theorem. Further, Proposition 3.9 connects non-uniqueness of minimal extrapolation with the existence of positive measures. We introduce the notion of *admissibility range* for ε immediately before Proposition 3.15, that in turn proves lower and upper bounds on ε. Propositions 3.19 and 3.20 examine whether it is possible to relate the minimal extrapolations of μ with those of $T\mu$ for various operators T. Section 4 contains the examples that we mentioned in Section 1.3, including the minimal extrapolations of discrete measures and of singular continuous measures.

2. Background

2.1. **The Beurling and the Candès and Fernandez-Granda theories.** Beurling called the super-resolution problem the *minimal extrapolation problem*, which explains our choice of terminology. Let $\widehat{\mathbb{R}} = \mathbb{R}$ be the dual group of $\mathbb{R}$. Note the similarities between (1.1)-(1.3) and (2.1)-(2.3).

Theorem 2.1 (Beurling, Theorem 2, page 362, [Beu89b])**.** *Let $\mu \in M_b(\mathbb{R})$ and let $\Lambda = [-\lambda, \lambda] \subseteq \widehat{\mathbb{R}}$ for some $\lambda > 0$. Define*

$$(2.1) \qquad m = \inf\{\|\nu\| : \nu \in M_b(\mathbb{R}) \text{ and } \widehat{\mu} = \widehat{\nu} \text{ on } \Lambda\},$$

$$(2.2) \qquad M = \{\nu \in M_b(\mathbb{R}) : \widehat{\mu} = \widehat{\nu} \text{ on } \Lambda \text{ and } \|\nu\| = m\},$$

$$(2.3) \qquad \Lambda_m = \{\gamma \in \Lambda : |\widehat{\mu}(\gamma)| = m\}.$$

Then, M is non-empty. We say ν is a minimal extrapolation if and only if $\nu \in M$. Let $\#\Lambda$ denote the cardinality of Λ.

(a) *Suppose $\#\Lambda_m = 0$. There exist sequences, $\{a_k\} \subseteq \mathbb{C}$ and $\{x_k\} \subseteq \mathbb{R}$, for which $\#\{x_k\colon |x_k| < r\} = O(r)$, $r \to \infty$, and such that*

$$\nu = \sum_{k=1}^{\infty} a_k \delta_{x_k}$$

is the unique minimal extrapolation of μ from Λ_m.

(b) *Suppose $\#\Lambda_m \geq 2$ and $\Lambda_m \neq \Lambda$. Then, Λ_m is a finite set, which allows us to define $\tau > 0$ as the smallest distance between any two points in Λ_m. Further, there exist $\{a_k\} \subseteq \mathbb{C}$ and $x_0 \in \mathbb{R}$, such that*

$$\nu = \sum_{k=-\infty}^{\infty} a_k \delta_{x_0 + \frac{k}{\tau}}$$

is the unique minimal extrapolation of μ from Λ_m.

(c) *If $\Lambda_m = \Lambda$, then there exist $\alpha \in \mathbb{R}/\mathbb{Z}$ and $x \in \mathbb{R}$, such that $\nu = m e^{2\pi i \alpha} \delta_x$ is the unique minimal extrapolation of μ from Λ_m.*

Remark 2.2. It is difficult to deduce information about the minimal extrapolations when $\#\Lambda_m = 1$. In this case, the minimal extrapolations might not be unique and there may exist positive absolutely continuous minimal extrapolations, e.g., see [Beu89b, Ess45] for specific examples.

The important paper by Candés and Fernandez-Granda [CFG14] shows that super-resolution reconstruction of a discrete μ is possible if μ satisfies a certain minimal separation condition. More precisely, given $\mu = \sum_{k=1}^{K} a_k \delta_{x_k} \in M(\mathbb{T}^d)$ and $\Lambda = \{-\lambda, -\lambda + 1, \ldots, \lambda\}^d \subseteq \mathbb{Z}^d$, we say μ *satisfies the minimum separation condition with constant $C > 0$* if

$$(2.4) \qquad \inf_{\substack{1 \leq k, k' \leq K \\ k \neq k'}} \|x_k - x_{k'}\|_{\ell^\infty(\mathbb{T}^d)} \geq \frac{C}{\lambda}.$$

Theorem 2.3 (Candès and Fernandez-Granda, Theorems 1.2-1.3, [CFG14]. Fernandez-Granda, Theorem 2.2, [FG15]). *Let $\mu = \sum_{k=1}^{K} a_k \delta_{x_k} \in M(\mathbb{T}^d)$ and $\Lambda = \{-\lambda, -\lambda + 1, \ldots, \lambda\}^d \subseteq \mathbb{Z}^d$.*

(a) *If $d = 1$, $\lambda \geq 128$, and μ satisfies the minimum separation condition with $C = 2$, then μ is the unique solution to Problem (SR). If additionally, μ is real valued, then the constant can be reduced to $C = 1.87$.*

(b) *If $d = 1$, $\lambda \geq 10^3$, and μ satisfies the minimum separation condition with $C = 1.26$, then μ is the unique solution to Problem (SR).*

(c) *If $d = 2$, μ is real valued, $\lambda \geq 512$, and μ satisfies the minimum separation condition for $C = 2.38$, then μ is the unique solution to Problem (SR).*

Remark 2.4. The separation condition (2.4) for $d \geq 2$ might initially seem strange because the distance between two points in the support of a discrete μ is measured with respect to the ℓ^∞ norm as opposed to the more natural ℓ^2 norm. This is a consequence of the authors' proof of [CFG14, Theorem 1.2 and Theorem 1.3]. Their argument is based on interpolating $\text{sign}(\mu)$, see Section 3.1 for the definition, under the assumption that μ satisfies the minimum separation condition for a sufficiently large constant. To accomplish this task for $d = 1$, they use the one dimensional Fejér kernel because the Fejér kernel, and its derivatives, decay rapidly away from the origin. For $d = 2$, the authors used the two-dimensional Fejér kernel, which is a tensor product of one-dimensional Fejér kernels. To guarantee that the two-dimensional Fejér kernel and its partial derivatives are small away from the origin, it is sufficient that only one of the two coordinates is far from the origin, which explains the appearance of the ℓ^∞ norm in the separation condition. The authors remarked that if a radial kernel was chosen instead, then it would be more natural to use the ℓ^2 norm in the minimum separation condition instead of the ℓ^∞ norm.

Recent interest in the super-resolution problem has grown significantly. We mention a few papers, and this list is certainly not exhaustive. As mentioned previously, the current literature focuses on super-resolution reconstruction of a discrete μ. Results of this type can be traced back to Donoho [Don92], who showed that it is possible to super-resolve a discrete measure supported in a one-dimensional lattice, if the distance between lattice points is sufficiently large. Since then, variations and generalizations of this minimal separation assumption have appeared, e.g., [Kah12, DCG12, CFG14, ASB14, ASB15, FG15].

De Castro-Gamboa [DCG12] also invokes Beurling. Adapting Beurling's ideas, they showed that it is possible to super-resolve a *discrete* measure μ, whose support is contained in the level set of a certain family of generalized polynomials, given partial *generalized moments* of μ. In contrast to their problem and techniques, we shall use Beurling's ideas to study super-resolution reconstruction of an *arbitrary* bounded measure μ, given a finite subset of the *Fourier coefficients* of μ.

Aubel-Stotz-Bölcskei [ASB14, ASB15] considered the super-resolution problem for both $\mathbb{R}$ and $\mathbb{T}$, for the *short-time Fourier transform* instead of the Fourier transform. We state their result for $\mathbb{T}$. For $g \in C^\infty(\mathbb{T})$, the short-time Fourier transform $A_g \colon M(\mathbb{T}) \to L^1(\mathbb{T} \times \mathbb{Z})$ is defined as

$$A_g \mu(x, m) = \int_{\mathbb{T}} \overline{g(y - x)} e^{-2\pi i \gamma y} \, d\mu(y).$$

Then, given a discrete $\mu \in M(\mathbb{T})$, $\Lambda = \{-\lambda, -\lambda + 1, \ldots, \lambda\} \subseteq \mathbb{Z}$, and an appropriate periodic Gaussian g depending on λ, if μ satisfies the minimum separation condition for a sufficiently large constant, then μ is the unique solution to

$$\underset{\nu \in M(\mathbb{T})}{\arg\min} \|\nu\|, \quad \text{subject to} \quad \forall x \in \mathbb{T} \text{ and } \forall m \in \Lambda, \quad A_g \nu(x, m) = A_g \mu(x, m).$$

There has not been much work on the super-resolution of non-discrete measures. Esseen [Ess45, Ess54] studied the same problem as Beurling, but for $\Lambda = (-\infty, \lambda) \subseteq \widehat{\mathbb{R}}$ and its applications to probability. His results showed that there is a fundamental difference between the case that Λ is compact and non-compact. For this reason, we cannot apply his ideas to our case where $\Lambda \subseteq \mathbb{Z}^d$ is finite; nonetheless his ideas are potentially helpful in the case when $\Lambda \subseteq \mathbb{Z}^d$ is infinite.

For the case of a locally compact abelian group G instead of $\mathbb{T}^d$, and its dual group $\widehat{G}$ instead of $\mathbb{Z}^d$, Hewitt-Rubin proved that Γ is either empty, a subgroup of $\widehat{G}$, or a translate of a subgroup of $\widehat{G}$ [HR55, Theorem 2.5, page 99].

2.2. **Discrete super-resolution on $\mathbb{T}$ and compressed sensing.** The super-resolution problem is naturally connected to deterministic compressed sensing. We shall explain this connection in the context of our theory, but we are not the first to notice this connection, e.g., [DCG12, CFG14]. Consider the subspace,

$$M_N(\mathbb{T}) = \left\{ \mu \in M(\mathbb{T}) \colon \mu = \sum_{n=0}^{N-1} a_n \delta_{\frac{n}{N}}, \ a_n \in \mathbb{C} \right\} \subseteq M(\mathbb{T}),$$

and the following problem: For a given $\mu \in M_N(\mathbb{T})$ and a finite subset $\Lambda \subseteq \mathbb{Z}$, find

(SR′) $\underset{\nu \in M_N(\mathbb{T})}{\arg\min} \|\nu\|, \quad \text{subject to } \widehat{\mu} = \widehat{\nu} \text{ on } \Lambda.$

Since $M_N(\mathbb{T}) \subseteq M(\mathbb{T})$, we see that Problem (SR′) is a special case of Problem (SR). This implies that if $\nu \in M_N(\mathbb{T}^d)$ for some N is a solution to Problem (SR), then ν is automatically a solution to Problem (SR′).

Consider the compressed sensing problem for the *discrete Fourier transform*: For a given vector $x \in \mathbb{C}^N$ and finite set $\Lambda \subseteq \mathbb{Z}/N\mathbb{Z}$, find

(CS) $\underset{y \in \mathbb{C}^N}{\arg\min} \|y\|_1, \quad \text{subject to } \mathcal{F}_N x = \mathcal{F}_N y \text{ on } \Lambda,$

where $\|y\|_1 = \sum_{n=1}^N |y_n|$ and $\mathcal{F}_N \colon \mathbb{C}^N \to \mathbb{C}^N$ denotes the discrete Fourier transform. Notice that the Fourier transform of the measure $\mu = \sum_{n=0}^{N-1} a_n \delta_{n/N} \in M_N(\mathbb{T})$ equals the discrete Fourier transform of the vector $x \in \mathbb{C}^N$, where $x_n = a_n$ for $n = 0, 1, \ldots, N-1$. This observation implies that Problem (SR$'$) is equivalent to Problem (CS).

As a result of this connection, our theory has immediate applications to deterministic compressed sensing of vectors, and our examples with $\mu \in M_N(\mathbb{T})$ in Section 4 can be directly translated to the compressed sensing setting. This is surprising and interesting for the following reason. Loosely speaking, the celebrated result of Candès-Romberg-Tao [CRT06a] states that, with high probability, a sufficiently sparse $x \in \mathbb{C}^N$ can be recovered by solving Problem (CS). Notably, their result is probabilistic, not deterministic. The deterministic counterpart was studied by the same authors in [CRT06b] for matrices that satisfy a *Restricted Isometry Property* (RIP). However, the latter result does not apply to Problem (CS) because in general, a partial discrete Fourier transform matrix does not satisfy a RIP property, see [Sle78]. Thus, our theory provides RIP-less deterministic compressed sensing results.

In this paper, we do not fully explore the consequences of this connection in the context of our theory. For those interested in applications of super-resolution to deterministic compressed sensing, we refer the reader to [DCG12], in which the authors used this connection to construct deterministic sensing matrices for positive discrete measures.

3. Mathematical Theory

3.1. Preliminary functional analysis results.

The super-resolution problem is a minimization problem over the infinite dimensional space, $M(\mathbb{T}^d)$. A direct and naive approach is to find all extrapolations of μ from Λ, and then proceed to find the minimum extrapolations. The set of extrapolations is

$$X = X(\mu, \Lambda) = \{\nu \in M(\mathbb{T}^d) \colon \widehat{\nu} = \widehat{\mu} \text{ on } \Lambda\} = \mu + \{\nu \in M(\mathbb{T}^d) \colon \widehat{\nu} = 0 \text{ on } \Lambda\},$$

where the set $\{\nu \in M(\mathbb{T}^d) \colon \widehat{\nu} = 0 \text{ on } \Lambda\}$ is an ideal of the Banach algebra, $M(\mathbb{T}^d)$. However, characterizing such ideals is a difficult problem [Ben75]. Consequently, we forgo an algebraic approach in favor of a functional analysis approach.

Let $C(\mathbb{T}^d)$ be the space of complex-valued continuous functions on $\mathbb{T}^d$ equipped with the sup-norm $\|\cdot\|_\infty$. Then, $C(\mathbb{T}^d)$ is a Banach space, and let $C(\mathbb{T}^d)'$ be its dual space of continuous linear functionals with the usual norm $\|\cdot\|$. The celebrated Riesz representation theorem, e.g., see [BC09, Theorem 7.2.7, page 334] states that:

(a) For each $\mu \in M(\mathbb{T}^d)$, there exists a bounded linear functional $\ell_\mu \in C(\mathbb{T}^d)'$ such that $\|\mu\| = \|\ell_\mu\|$ and

$$\forall f \in C(\mathbb{T}^d), \quad \ell_\mu(f) = \langle f, \mu \rangle = \int_{\mathbb{T}^d} f(x) \, \overline{d\mu(x)}.$$

(b) For each bounded linear functional $\ell \in C(\mathbb{T}^d)'$, there exists a unique $\mu \in M(\mathbb{T}^d)$ such that $\|\mu\| = \|\ell\|$ and

$$\forall f \in C(\mathbb{T}^d), \quad \ell(f) = \langle f, \mu \rangle = \int_{\mathbb{T}^d} f(x) \, \overline{d\mu(x)}.$$

Proposition 3.1a shows that the super-resolution problem is well-posed, by using standard functional analysis. However, this type of argument does not yield useful statements about the minimal extrapolations. Instead of working with $C(\mathbb{T}^d)$, we shall work with a subspace. Since only $\widehat{\mu}\mid_\Lambda$ is important to the super-resolution problem, we consider the subspace

$$C(\mathbb{T}^d; \Lambda) = \left\{ f \in C(\mathbb{T}^d) \colon f(x) = \sum_{m \in \Lambda} a_m e^{2\pi i m \cdot x}, \ a_m \in \mathbb{C} \right\}.$$

Proposition 3.1b shows that $C(\mathbb{T}^d; \Lambda)$ is a closed subspace of $C(\mathbb{T}^d)$, and that implies $C(\mathbb{T}^d; \Lambda)$ is a Banach space. Further, Proposition 3.1c demonstrates that

$$U = U(\mathbb{T}^d; \Lambda) = \{f \in C(\mathbb{T}^d; \Lambda) \colon \|f\|_\infty \le 1\},$$

the closed unit ball of $C(\mathbb{T}^d; \Lambda)$, is compact.

The purpose of restricting to the subspace, $C(\mathbb{T}^d; \Lambda)$, is to identify $\mu \in M(\mathbb{T}^d)$ with the bounded linear functional, $L_\mu \in C(\mathbb{T}^d; \Lambda)'$, defined as

$$\forall f \in C(\mathbb{T}^d; \Lambda), \quad L_\mu(f) = \int_{\mathbb{T}^d} f(x)\, \overline{d\mu(x)}.$$

Although, by definition, $\|L_\mu\| = \sup_{f \in U} |L_\mu(f)|$, Proposition 3.1d,e show that we have the stronger statement,

$$\|L_\mu\| = \max_{f \in U} |L_\mu(f)| = \varepsilon.$$

Observe that for any extrapolation, $\nu \in M(\mathbb{T}^d)$,

$$\forall f \in C(\mathbb{T}^d; \Lambda), \quad L_\mu(f) = \langle f, \mu \rangle = \langle f, \nu \rangle = L_\nu(f).$$

Proposition 3.1. *Let $\mu \in M(\mathbb{T}^d)$ and let $\Lambda \subseteq \mathbb{Z}^d$ be a finite subset.*

(a) $\mathcal{E} \subseteq M(\mathbb{T}^d)$ is non-empty, weak-*compact, and convex.
(b) $C(\mathbb{T}^d; \Lambda)$ is a closed subspace of $C(\mathbb{T}^d)$.
(c) U is a compact subset of $C(\mathbb{T}^d; \Lambda)$.
(d) $\varepsilon = \|L_\mu\|$.
(e) $\varepsilon = \max_{f \in U} |L_\mu(f)|$.
(f) There exists $\varphi \in U$ such that $L_\mu(\varphi) = \varepsilon$.

Proof.

(a) By definition of ε, there exists a sequence $\{\nu_j\} \subseteq \{\nu \colon \widehat{\nu} = \widehat{\mu} \text{ on } \Lambda\}$ such that $\|\nu_j\| \to \varepsilon$. Then, this sequence is bounded. By Banach-Alaoglu, after passing to a subsequence, we can assume there exists $\nu \in M(\mathbb{T}^d)$ such that $\nu_j \to \nu$ in the weak-$*$ topology.

Let V be the closed unit ball of $C(\mathbb{T}^d)$. We have $\|\nu\| \le \varepsilon$ because

$$\|\nu\| = \sup_{f \in V} |\langle f, \nu \rangle| = \sup_{f \in V} \lim_{j \to \infty} |\langle f, \nu_j \rangle| \le \sup_{f \in V} \lim_{j \to \infty} \|f\|_\infty \|\nu_j\| \le \varepsilon.$$

Moreover, for $m \in \Lambda$, we have

$$\widehat{\mu}(m) = \lim_{j \to \infty} \widehat{\nu_j}(m) = \lim_{j \to \infty} \int_{\mathbb{T}^d} e^{-2\pi i m \cdot x}\, d\nu_j(x) = \int_{\mathbb{T}^d} e^{-2\pi i m \cdot x}\, d\nu(x) = \widehat{\nu}(m).$$

This shows that ν is an extrapolation, and thus, $\|\nu\| \ge \varepsilon$. Therefore, $\nu \in \mathcal{E}$.

The proof that $\mathcal{E}$ is weak-$*$ compact is similar. Pick any sequence $\{\nu_j\} \subseteq \mathcal{E}$ and after passing to a subsequence, we can assume $\nu_j \to \nu$ in the weak-$*$ topology for some $\nu \in M(\mathbb{T}^d)$. By the same argument, we see that $\nu \in \mathcal{E}$.

If $\mathcal{E}$ contains exactly one measure, then $\mathcal{E}$ is trivially convex. Otherwise, let $t \in [0, 1]$, $\nu_0, \nu_1 \in \mathcal{E}$, and $\nu_t = (1 - t)\nu_0 + t\nu_1$. Then, ν_t is an extrapolation and thus, $\|\nu_t\| \ge \varepsilon$. By the triangle inequality, we have $\|\nu_t\| \le (1 - t)\|\nu_0\| + t\|\nu_1\| = \varepsilon$. Thus, $\nu_t \in \mathcal{E}$ for each $t \in [0, 1]$.

(b) Suppose $\{f_j\} \subseteq C(\mathbb{T}^d)$ and that there exists $f \in C(\mathbb{T}^d)$ such that $f_j \to f$ uniformly. Then, $\widehat{f_j}(m) \to \widehat{f}(m)$ for all $m \in \mathbb{Z}^d$. Since $\widehat{f_j}(m) = 0$ if $m \notin \Lambda$, we deduce that $\widehat{f}(m) = 0$ if $m \notin \Lambda$. This shows that $f \in C(\mathbb{T}^d; \Lambda)$ and thus, $C(\mathbb{T}^d; \Lambda)$ is a closed subspace of $C(\mathbb{T}^d)$.

(c) Let $\{f_j\} \subseteq U$. We first show that $\{f_j\}$ is a uniformly bounded equicontinuous family. By definition, $\|f_j\|_\infty \le 1$, and there exist $a_{j,m} \in \mathbb{C}$ such that $f_j(x) = \sum_{m \in \Lambda} a_{j,m} e^{2\pi i m \cdot x}$. Note that

$|a_{j,m}| \le \|f_j\|_\infty \le 1$. Then, for any $x, y \in \mathbb{T}^d$, we have

$$|f_j(x) - f_j(y)| = \Big| \sum_{m \in \Lambda} a_{j,m}(e^{2\pi i m \cdot x} - e^{2\pi i m \cdot y}) \Big| \le \sum_{m \in \Lambda} |e^{2\pi i m \cdot x} - e^{2\pi i m \cdot y}|.$$

Let $\varepsilon > 0$ and $m \in \Lambda$. There exists $\delta_m > 0$ such that $|e^{2\pi i m \cdot x} - e^{2\pi i m \cdot y}| < \varepsilon$ whenever $|x - y| < \delta_m$. Let $\delta = \min_{m \in \Lambda} \delta_m$. Combining this with the previous inequalities, we have

$$|f_j(x) - f_j(y)| \le \sum_{m \in \Lambda} |e^{2\pi i m \cdot x} - e^{2\pi i m \cdot y}| \le \#\Lambda \varepsilon,$$

whenever $|x - y| < \delta$. This shows that $\{f_j\}$ is a uniformly bounded equicontinuous family.

By the Arzelà-Ascoli theorem, there exists $f \in C(\mathbb{T}^d)$, with $\|f\|_\infty \le 1$, and a subsequence $\{f_{j_k}\}$ such that $f_{j_k} \to f$ uniformly. Thus, $\widehat{f_{j_k}}(m) \to \widehat{f}(m)$ for all $m \in \mathbb{Z}^d$, which shows that $f \in C(\mathbb{T}^d; \Lambda)$.

(d) Let $\nu \in \mathcal{E}$. Then,

$$\forall f \in U, \quad |L_\mu(f)| = |\langle f, \mu \rangle| = |\langle f, \nu \rangle| \le \|f\|_\infty \|\nu\| \le \varepsilon,$$

which proves the upper bound, $\|L_\mu\| \le \varepsilon$.

For the lower bound, we use the Hahn-Banach theorem to extend $L_\mu \in C(\mathbb{T}^d; \Lambda)'$ to $\ell \in C(\mathbb{T}^d)'$, where $\|L_\mu\| = \|\ell\|$. By the Riesz representation theorem, there exists a unique $\sigma \in M(\mathbb{T}^d)$ such $\ell(f) = \langle f, \sigma \rangle$ for all $f \in C(\mathbb{T}^d)$ and $\|\sigma\| = \|\ell\|$. In particular,

$$\forall f \in C(\mathbb{T}^d; \Lambda), \quad \langle f, \sigma \rangle = \ell(f) = L_\mu(f) = \langle f, \mu \rangle.$$

Set $f(x) = e^{-2\pi i m \cdot x}$, where $m \in \Lambda$, to deduce that $\widehat{\mu} = \widehat{\sigma}$ on Λ. This implies $\|\sigma\| \ge \varepsilon$. Combining these facts, we have

$$\varepsilon \le \|\sigma\| = \|\ell\| = \|L_\mu\|.$$

This proves the lower bound.

(e) We know that $\|L_\mu\| = \varepsilon$. By definition, there exists $\{f_j\} \subseteq U$ such that $|L_\mu(f_j)| \ge \varepsilon - 1/j$. By compactness of U, there exists a subsequence $\{f_{j_k}\} \subseteq U$ and $f \in U$ such that $f_{j_k} \to f$ uniformly. We immediately have $|L_\mu(f)| \le \|f\|_\infty \varepsilon \le \varepsilon$. For the reverse inequality, as $k \to \infty$,

$$|L_\mu(f)| \ge |L_\mu(f_{j_k})| - |L_\mu(f - f_{j_k})| \ge \varepsilon - \frac{1}{j_k} - \|\mu\| \|f - f_{j_k}\|_\infty \to \varepsilon.$$

This proves that $|L_\mu(f)| = \varepsilon$.

(f) There exists $f \in U$ such that $|L_\mu(f)| = \varepsilon$. This implies $e^{i\theta} L_\mu(f) = \varepsilon$ for some $\theta \in \mathbb{R}$. Hence, let $\varphi = e^{i\theta} f \in U$, and thus, $L_\mu(\varphi) = \varepsilon$.

$\square$

The next step is to deduce information about the minimal extrapolations from the previous proposition. As a consequence of the Radon-Nikodym theorem, for each $\mu \in M(\mathbb{T}^d)$, there exists a μ-measurable function $\mathrm{sign}(\mu)$ such that $|\mathrm{sign}(\mu)| = 1$ μ-a.e. and satisfies the identity

$$\forall f \in L^1_{|\mu|}(\mathbb{T}^d), \quad \int_{\mathbb{T}^d} f \, d|\mu| = \int_{\mathbb{T}^d} f \, \overline{\mathrm{sign}(\mu)} \, d\mu.$$

See [BC09, Theorem 5.3.2, page 242, and Theorem 5.3.5, page 244] for further details. The support of $\mu \in M(\mathbb{T}^d)$, denoted $\mathrm{supp}(\mu)$, is the complement of all open sets $A \subseteq \mathbb{T}^d$ such that $\mu(A) = 0$.

By Proposition 3.1f, there exists $\varphi \in U$ such that $L_\mu(\varphi) = \varepsilon$. Proposition 3.2a shows that φ interpolates $\mathrm{sign}(\nu)$ ν-a.e. for all $\nu \in \mathcal{E}$. In theory, φ is not the unique function belonging to U that interpolates $\mathrm{sign}(\nu)$, and it would be desirable to pick such a φ that has the best properties, or use every φ available. We are unable to do this, but Proposition 3.2b, which is a special case of Proposition 3.2a, shows that it is possible to characterize certain types of φ by the set Γ. This explains why Γ is so important in our analysis.

Proposition 3.2. *Let $\mu \in M(\mathbb{T}^d)$ and let $\Lambda \subseteq \mathbb{Z}^d$ be a finite subset.*

(a) There exists $\varphi \in U$ such that,

$$\forall \nu \in \mathcal{E}, \quad \varphi = sign(\nu) \quad \nu\text{-a.e.}, \quad and \quad supp(\nu) \subseteq \{x \in \mathbb{T}^d \colon |\varphi(x)| = 1\}.$$

(b) For each $m \in \mathbb{Z}^d$, define $\alpha_m \in \mathbb{R}/\mathbb{Z}$ by the formula $e^{-2\pi i \alpha_m} \widehat{\mu}(m) = |\widehat{\mu}(m)|$. Then, $m \in \Gamma$ if and only if

$$\forall \nu \in \mathcal{E}, \quad sign(\nu)(x) = e^{2\pi i \alpha_m} e^{2\pi i m \cdot x} \quad \nu\text{-a.e.}$$

Proof.

(a) By Proposition 3.1f, there exists $\varphi \in U$ such that $L_\mu(\varphi) = \varepsilon$. Let $\nu \in \mathcal{E}$. Since $\varphi \in U \subseteq C(\mathbb{T}^d; \Lambda)$, we have

$$\langle \varphi, \nu \rangle = \langle \varphi, \mu \rangle = L_\mu(\varphi) = \varepsilon = \|\nu\|.$$

Since $\|\varphi\|_\infty \leq 1$ and $\langle \varphi, \nu \rangle = \|\nu\|$, we must have $\varphi = \text{sign}(\nu)$ ν-a.e.

Since $|\varphi| = |\text{sign}(\nu)| = 1$ ν-a.e. and ν is a Radon measure,

$$\text{supp}(\nu) \subseteq \overline{\{x \in \mathbb{T}^d \colon |\varphi(x)| = 1\}} = \{x \in \mathbb{T}^d \colon |\varphi(x)| = 1\}.$$

The last equality holds because the inverse image of the closed set $\{1\}$ under the continuous function $|\varphi|$ is closed.

(b) Suppose $m \in \Gamma$ and $\nu \in \mathcal{E}$. Then

$$\int_{\mathbb{T}^d} e^{-2\pi i \alpha_m} e^{-2\pi i m \cdot x} \, d\nu(x) = e^{-2\pi i \alpha_m} \widehat{\nu}(m) = e^{-2\pi i \alpha_m} \widehat{\mu}(m) = |\widehat{\mu}(m)| = \varepsilon = \|\nu\|.$$

This shows $\text{sign}(\nu)(x) = e^{2\pi i \alpha_m} e^{2\pi i m \cdot x}$ ν-a.e. The converse follows by reversing the steps.

$\square$

3.2. **An analogue of Beurling's theorem.** We are ready to prove the main theorem. By Proposition 3.2a, there exists a $\varphi \in U$ such that the minimal extrapolations are supported in the closed set $S = \{x \in \mathbb{T}^d \colon |\varphi(x)| = 1\}$. By itself, this statement is not useful because if $|\varphi| \equiv 1$, then $S = \mathbb{T}^d$. However, Proposition 3.2b characterizes the case that $|\varphi| \equiv 1$ in terms of the set Γ. Intuitively, $\#\Gamma$ represents the number of "bad" functions in U that interpolate the signs of the minimal extrapolations. While it is desirable to have $\Gamma = \emptyset$, perhaps surprisingly, we can make strong statements when $\#\Gamma$ is large.

Theorem 3.3. *Let $\mu \in M(\mathbb{T}^d)$ and let $\Lambda \subseteq \mathbb{Z}^d$ be a finite subset.*

(a) Suppose $\#\Gamma = 0$. Then, there exists a closed set S of d-dimensional Lebesgue measure zero that contains the support of each minimal extrapolation. If $d = 1$, then S is a finite number of points.

(b) Suppose $\#\Gamma \geq 2$. For each distinct pair $m, n \in \Gamma$, define $\alpha_{m,n} \in \mathbb{R}/\mathbb{Z}$ by $e^{2\pi i \alpha_{m,n}} = \widehat{\mu}(m)/\widehat{\mu}(n)$. Then, every minimal extrapolation is supported in the set,

$$S = \bigcap_{\substack{m,n \in \Gamma \\ m \neq n}} \{x \in \mathbb{T}^d \colon x \cdot (m - n) + \alpha_{m,n} \in \mathbb{Z}\}.$$

Further, if $d \geq 2$ and there exist d linearly independent vectors, $p_1, p_2, \ldots, p_d \in \mathbb{Z}^d$, such that

$$\{p_1, p_2, \ldots, p_d\} \subseteq \{m - n \colon m, n \in \Gamma\},$$

then S is a lattice on $\mathbb{T}^d$.

Proof.

(a) By Proposition 3.2a, there exists $\varphi \in U$ such that each minimal extrapolation is supported in the closed set

$$S = \{x \in \mathbb{T}^d \colon |\varphi(x)| = 1\}.$$

Note that $|\varphi| \not\equiv 1$. In fact, if $|\varphi| \equiv 1$, since $\varphi \in U$, we must have $\varphi(x) = e^{2\pi i m \cdot x}$ for some $m \in \Lambda$. Then, by Proposition 3.2b, this implies $m \in \Gamma$, which contradicts the assumption that $\Gamma = \emptyset$.

We have $\varphi(x) = \sum_{m \in \Lambda} a_m e^{2\pi i m \cdot x}$ for some $a_m \in \mathbb{C}$. Consider the function

(3.1) $$\Phi(x) = 1 - |\varphi(x)|^2 = 1 - \sum_{m \in \Lambda} \sum_{n \in \Lambda} a_m \overline{a_n} e^{2\pi i (m-n) \cdot x}.$$

Then, the minimal extrapolations are supported in the closed set

$$S = \{x \in \mathbb{T}^d \colon \Phi(x) = 0\}.$$

Note that $\Phi \not\equiv 0$ because $|\varphi| \not\equiv 1$. Since Φ is a non-trivial real-analytic function, S is a set of d-dimensional Lebesgue measure zero. In particular, if $d = 1$, then S is a finite set of points.

(b) Let $m, n \in \Gamma$. There exist $\alpha_m, \alpha_n \in \mathbb{R}/\mathbb{Z}$ defined in Proposition 3.2b, such that

$$\forall \nu \in \mathcal{E}, \quad \text{sign}(\nu)(x) = e^{2\pi i \alpha_m} e^{2\pi i m \cdot x} = e^{2\pi i \alpha_n} e^{2\pi i n \cdot x} \quad \nu\text{-a.e.}$$

Set $\alpha_{m,n} = \alpha_m - \alpha_n \in \mathbb{R}/\mathbb{Z}$. Then each minimal extrapolation is supported in

$$S_{m,n} = \{x \in \mathbb{T}^d \colon x \cdot (m-n) + \alpha_{m,n} \in \mathbb{Z}\}.$$

Thus, *each minimal extrapolation is supported in the set*

$$S = \bigcap_{\substack{m,n \in \Gamma \\ m \neq n}} S_{m,n} = \bigcap_{\substack{m,n \in \Gamma \\ m \neq n}} \{x \in \mathbb{T}^d \colon x \cdot (m-n) + \alpha_{m,n} \in \mathbb{Z}\}.$$

Note that $e^{2\pi i \alpha_{m,n}} = \widehat{\mu}(m)/\widehat{\mu}(n)$ because

$$e^{-2\pi i \alpha_m} \widehat{\mu}(m) = |\widehat{\mu}(m)| = \varepsilon = |\widehat{\mu}(n)| = e^{-2\pi i \alpha_n} \widehat{\mu}(n).$$

Suppose $\{p_1, \ldots, p_d\}$ satisfies the hypothesis. By the support assertion that we just proved, there exists $\beta = (\beta_1, \beta_2, \ldots, \beta_d) \in \mathbb{T}^d$ such that every minimal extrapolation is supported in

$$S = \bigcap_{j=1}^{d} \{x \in \mathbb{T}^d \colon x \cdot p_j + \beta_j \in \mathbb{Z}\}.$$

Let us explain the geometry of the situation before we proceed with the proof that S is a lattice. Note that $\{x \in \mathbb{T}^d \colon x \cdot p_j + \beta_j \in \mathbb{Z}\}$ is a family of parallel and periodically spaced hyperplanes. Since the vectors, $p_1, p_2, \ldots, p_d$, are assumed to be linearly independent, one family of hyperplanes is not parallel to any other family of hyperplanes. Hence, the intersection of d non-parallel and periodically spaced hyperplanes is a lattice, see Figure 3.1 for an illustration.

For the rigorous proof of this observation, first note that

$$S = \{x \in \mathbb{T}^d \colon Px + \beta \in \mathbb{Z}^d\},$$

where $P = (p_1, p_2, \ldots, p_d)^t \in \mathbb{Z}^{d \times d}$ is invertible because its rows are linearly independent. Let $x_0 \in \mathbb{R}^d$ be the solution to $Px + \beta = 0$, and let $q_j \in \mathbb{Q}^d$ be the solution to $Px = e_j$, where e_j is the standard basis vector for $\mathbb{R}^d$. Then $p_j \cdot q_k = \delta_{j,k}$, and S is generated by the point x_0 and the lattice vectors, $q_1, q_2, \ldots, q_d$.
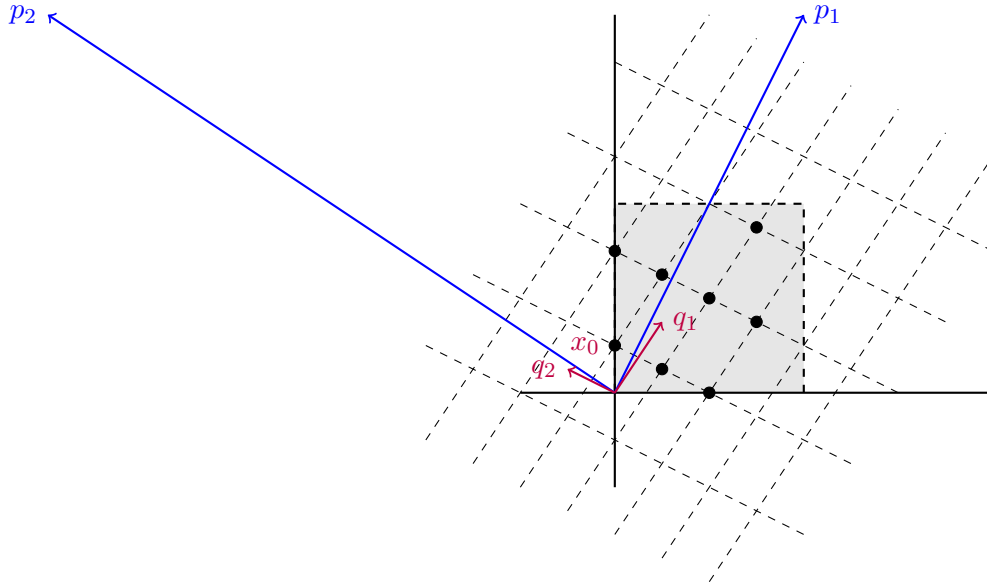
$\square$

FIGURE 3.1. An illustration of Theorem 3.3b, where $d = 2$, $p_1 = (1, 2)$, $p_2 = (-3, 2)$, $\beta_1 = 1/2$, $\beta_2 = -1/2$, $q_1 = (1/4, 3/8)$, and $q_2 = (-1/4, 1/8)$. The family of hyperplanes are the dashed lines, the lattice S is the black dots, and the shaded region is $[0, 1)^2$.

Remark 3.4. The statement, "the minimal extrapolations are supported in a discrete set (respectively, set of measure zero)," is stronger than the statement, "the minimal extrapolations are discrete measures (respectively, singular measures)," because the former statement guarantees that the minimal extrapolations are supported in the *same* discrete set (respectively, set of measure zero). Having a common support set for the minimal extrapolations is important in determining uniqueness, see Proposition 3.8.

Remark 3.5. By Theorem 3.3b, when $d = 1$ and $\#\Gamma \geq 2$, the minimal extrapolations are discrete measures. However, Example 4.7 shows that for higher dimensions, the minimal extrapolations are not necessarily discrete. This demonstrates that the conclusion of Theorem 3.3b for $d \geq 2$ is optimal.

Remark 3.6. Let $\{p_1, p_2, \ldots, p_d\}$ and $\{q_1, q_2, \ldots, q_d\}$ be the vectors defined in the proof of Theorem 3.3b. Since $p_j \cdot q_j = 1$ for $j = 1, 2 \ldots, d$, then, by Cauchy-Schwarz, we have $\|q_j\|_2 \geq \|p_j\|_2^{-1}$. Since the set $\{q_1, q_2, \ldots, q_d\}$ generates the lattice S and $\|q_j\|_2 \geq \|p_j\|_2^{-1}$, we see that $\#S$ is non-decreasing as a function of $\|p_j\|_2$, when the set $\{p_k \colon k \neq j\}$ is fixed. Thus, it is best to choose the set $\{p_1, p_2, \ldots, p_d\}$ such that it satisfies the hypothesis of Theorem 3.3b and each $\|p_j\|_2$ is small. This observation is consistent with Beurling's observation in Theorem 2.1b; he defined τ to be the smallest distance between any two points in Λ_m, which is optimal.

Remark 3.7. Equivalently, in the Fourier domain, problem (SR) can be interpreted as an extension problem: Of all possible extensions of $\widehat{\mu}\,|_\Lambda$ to $\mathbb{Z}^d$ such that each extension corresponds to the Fourier transform of a bounded measure, find the extension whose corresponding bounded measure has the smallest total variation. When interpreted this way, Theorem 3.3 for $d = 1$ says that if $\#\Gamma \neq 1$, then the Fourier transforms of the minimal extrapolations are of the form $\sum_{k=1}^{K} a_k e^{2\pi i m \cdot x_k}$, so the optimal extensions of $\widehat{\mu}\,|_\Lambda$ to $\mathbb{Z}^d$ are *almost periodic*, e.g., see [Rud62, Ben75].

3.3. **On uniqueness and non-uniqueness of minimal extrapolations.** Theorem 3.3 provides sufficient conditions for the minimal extrapolations to be supported in a discrete set, but it does

not provide sufficient conditions for uniqueness. Since a family of discrete measures supported on a common set behaves essentially like a vector, we use basic linear algebra to address the question of uniqueness when the minimal extrapolations are supported in a discrete set. Note that matrices of the form (3.2) have applications to the compressed sensing of positive, sparse, discrete measures, e.g., see [DCG12].

Proposition 3.8. *Let $\mu \in M(\mathbb{T}^d)$ and let $\Lambda = \{m_1, m_2, \ldots, m_J\} \subseteq \mathbb{Z}^d$ be a finite subset. Suppose there exists a finite set, $\{x_k \in \mathbb{T}^d \colon k = 1, 2, \ldots, K\}$, such that each minimal extrapolation is supported in this set. Suppose that the matrix,*

$$(3.2) \qquad E(m_1, \ldots, m_J; x_1, \ldots, x_K) = \begin{pmatrix} e^{-2\pi i m_1 \cdot x_1} & e^{-2\pi i m_1 \cdot x_2} & \cdots & e^{-2\pi i m_1 \cdot x_K} \\ e^{-2\pi i m_2 \cdot x_1} & e^{-2\pi i m_2 \cdot x_2} & \cdots & e^{-2\pi i m_2 \cdot x_K} \\ \vdots & \vdots & & \vdots \\ e^{-2\pi i m_J \cdot x_1} & e^{-2\pi i m_J \cdot x_2} & \cdots & e^{-2\pi i m_J \cdot x_K} \end{pmatrix},$$

has full column rank (this can only occur if $J \geq K$). Then, the minimal extrapolation is unique.

Proof. Let ν be the difference of any two minimal extrapolations. Then ν is also supported in $\{x_k \colon k = 1, 2, \ldots, K\}$ and it is of the form $\nu = \sum_{k=1}^{K} a_k \delta_{x_k}$. Since $\widehat{\nu} = 0$ on Λ, we have $0 = \widehat{\nu}(m_j) = \sum_{k=1}^{K} a_k e^{-2\pi i m_j \cdot x_k}$ for $j = 1, 2, \ldots, J$, which is equivalent to the linear system $Ea = 0$, where $a = (a_1, \ldots, a_K) \in \mathbb{C}^K$. By assumption, E has full column rank. This implies $a = 0$. $\qquad \square$

To finish this subsection, we address the situation when $\#\Gamma = 1$, i.e., the missing case of Theorem 3.3. We say a measure $\mu \in M(\mathbb{T}^d)$ is a *positive measure* if $\mu(A) \geq 0$ for all Borel sets $A \subseteq \mathbb{T}^d$. A sequence $\{a_m \in \mathbb{C} \colon m \in \mathbb{Z}^d\}$ is a *positive-definite sequence* if for all sequences $\{b_m \in \mathbb{C} \colon m \in \mathbb{Z}^d\}$ of finite support, we have

$$\sum_{m,n \in \mathbb{Z}^d} a_{m-n} b_m \overline{b_n} \geq 0.$$

For $y \in \mathbb{T}^d$, let $T_y \colon M(\mathbb{T}^d) \to M(\mathbb{T}^d)$ be the translation operator defined by $T_y \mu(x) = \mu(x - y)$. For $n \in \mathbb{Z}^d$, let $M_n \colon M(\mathbb{T}^d) \to M(\mathbb{T}^d)$ be the modulation operator defined by $M_n \mu(x) = e^{2\pi i n \cdot x} \mu(x)$.

Proposition 3.9. *Let $\mu \in M(\mathbb{T}^d)$ and let $\Lambda \subseteq \mathbb{Z}^d$ be a finite subset. Suppose there exists $n \in \Lambda$ such that $(M_{-n}\mu)^\wedge \mid_{\Lambda - n}$ extends to a positive-definite sequence on $\mathbb{Z}^d$. Then, $n \in \Gamma$, and each positive-definite extension of $(M_{-n}\mu)^\wedge \mid_{\Lambda - n}$ corresponds to a positive measure ν, such that $M_n \nu$ is a minimal extrapolation of μ from Λ.*

Proof. Extend $(M_{-n}\mu)^\wedge \mid_{\Lambda - n}$ to a positive-definite sequence. By Herglotz' theorem, there exists a positive measure $\nu \in M(\mathbb{T}^d)$ such that $\widehat{\nu} = (M_{-n}\mu)^\wedge = T_{-n}\mu$ on $\Lambda - n$. Then, $M_n \nu$ is an extrapolation of μ from Λ, which shows that

$$\|\nu\| = \|M_n \nu\| \geq \varepsilon(\mu, \Lambda).$$

For the reverse inequality, since ν is a positive measure, we have $\|\nu\| = \widehat{\nu}(0)$. Then,

$$\|M_n \nu\| = \|\nu\| = \widehat{\nu}(0) = |\widehat{\nu}(0)| = |\widehat{\mu}(n)| \leq \|\widehat{\mu}\|_{\ell^\infty(\Lambda)} \leq \varepsilon(\mu, \Lambda),$$

where the last inequality follows by Proposition 3.15. This shows that $|\widehat{\mu}(n)| = \|\nu\| = \varepsilon$, which proves that $n \in \Gamma$ and $M_n \nu$ is a minimal extrapolation of μ from Λ. $\qquad \square$

Corollary 3.10. *Let $\mu \in M(\mathbb{T}^d)$ and let $\Lambda \subseteq \mathbb{Z}^d$ be a finite subset. Suppose $0 \in \Lambda$ and $\widehat{\mu} \mid_\Lambda$ extends to a positive-definite sequence on $\mathbb{Z}^d$. Then, $0 \in \Gamma$, and each positive-definite extension of $\widehat{\mu} \mid_\Lambda$ corresponds to a positive minimal extrapolation.*

Remark 3.11. Beurling [Beu89b] and Esseen [Ess45] essentially proved the analogue of Corollary 3.10 for $\mathbb{R}$ instead of $\mathbb{T}^d$. Proposition 3.9 generalizes their result to handle situations when $0 \notin \Lambda$. This is important because from the viewpoint of Proposition 3.19c, the super-resolution problem is

invariant under simultaneous translations of $\widehat{\mu}$ and Λ, which means that $0 \in \mathbb{Z}^d$ is no more special than any other point $n \in \mathbb{Z}^d$.

Remark 3.12. Proposition 3.9 suggests that the case $\#\Gamma = 1$ is special compared to the cases $\#\Gamma = 0$ or $\#\Gamma \geq 2$ because, when $\#\Gamma = 1$, there may exist absolutely continuous minimal extrapolations. In Example 4.2, $\#\Gamma = 1$, and there exist uncountably many discrete and positive absolutely continuous minimal extrapolations.

Remark 3.13. Suppose that $\mu \in M(\mathbb{T})$ and $\widehat{\mu}\mid_\Lambda$ can be extended to a positive-definite sequence on $\mathbb{Z}$. In theory, there are an infinite number of such extensions, and one particular method of choosing such an extension is called the *Maximum Entropy Method* (MEM). According to MEM, one extends $\widehat{\mu}\mid_\Lambda$ to the positive-definite sequence $\{a_m \in \mathbb{C} \colon m \in \mathbb{Z}\}$ whose corresponding density function $f \in L^1(\mathbb{T})$ is the unique maximizer of a specific logarithmic integral associated with the physical notion of entropy, e.g., see [Ben96, Theorems 3.6.3 and 3.6.6]. MEM is related to spectral estimation methods [Chi78], the maximum likelihood method [Chi78], and moment problems [Lan87].

3.4. **The admissibility range for ε.** While the mathematical theory we have developed connects the set Γ with the support of the minimal extrapolations, the difficulty of applying the theory is that finding ε can be difficult. One approach to computing ε is to use Proposition 3.1e. It states that $\varepsilon = \max_{f \in U} |\langle f, \mu \rangle|$. By definition, $f \in U$ if and only if $f(x) = \sum_{m \in \Lambda} a_m e^{2\pi i m \cdot x}$ for some $a_m \in \mathbb{C}$ and $\|f\|_\infty \leq 1$. Also, $f \in U$ if and only if $e^{i\theta} f \in U$ for all $\theta \in \mathbb{R}$. Applying Parseval's theorem, we have

$$(3.3) \qquad \varepsilon = \max \left\{ \mathrm{Re}\Big(\sum_{m \in \Lambda} a_m \overline{\widehat{\mu}(m)} \Big) \colon a_m \in \mathbb{C}, \ \sup_{x \in \mathbb{T}^d} \Big| \sum_{m \in \Lambda} a_m e^{2\pi i m \cdot x} \Big| \leq 1 \right\}.$$

Remark 3.14. By using results of Dumitrescu [Dum07, Theorem 4.24, Corollary 4.25, Remark 4.28], Candès and Fernandez-Granda [CFG14, Corollary 4.1] showed that if $\Lambda = \{-\lambda, -\lambda+1, \ldots, \lambda\}$ and $\mu \in M(\mathbb{T})$ satisfies a minimal extrapolation condition, then (3.3) can be written as a semi-definite program. Thus, under their hypotheses, both ε and a function $f \in U$ that attains the maximum can be computed using standard convex optimization software.

They provided a partial algorithm that recovers the unique minimal extrapolation. Their algorithm does not work when their function $p_{2n-2} \colon \mathbb{T} \to \mathbb{R}$, see [CFG14, (4.4), page 937], identically vanishes. Note that their p_{2n-2} is equivalent to our Φ, when $d = 1$, see (3.1). Recall from the proof of Theorem 3.3 that $\Gamma = \emptyset$ implies $\Phi \not\equiv 0$, but when $\Gamma \neq \emptyset$, it is possible that $\Phi \equiv 0$. To summarize, their algorithm is only guaranteed to output the unique minimal extrapolation when $\Gamma = \emptyset$, which is not guaranteed, even if μ satisfies a minimum separation condition.

On other hand, Theorem 3.3b is able to handle the case when $\#\Gamma \geq 2$, and for arbitrary $\mu \in M(\mathbb{T}^d)$. The conclusions of Theorem 3.3b can be directly translated into an algorithm, that not only handles general measures and for $d \geq 2$, but also improves upon the algorithm that Candès and Fernandez-Granda proposed for $\mu \in M(\mathbb{T})$ satisfying a minimum separation condition. The basic idea for improving their algorithm is that after obtaining ε, compute Γ, which can be done because Λ is a finite set and $\widehat{\mu}\mid_\Lambda$ is given. If $\Gamma = \emptyset$, then follow their algorithm, which is guaranteed to work in this case. If $\#\Gamma \geq 2$, then determine the support set of the minimal extrapolations by solving the system of linear equations as shown in the proof of Theorem 3.3. If $\#\Gamma = 1$, then perhaps, the best approach is to use a MEM-type method, see Remark 3.13. Of course, our discussion in this paragraph does not deal with round-off error and noise, and we have not provided the details of our proposed algorithm, see [BL16b].

While the optimization formulation of ε in (3.3) can sometimes be solved numerically, this is not necessary for theoretical purposes. Instead, we can always obtain useful lower and upper bounds for ε, and, under special circumstances, we can determine ε exactly. We say $[A, B] \subseteq \mathbb{R}^+$ is an *admissibility range* for ε provided that $0 \leq A \leq \varepsilon \leq B$. Proposition 3.15 shows that we always

have $A = \|\widehat{\mu}\|_{\ell^\infty}$ and $B = \|\mu\|$, and B can be improved in certain situations. We call $\left[\|\widehat{\mu}\|_{\ell^\infty}, \|\mu\|\right]$ the *naive admissibility range* for ε. See Figure 3.2 for an illustration of the admissibility range for ε and Proposition 3.15.

Proposition 3.15. *Let $\mu \in M(\mathbb{T}^d)$ and let $\Lambda \subseteq \mathbb{Z}^d$ be a finite subset. We have the lower and upper bounds,*

$$(3.4) \qquad \qquad \|\widehat{\mu}\|_{\ell^\infty(\Lambda)} \le \varepsilon \le \|\mu\|.$$

Further, if there exists an extrapolation ν such that $\|\nu\| < \|\mu\|$, then

$$(3.5) \qquad \qquad \|\widehat{\mu}\|_{\ell^\infty(\Lambda)} \le \varepsilon \le \|\nu\| < \|\mu\|.$$

Proof. To see the lower bound for ε in (3.4) and (3.5), let σ be a minimal extrapolation. Then,

$$\sup_{m \in \Lambda} |\widehat{\mu}(m)| = \sup_{m \in \Lambda} |\widehat{\sigma}(m)| \le \|\sigma\| = \varepsilon.$$

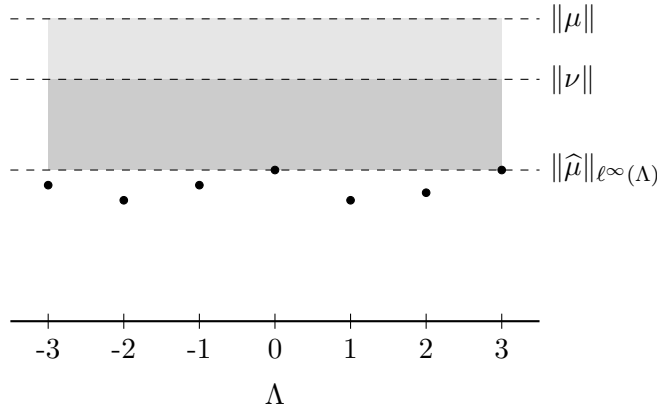The upper bounds, $\varepsilon \le \|\mu\|$ and $\varepsilon \le \|\nu\|$ in (3.4) and (3.5), follow by definition of ε. $\square$

FIGURE 3.2. The black points represent the values of $|\widehat{\mu}|$ on Λ. The union of the light and dark gray regions is the naive admissibility range for ε, whereas the dark gray region is an improved admissibility range for ε.

Remark 3.16. By tightening the admissibility range for ε, we can deduce information about the minimal extrapolations. The simplest case is when $\varepsilon = \|\widehat{\mu}\|_{\ell^\infty(\Lambda)} = \|\mu\|$, see Examples 4.2 and 4.3. The next simplest case is when there exists an extrapolation ν, such that $\varepsilon = \|\widehat{\mu}\|_{\ell^\infty(\Lambda)} = \|\nu\|$, see Example 4.4. A more complicated case is when $\|\widehat{\mu}\|_{\ell^\infty(\Lambda)} < \varepsilon$, see Example 4.5.

Remark 3.17. Candès and Fernandez-Granda [CFG14] focused entirely on the case that $\|\mu\| = \varepsilon$ because this is a necessary condition to super-resolve μ. In contrast to their result, our theory is better suited for situations when $\|\widehat{\mu}\|_{\ell^\infty} = \varepsilon$. This is because Theorem 3.3b is strongest when $|\widehat{\mu}(m)| = \varepsilon$ for many points $m \in \Lambda$, i.e., when $\#\Gamma$ is large. If $\|\mu\| = \|\widehat{\mu}\|_{\ell^\infty} = \varepsilon$, then our theorem has applications in the direction of deducing the *possibility* of super-resolution. Otherwise, our theorem has applications in the direction of deducing the *impossibility* of super-resolution reconstruction. In Section 4, we shall see applications of our theory to both directions.

Remark 3.18. Uniqueness of the minimal extrapolation of a discrete measure satisfying the minimum separation condition (2.4) is a consequence of the existence of a *dual certificate*, which is closely related to the optimization problem (3.3). The dual certificate for vectors was introduced in [CRT06a], then extended to the class of discrete measures [CFG14, TBSR13], and further generalized to non-Fourier transform measurements [DCG12]. To our knowledge, sufficient conditions for uniqueness of non-discrete measures given partial frequency information is not well-understood.

3.5. **Basic properties of minimal extrapolation.** Our next goal is to examine the symmetries of the minimal extrapolations. We are interested in the vector space operations, namely, addition of measures and the multiplication of measures by complex constants. We are also interested in the operations that are well-behaved under the Fourier transform on $\mathbb{T}^d$, namely, translation on the torus, modulation by integers, convolution of measures, and product of measures.

Proposition 3.19. *Let $\mu \in M(\mathbb{T}^d)$, $\Lambda \subseteq \mathbb{Z}^d$ be a finite subset, $c \in \mathbb{C}$ non-zero, and $y \in \mathbb{T}^d$.*

(a) *Multiplication by constants is bijective: $c\, \varepsilon(\mu, \Lambda) = \varepsilon(c\mu, \Lambda)$, and $\nu \in \mathcal{E}(\mu, \Lambda)$ if and only if $c\nu \in \mathcal{E}(c\mu, \Lambda)$.*

(b) *Translation is bijective: $\varepsilon(\mu, \Lambda) = \varepsilon(T_y\mu, \Lambda)$, and $\nu \in \mathcal{E}(\mu, \Lambda)$ if and only if $T_y\nu \in \mathcal{E}(T_y\mu, \Lambda)$.*

(c) *Minimal extrapolation is invariant under simultaneous shifts of $\widehat{\mu}$ and Λ: $\varepsilon(\mu, \Lambda) = \varepsilon(M_n\mu, \Lambda + n)$, and $\mathcal{E}(\mu, \Lambda) = \mathcal{E}(M_n\mu, \Lambda + n)$.*

(d) *The product of minimal extrapolations is a minimal extrapolation for the product: For $j = 1, 2$, let $\mu_j \in M(\mathbb{T}^{d_j})$, let $\Lambda_j \subseteq \mathbb{Z}^{d_j}$ be a finite subset, and let $\nu_j \in \mathcal{E}(\mu_j, \Lambda_j)$. Then $\varepsilon(\mu_1, \Lambda_1)\varepsilon(\mu_2, \Lambda_2) = \varepsilon(\mu_1 \times \mu_2, \Lambda_1 \times \Lambda_2)$, and $\nu_1 \times \nu_2 \in \mathcal{E}(\mu_1 \times \mu_2, \Lambda_1 \times \Lambda_2)$.*

Proof.

(a) If $\nu \in \mathcal{E}(\mu, \Lambda)$, then $c\nu$ is an extrapolation of $c\mu$ from Λ. Suppose $c\nu \notin \mathcal{E}(c\mu, \Lambda)$. Then there exists σ such that $\widehat{\sigma} = c\widehat{\mu}$ on Λ and $\|\sigma\| < \|c\nu\|$. Thus, $\widehat{\sigma}/c = \widehat{\mu}$ on Λ and $\|\sigma/c\| < \|\nu\|$, and this contradicts the assumption that $\nu \in \mathcal{E}(\mu, \Lambda)$. The converse follows by a similar argument.

(b) If $\nu \in \mathcal{E}(\mu, \Lambda)$, then $T_y\nu$ is an extrapolation of $T_y\mu$ from Λ. Suppose $T_y\nu \notin \mathcal{E}(T_y\mu, \Lambda)$. Then there exists σ such that $\widehat{\sigma} = (T_y\mu)^\wedge$ on Λ and $\|\sigma\| < \|T_y\nu\| = \|\nu\|$. Then $(T_{-y}\sigma)^\wedge = \widehat{\mu}$ on Λ and $\|T_{-y}\sigma\| = \|\sigma\| < \|\nu\|$, which contradicts the assumption that $\nu \in \mathcal{E}(\mu, \Lambda)$. The converse follows by a similar argument.

(c) If $f \in U(\mathbb{T}^d; \Lambda)$, then $M_n f \in U(\mathbb{T}^d; \Lambda + n)$. By Parseval's theorem,

$$\langle f, \mu \rangle = \sum_{m \in \Lambda} \widehat{f}(m)\overline{\widehat{\mu}(m)} = \sum_{m \in \Lambda + n} (M_n f)^\wedge(m)\overline{(M_n\mu)^\wedge(m)} = \langle M_n f, M_n\mu \rangle.$$

Using Proposition 3.1e, we see that $\varepsilon(\mu, \Lambda) = \varepsilon(M_n\mu, \Lambda + n)$.

If $\nu \in \mathcal{E}(\mu, \Lambda)$, then $M_n\nu$ is an extrapolation of $M_n\mu$ from $\Lambda + n$, and $\|M_n\| = \|\nu\| = \varepsilon(\mu, \Lambda) = \varepsilon(M_n\mu, \Lambda + n)$. The converse follows similarly.

(d) For convenience, let $\mu = \mu_1 \times \mu_2$, $\nu = \nu_1 \times \nu_2$, $\Lambda = \Lambda_1 \times \Lambda_2$, $\varepsilon_j = \varepsilon(\mu_j, \Lambda_j)$, and $\varepsilon = \varepsilon(\mu, \Lambda)$. Since ν is an extrapolation of μ, by Proposition 3.15, we have

$$\varepsilon \leq \|\nu\| = \|\nu_1\|\|\nu_2\| = \varepsilon_1\varepsilon_2.$$

To see the reverse inequality, we use Proposition 3.1f, and see that there exist $\varphi_j \in U(\mathbb{T}^{d_j}; \Lambda_j)$ such that $\varepsilon_j = \langle \varphi_j, \mu_j \rangle$, for $j = 1, 2$. Let $\varphi = \varphi_1 \otimes \varphi_2$ and observe that $\varphi \in U(\mathbb{T}^d; \Lambda)$. By Proposition 3.1e, we have

$$\varepsilon = \max_{f \in U(\mathbb{T}^d; \Lambda)} |\langle f, \mu \rangle| \geq |\langle \varphi, \mu \rangle| \geq \langle \varphi, \mu \rangle = \langle \varphi_1, \mu_1 \rangle\langle \varphi_2, \mu_2 \rangle = \varepsilon_1\varepsilon_2.$$

This shows that $\varepsilon = \varepsilon_1\varepsilon_2$, and, since $\|\nu\| = \varepsilon_1\varepsilon_2$, we conclude that $\nu \in \mathcal{E}(\mu, \Lambda)$.

$\square$

While minimal extrapolation is well-behaved under translation, it not well behaved under modulation. This is because the Fourier transform of modulation is translation, and so $\widehat{\mu}\,|_\Lambda$ and $(M_n\mu)^\wedge\,|_\Lambda$ are, in general, not equal. In contrast, the Fourier transform of translation is modulation, and so $\widehat{\mu}\,|_\Lambda$ and $(T_y\mu)^\wedge\,|_\Lambda$ only differ by a phase factor. We shall prove these statements in Proposition 3.20.

Proposition 3.20.

(a) *For $j = 1, 2$, there exist $\mu_j \in M(\mathbb{T})$, a finite subset $\Lambda \subseteq \mathbb{Z}$, and $\nu_j \in \mathcal{E}(\mu_j, \Lambda)$, such that $\nu_1 + \nu_2 \notin \mathcal{E}(\mu_1 + \mu_2, \Lambda)$.*

(b) *There exist $\mu \in M(\mathbb{T})$, a finite subset $\Lambda \subseteq \mathbb{Z}$, $\nu \in \mathcal{E}(\mu, \Lambda)$, and $n \in \mathbb{Z}$, such that $M_n \nu \notin \mathcal{E}(M_n \mu, \Lambda)$.*

(c) *For $j = 1, 2$, there exist $\mu_j \in M(\mathbb{T})$, a finite subset $\Lambda \subseteq \mathbb{Z}$, and $\nu_j \in \mathcal{E}(\mu_j, \Lambda)$, such that $\nu_1 * \nu_2 \notin \mathcal{E}(\mu_1 * \mu_2, \Lambda)$.*

Proof.

(a) Let $\mu_1 = \delta_0 + \delta_{1/2}$, $\mu_2 = -\delta_0 - \delta_{1/2}$, and $\Lambda = \{-1, 0, 1\}$. By Example 4.2, we have $\nu_1 = \delta_0 + \delta_{1/2} \in \mathcal{E}(\mu_1, \Lambda)$, and $\nu_2 = -\delta_{1/4} - \delta_{3/4} \in \mathcal{E}(\mu_2, \Lambda)$. Then, $\mu_1 + \mu_2 = 0$, and so $\varepsilon(\mu_1 + \mu_2, \Lambda) = 0$. However, $\nu_1 + \nu_2 \notin \mathcal{E}(\mu_1 + \mu_2, \Lambda)$ because $\|\nu_1 + \nu_2\| = \|\delta_0 - \delta_{1/4} + \delta_{1/2} - \delta_{3/4}\| > 0$.

(b) Let $\mu = \delta_0 + \delta_{1/2}$, $\Lambda = \{-1, 0, 1\}$, and $n = -1$. By Example 4.2, we have $\nu = \delta_{1/4} + \delta_{3/2} \in \mathcal{E}(\mu, \Lambda)$. However, $M_{-1}\mu = \delta_0 - \delta_{1/2}$, and by Example 4.3, $\mathcal{E}(M_{-1}\mu, \Lambda) = \{\delta_0 - \delta_{1/2}\}$. Thus, $M_{-1}\nu \notin \mathcal{E}(M_{-1}\mu, \Lambda)$.

(c) Let $\mu_1 = \delta_0 + \delta_{1/2}$, $\mu_2 = \delta_0 - \delta_{1/2}$, and $\Lambda = \{-1, 0, 1\}$. Then $(\mu_1 * \mu_2)^\wedge = 0$ on Λ, which implies $\varepsilon(\mu_1 * \mu_2, \Lambda) = 0$. By Examples 4.2 and 4.3, $\nu_1 = \mu_1 \in \mathcal{E}(\mu_1, \Lambda)$ and $\nu_2 = \mu_2 \in \mathcal{E}(\mu_2, \Lambda)$. However, $\nu_1 * \nu_2 \notin \mathcal{E}(\mu_1 * \mu_2, \Lambda)$ because $\|\nu_1 * \nu_2\| = \|\delta_0 - \delta_1\| > 0$.

$\square$

4. EXAMPLES

4.1. **Discrete measures.** There are several reasons why we are interested in computing the minimal extrapolations of discrete measures. They are the simplest types of measures, and so, their minimal extrapolations can be computed rather easily. By Theorem 3.3, the minimal extrapolations of a non-discrete measure are sometimes discrete measures, so they appear naturally in our analysis. As discussed in Section 2.2, examples of μ that have minimal extrapolations supported in a lattice can be interpreted in the context of deterministic compressed sensing. Examples 4.2-4.5 and Example 4.7 can be written in the context of compressed sensing.

Remark 4.1. In view of Proposition 3.19 a,b, and without loss of generality, we can assume any discrete measure $\mu = \sum_{k=1}^{\infty} a_k \delta_{x_k} \in M(\mathbb{T}^d)$, where $\sum_{k=1}^{\infty} |a_k| < \infty$, can be written as $\mu = \delta_0 + \sum_{k=2}^{\infty} a'_k \delta_{x'_k} \in M(\mathbb{T}^d)$, where $\sum_{k=2}^{\infty} |a'_k| < \infty$.

Example 4.2. Let $\mu = \delta_0 + \delta_{1/2}$, and $\Lambda = \{-1, 0, 1\}$. We have $\widehat{\mu}(0) = 2$, and $\widehat{\mu}(\pm 1) = 0$. Clearly $\|\widehat{\mu}\|_{\ell^\infty(\Lambda)} = \|\mu\| = 2$. By Proposition 3.15, $\varepsilon = \|\widehat{\mu}\|_{\ell^\infty(\Lambda)} = \|\mu\| = 2$, which implies $\mu \in \mathcal{E}$.

Further, there is an uncountable number of discrete minimal extrapolations. To see this, for each $y \in \mathbb{T}$ and any integer $K \geq 2$, define

$$\nu_{y,K} = \frac{2}{K} \sum_{k=0}^{K-1} \delta_{y + \frac{k}{K}}.$$

A straightforward calculation shows that $\nu_{y,K}$ is an extrapolation and that $\|\nu_{y,K}\| = \varepsilon$.

Also, we can construct positive absolutely continuous minimal extrapolations. One example is the constant function $f \equiv 2$ on $\mathbb{T}$. For other examples, let $N \geq 2$ and let $F_N \in C^\infty(\mathbb{T})$ be the Fejér kernel,

$$F_N(x) = \sum_{n=-N}^{N} \left(1 - \frac{|n|}{N+1}\right) e^{2\pi i n x}.$$

For any $c > 0$ such that $c \leq (2N+2)/(3N+1)$, extend $\widehat{\mu}|_\Lambda$ to the sequence $\{(a_{N,c})_m : m \in \mathbb{Z}\}$, where

$$(a_{N,c})_m = \begin{cases} 2 & m = 0, \\ c\left(1 - \frac{|m|}{N+1}\right) & 2 \leq |m| \leq N, \\ 0 & \text{otherwise.} \end{cases}$$

Consider the real-valued function

$$f_{N,c}(x) = 2 + \sum_{n=-N}^{-2} (a_{N,c})_n e^{2\pi inx} + \sum_{n=2}^{N} (a_{N,c})_n e^{2\pi inx}$$

$$= 2 + c\sum_{n=-N}^{-2} \Big(1 - \frac{|n|}{N+1}\Big) e^{2\pi inx} + c\sum_{n=2}^{N} \Big(1 - \frac{|n|}{N+1}\Big) e^{2\pi inx}.$$

We check that $\widehat{f_{N,c}}(m) = (a_{N,c})_m$ for all $m \in \mathbb{Z}$, which implies $f_{N,c}$ is an extrapolation of μ. Using the upper bound on c, we have, for all $x \in \mathbb{T}$, that

$$2 \geq c + 2c\Big(1 - \frac{1}{N+1}\Big)\cos(2\pi x) = c + c\Big(1 - \frac{1}{N+1}\Big)e^{2\pi ix} + c\Big(1 - \frac{1}{N+1}\Big)e^{-2\pi ix}.$$

Using this inequality, and the definitions of F_N and $f_{N,c}$, we have

$$f_{N,c}(x) \geq cF_N \geq 0.$$

Since $f_{N,c} \geq 0$, we also have

$$\|f_{N,c}\|_1 = \int_{\mathbb{T}} f_{N,c}(x)\,dx = 2 + \int_{\mathbb{T}} \Big(\sum_{n=-N}^{-2} (a_{N,c})_n e^{2\pi inx} + \sum_{n=2}^{N}(a_{N,c})_n e^{2\pi inx}\Big)\,dx = 2 = \varepsilon.$$

Thus, for any $N \geq 2$ and $c \leq (2N+2)/(3N+1)$, $f_{N,c}$ is a *positive absolutely continuous* minimal extrapolation. Hence, we have constructed an uncountable number of positive absolutely continuous minimal extrapolations.

Example 4.3. Let $\mu = \delta_0 - \delta_{1/2}$, and $\Lambda = \{-1, 0, 1\}$. We have $\widehat{\mu}(0) = 0$, and $\widehat{\mu}(\pm 1) = 2$. Further, we have $\|\widehat{\mu}\|_{\ell^\infty(\Lambda)} = \|\mu\| = 2$, so that by Proposition 3.15, we have $\varepsilon = \|\widehat{\mu}\|_{\ell^\infty(\Lambda)} = \|\mu\| = 2$ and $\mu \in \mathcal{E}$.

Consequently, $\Gamma = \{-1, 1\}$. By Theorem 3.3b, there exists $\alpha_{-1,1} \in \mathbb{R}/\mathbb{Z}$ satisfying

$$e^{2\pi i\alpha_{-1,1}} = \frac{\widehat{\mu}(-1)}{\widehat{\mu}(1)} = 1,$$

and the minimal extrapolations are supported in the set $\{x \in \mathbb{T} : 2x \in \mathbb{Z}\} = \{0, 1/2\}$. This implies each $\nu \in \mathcal{E}$ is discrete and can be written as $\nu = a_1\delta_0 + a_2\delta_{1/2}$. In theory, a_1, a_2 depend on ν, so we cannot conclude uniqueness yet.

The matrix E from (3.2) is

$$E(-1, 0, 1; 0, 1/2) = \begin{pmatrix} 1 & -1 \\ 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Clearly, E has full column rank, so, by Proposition 3.8, μ is the unique minimal extrapolation. Thus, super-resolution reconstruction of μ from Λ is possible.

Example 4.4. Let $\mu = \delta_0 - \delta_{1/4}$, and let $\Lambda = \{-1, 0, 1\}$. We have $\widehat{\mu}(\pm 1) = 1 \pm i = \sqrt{2}e^{\pm\pi i/4}$, and $\widehat{\mu}(0) = 0$. Note that $\|\widehat{\mu}\|_{\ell^\infty(\Lambda)} = \sqrt{2} < 2 = \|\mu\|$, which shows that $\sqrt{2} \leq \varepsilon \leq 2$. We claim that $\varepsilon = \sqrt{2}$. To see this, consider $\nu = (-\delta_{3/8} + \delta_{7/8})/\sqrt{2}$. We verify that $\|\nu\| = \sqrt{2}$ and that ν is an extrapolation. By Proposition 3.15, $\varepsilon = \sqrt{2}$ and $\nu \in \mathcal{E}$. This also implies $\mu \notin \mathcal{E}$, and so super-resolution reconstruction of μ from Λ is impossible.

We claim that ν is the unique minimal extrapolation. The matrix E from (3.2) is

$$E(-1, 0, 1; 3/8, 7/8) = \begin{pmatrix} e^{2\pi i3/8} & e^{2\pi i7/8} \\ 1 & 1 \\ e^{-2\pi i3/8} & e^{-2\pi i7/8} \end{pmatrix},$$

which we observe to have full column rank. By Proposition 3.8, we conclude that ν is the unique minimal extrapolation. Thus, super-resolution reconstruction of ν from Λ is possible.

We explain the derivation of ν. We guess that $\varepsilon = \sqrt{2}$ and see what Theorem 3.3b implies. Under this assumption that $\varepsilon = \sqrt{2}$, we have $\Gamma = \{-1, 1\}$. By Theorem 3.3b, there exists $\alpha_{-1,1} \in \mathbb{R}/\mathbb{Z}$ satisfying

$$e^{2\pi i \alpha_{1,-1}} = \frac{\widehat{\mu}(1)}{\widehat{\mu}(-1)} = e^{\pi i/2},$$

and the minimal extrapolations are supported in $\{x \in \mathbb{T} : 2x + 1/4 \in \mathbb{Z}\} = \{3/8, 7/8\}$. Hence, if $\varepsilon = \sqrt{2}$, then every $\sigma \in \mathcal{E}$ is of the form $\sigma = a_1 \delta_{3/8} + a_2 \delta_{7/8}$. Thus, by definition of a minimal extrapolation, $\|\sigma\| = \sqrt{2}$ and $\widehat{\mu} = \widehat{\sigma}$ on Λ. Using this information, we solve for the coefficients a_1, a_2, and compute that $|a_1| = |a_2| = \sqrt{2}$, $a_1 = -a_2$, and $a_1 = -\sqrt{2}/2$. Thus, we obtain that $\sigma = (-\delta_{3/8} + \delta_{7/8})/\sqrt{2}$. From here, we simply check that $\nu = \sigma$ is, in fact, a minimal extrapolation.

Example 4.5. Let $\mu = \delta_0 + e^{\pi i/3} \delta_{1/3}$ and let $\Lambda = \{-1, 0, 1\}$. We have $\widehat{\mu}(-1) = 0$, $\widehat{\mu}(0) = 1 + e^{\pi i/3} = \sqrt{3} \, e^{\pi i/6}$, and $\widehat{\mu}(1) = 1 + e^{-\pi i/3} = \sqrt{3} \, e^{-\pi i/6}$.

Suppose, for the purpose of obtaining a contradiction, that $\varepsilon = \|\widehat{\mu}\|_{\ell^\infty(\Lambda)} = \sqrt{3}$. Then $\Gamma = \{0, 1\}$. By Theorem 3.3b, there is $\alpha_{0,1} \in \mathbb{R}/\mathbb{Z}$ such that

$$e^{2\pi i \alpha_{0,1}} = \frac{\widehat{\mu}(0)}{\widehat{\mu}(1)} = e^{\pi i/3},$$

and each $\nu \in \mathcal{E}$ is of the form $\nu = a \delta_{1/6}$ for some $a \in \mathbb{C}$. Then, $|\widehat{\nu}| = |a|$ on $\mathbb{Z}$ and, in particular, $\widehat{\mu} \neq \widehat{\nu}$ on Λ, which is a contradiction.

Thus, $\varepsilon > \sqrt{3}$, i.e., $\Gamma = \emptyset$. Therefore, Theorem 3.3a applies, and so there is a finite set S such that $\mathrm{supp}(\nu) \subseteq S$ for each $\nu \in \mathcal{E}$. In particular, each $\nu \in \mathcal{E}$ is discrete. Hence, we have to solve the optimization problem (3.3), which is

$$\varepsilon = \max\left\{ a\sqrt{3} \, e^{\pi i/6} + b\sqrt{3} \, e^{-\pi i/6} : \forall x \in \mathbb{T}, \ |a e^{2\pi i x} + b + c e^{-2\pi i x}| \leq 1, \ a, b, c \in \mathbb{C} \right\}.$$

This optimization problem can be written as a semi-definite program, see [CFG14, Corollary 4.1, page 936]. After obtaining numerical approximations to the optimizers of this problem, we guess that the the exact optimizers are

$$a = \frac{2}{3\sqrt{3}} e^{-\pi i/6}, \quad b = \frac{4}{3\sqrt{3}} e^{\pi i/6}, \quad c = -\frac{i}{3\sqrt{3}}.$$

These values of a, b, c are, in fact, the optimizers because $a\sqrt{3} \, e^{\pi i/6} + b\sqrt{3} \, e^{-\pi i/6} = 2$ and $|a e^{2\pi i x} + b + c e^{-2\pi i x}| \leq 1$ for all $x \in \mathbb{T}$. Thus, $\varepsilon = 2$ and $\mu \in \mathcal{E}$. Since $|a e^{2\pi i x} + b + c e^{-2\pi i x}| = 1$ precisely on $S = \{0, 1/3\}$, by Theorem 3.3a, the minimal extrapolations are supported in S.

The matrix E from (3.2) is

$$E(-1, 0, 1; 0, 1/3) = \begin{pmatrix} 1 & e^{-2\pi i/3} \\ 1 & 1 \\ 1 & e^{2\pi i/3} \end{pmatrix}.$$

Since E has full rank, by Proposition 3.8, μ is the unique minimal extrapolation. Thus, super-resolution reconstruction of μ from Λ is possible.

The following example illustrates that if μ is a sum of two Dirac measures, then their supports have to be sufficiently spaced apart in order for super-resolution of μ to be possible. This shows that, in general, a minimum separation condition is necessary to super-resolve a sum of two Dirac measures, see [CFG14].

Example 4.6. Let $\mu_y = \delta_0 - \delta_y$ for some non-zero $y \in \mathbb{T}^d$ and let $\Lambda \subseteq \mathbb{Z}^d$ be a finite subset. We claim that if y is sufficiently small depending on Λ, then $\mu_y \notin \mathcal{E}(\mu_y, \Lambda)$. Note that $\|\mu_y\| = 2$ for any $y \in \mathbb{T}^d$. Let η denote the normalized Lebesgue measure on $\mathbb{T}^d$ and define the measures ν_y by the formula

$$\nu_y(x) = \sum_{m \in \Lambda} \widehat{\mu_y}(m) e^{2\pi i m \cdot x} \, \eta(x).$$

By construction, ν_y is an extrapolation of μ_y because, for each $n \in \Lambda$,

$$\widehat{\nu_y}(n) = \int_{\mathbb{T}^d} e^{-2\pi i n \cdot x} \, d\nu_y(x) = \sum_{m \in \Lambda} \widehat{\mu_y}(m) \int_{\mathbb{T}^d} e^{-2\pi i (n-m) \cdot x} \, dx = \widehat{\mu_y}(n).$$

Then, as $y \to 0$,

$$\|\nu_y\| = \int_{\mathbb{T}^d} \Big| \sum_{m \in \Lambda} \widehat{\mu_y}(m) e^{2\pi i m \cdot x} \Big| \, dx \ = \int_{\mathbb{T}^d} \Big| \sum_{m \in \Lambda} (1 - e^{-2\pi i m \cdot y}) e^{2\pi i m \cdot x} \Big| \, dx \to 0.$$

Thus, we take y sufficiently small so that $\|\nu_y\| < 2 = \|\mu_y\|$, and, hence, $\mu_y \notin \mathcal{E}(\mu_y, \Lambda)$. Note that this argument does not contradict Proposition 3.15 because $\|\widehat{\mu_y}\|_{\ell^\infty(\Lambda)} \to 0$ as $y \to 0$. Thus, for y sufficiently small, super-resolution reconstruction of μ_y from Λ is impossible.

4.2. Singular continuous measures.

The following example is an analogue of Example 4.2 for higher dimensions.

Example 4.7. Let $\mu = \delta_{(0,0)} + \delta_{(1/2,1/2)}$, and $\Lambda = \{-1, 0, 1\}^2 \setminus \{(1,-1), (-1,1)\}$. Then, $\widehat{\mu}(m) = 1 + e^{-\pi i (m_1 + m_2)}$, and, in particular, $\widehat{\mu}(1,1) = \widehat{\mu}(-1,-1) = \widehat{\mu}(0,0) = 2$ and $\widehat{\mu}(\pm 1, 0) = \widehat{\mu}(0, \pm 1) = 0$. We deduce that $\varepsilon = \|\mu\| = \|\widehat{\mu}\|_{\ell^\infty(\Lambda)} = 2$ from Proposition 3.15, and so $\mu \in \mathcal{E}$.

Further, $\Gamma = \{(0,0), (1,1), (-1,-1)\}$, and so $\#\Gamma = 3$. According to the definition of $\alpha_{m,n}$ in Theorem 3.3b, set $\alpha_{(-1,-1),(0,0)} = \alpha_{(0,0),(1,1)} = \alpha_{(-1,-1),(1,1)} = 0$. By the conclusion of Theorem 3.3b, the minimal extrapolations are supported in the set $S = S_{(-1,-1),(0,0)} \cap S_{(0,0),(1,1)} \cap S_{(-1,-1),(1,1)}$, where

$$S_{(-1,-1),(0,0)} = \{x \in \mathbb{T}^2 : x \cdot (-1,-1) \in \mathbb{Z}\} = \{x \in \mathbb{T}^2 : x_1 + x_2 \in \mathbb{Z}\},$$
$$S_{(0,0),(1,1)} = \{x \in \mathbb{T}^2 : x \cdot (-1,-1) \in \mathbb{Z}\} = \{x \in \mathbb{T}^2 : x_1 + x_2 \in \mathbb{Z}\},$$
$$S_{(-1,-1),(1,1)} = \{x \in \mathbb{T}^2 : x \cdot (-2,-2) \in \mathbb{Z}\} = \{x \in \mathbb{T}^2 : 2x_1 + 2x_2 \in \mathbb{Z}\}.$$

It follows that the minimal extrapolations are supported in

$$S = S_{(-1,-1),(0,0)} \cap S_{(0,0),(1,1)} \cap S_{(-1,-1),(1,1)} = \{x \in \mathbb{T}^2 : x_1 + x_2 = 1\}.$$

We can construct other discrete minimal extrapolations besides μ. For each $y \in \mathbb{T}$ and for each integer $K \geq 2$, define the measure

$$\nu_{y,K} = \frac{2}{K} \sum_{k=0}^{K-1} \delta_{\left(y + \frac{k}{K}, 1 - y - \frac{k}{K}\right)}.$$

We claim $\nu_{y,K}$ is a minimal extrapolation. We have $\|\nu_{y,K}\| = \varepsilon$, and

$$\widehat{\nu_{y,K}}(m) = e^{-2\pi i (m_1 y - m_2 y)} e^{-2\pi i m_2} \frac{2}{K} \sum_{k=0}^{K-1} e^{-2\pi i (m_1 - m_2) k / K}.$$

We see that $\widehat{\nu_{y,K}} = \widehat{\mu}$ on Λ, which proves the claim.

We can also construct continuous singular minimal extrapolations. Let $\sigma = \sqrt{2}\sigma_S$, where σ_S is the *surface measure* of the Borel set S. We readily verify that $\|\sigma\| = \varepsilon$ and

$$\widehat{\sigma}(m) = \sqrt{2} \int_{\mathbb{T}^2} e^{-2\pi i m \cdot x} \, d\sigma_S = 2e^{-2\pi i m_2} \int_0^1 e^{-2\pi i (m_1 - m_2) t} \, dt = 2\delta_{m_1, m_2},$$

which proves that $\sigma \in \mathcal{E}$. In particular, S is the smallest set that contains the support of all the minimal extrapolations.

Since μ is not the unique minimal extrapolation, super-resolution reconstruction of μ from Λ is impossible.

Example 4.8. For an integer $q \geq 3$, let C_q be the *middle $1/q$-Cantor set*, which is defined by $C_q = \bigcap_{k=0}^{\infty} C_{q,k}$, where $C_{q,0} = [0,1]$ and

$$C_{q,k+1} = \frac{C_{q,k}}{q} \cup \left((1-q) + \frac{C_{q,k}}{q} \right).$$

Let $F_q \colon [0,1] \to [0,1]$ be the *Cantor-Lebesgue* function on C_q, which is defined by the point-wise limit of the sequence $\{F_{q,k}\}$, where $F_{q,0}(x) = x$ and

$$F_{q,k+1}(x) = \begin{cases} \frac{1}{2} F_{q,k}(qx) & 0 \leq x \leq \frac{1}{q}, \\ \frac{1}{2} & \frac{1}{q} \leq x \leq \frac{q-1}{q}, \\ \frac{1}{2} F_{q,k}(qx - (q-1)) + \frac{1}{2} & \frac{q-1}{q} \leq x \leq 1. \end{cases}$$

By construction, $F_q(0) = 0$, $F_q(1) = 1$, and F_q is non-decreasing and uniformly continuous on $[0,1]$. Thus, F_q can be uniquely identified with the measure $\sigma_q \in M(\mathbb{T})$, and $\|\sigma_q\| = 1$. Since $F_q' = 0$ a.e. and F_q does not have any jump discontinuities, σ_q is a continuous singular measure, with zero discrete part. The Fourier coefficients of σ_q are

$$\widehat{\sigma_q}(m) = (-1)^m \prod_{k=1}^{\infty} \cos(\pi m q^{-k}(1-q)),$$

see [Zyg59, pages 195-196]. In particular, for any integer $n \geq 1$, we have

$$\widehat{\sigma_q}(q^n) = (-1)^{q^n} \prod_{k=1}^{\infty} \cos(\pi q^{-k}(1-q)),$$

which is convergent and independent of n. Since $\widehat{\sigma_q}(0) = \|\sigma_q\| = 1$, we immediately see that $\varepsilon = 1$ and $\sigma_q \in \mathcal{E}$. Again, we cannot determine whether σ_q is the unique minimal extrapolation because Theorem 3.3 cannot handle the case $\#\Gamma = 1$, see Remark 3.12.

Example 4.9. Let $\sigma_A, \sigma_B \in M(\mathbb{T}^d)$ be the surface measures of the Borel sets $A = \{x \in \mathbb{T}^2 : x_2 = 0\}$ and $B = \{x \in \mathbb{T}^2 : x_2 = 1/2\}$, respectively. Let $\mu = \sigma_A + \sigma_B$, and $\Lambda = \{-2, -1, \ldots, 1\}^2$. Then,

$$\widehat{\mu}(m) = \int_0^1 e^{2\pi i m_1 t} \, dt + \int_0^1 e^{2\pi i (m_1 t + m_2/2)} \, dt = \delta_{m_1,0} + (-1)^{m_2} \delta_{m_1,0}.$$

We immediately see that $\varepsilon = \|\widehat{\mu}\|_{\ell^\infty(\Lambda)} = \|\mu\| = 2$, which implies $\mu \in \mathcal{E}$. Then, $\Gamma = \{(0,0), (0,2), (0,-2)\}$, and, by Theorem 3.3b, the minimal extrapolations are supported in $\{x \in \mathbb{T}^2 : x_2 = 0\} \cup \{x \in \mathbb{T}^2 : x_2 = 1/2\}$. Determining whether μ is the unique minimal extrapolation is beyond the theory we have developed herein, see Remark 3.18. We shall examine this uniqueness problem in [BL16b].

5. Acknowledgements

References

[ADCG15] Jean-Marc Azaïs, Yohann De Castro, and Fabrice Gamboa. Spike detection from inaccurate samplings. *Applied and Computational Harmonic Analysis*, 38(2):177–195, 2015.

[ASB14] Céline Aubel, David Stotz, and Helmut Bolcskei. Super-resolution from short-time Fourier transform measurements. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 36–40. IEEE, 2014.

[ASB15] Céline Aubel, David Stotz, and Helmut Bölcskei. A theory of super-resolution from short-time Fourier transform measurements. *arXiv preprint arXiv:1509.01047*, 2015.

[AYB15] Enrico Au-Yeung and John J. Benedetto. Generalized Fourier frames in terms of balayage. *Journal of Fourier Analysis and Applications*, 21(3):472–508, 2015.

[BC09]   John J. Benedetto and Wojciech Czaja. *Integration and Modern Analysis*. Birkhäuser Boston, Inc., Boston, 2009.

[Ben75] John J. Benedetto. *Spectral Synthesis*. Academic Press, Inc., New York-London, 1975.

[Ben96] John J. Benedetto. *Harmonic Analysis and Applications*. CRC Press Inc., 1996.

[Beu89a] Arne Beurling. Balayage of Fourier-Stieltjes transforms. *The Collected Works of Arne Beurling*, 2:341–350, 1989.

[Beu89b] Arne Beurling. Interpolation for an interval in $\mathbb{R}^1$. *The Collected Works of Arne Beurling*, 2:351–365, 1989.

[BL16a] John J. Benedetto and Weilin Li. Super-resolution of bounded measures on $\mathbb{R}^d$. *In Preparation*, 2016.

[BL16b] John J. Benedetto and Weilin Li. Super-resolution of singular continuous measures on $\mathbb{T}^d$. *In Preparation*, 2016.

[BTR13] Badri Narayan Bhaskar, Gongguo Tang, and Benjamin Recht. Atomic norm denoising with applications to line spectral estimation. *IEEE Transactions on Signal Processing*, 61(23):5987–5999, 2013.

[CFG13] Emmanuel J. Candès and Carlos Fernandez-Granda. Super-resolution from noisy data. *Journal of Fourier Analysis and Applications*, 19(6):1229–1254, 2013.

[CFG14] Emmanuel J. Candès and Carlos Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, 2014.

[Chi78] Donald G. Childers. *Modern Spectrum Analysis*. IEEE Press, Piscataway, New Jersey, 1978.

[CRT06a] Emmanuel J. Candès, Justin K. Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Information Theory*, 52(2):489–509, 2006.

[CRT06b] Emmanuel J. Candès, Justin K. Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006.

[DCG12] Yohann De Castro and Fabrice Gamboa. Exact reconstruction using Beurling minimal extrapolation. *Journal of Mathematical Analysis and Applications*, 395(1):336–354, 2012.

[Don92] David L. Donoho. Superresolution via sparsity constraints. *SIAM Journal on Mathematical Analysis*, 23(5):1309–1331, 1992.

[Don06] David L. Donoho. Compressed sensing. *IEEE Trans. Information Theory*, 52(4):1289–1306, 2006.

[Dum07] Bogdan Dumitrescu. *Positive Trigonometric Polynomials and Signal Processing Applications*. Springer, Dordrecht, 2007.

[Ess45] Carl-Gustav Esseen. Fourier analysis of distribution functions. a mathematical study of the Laplace-Gaussian law. *Acta Mathematica*, 77(1):1–125, 1945.

[Ess54] Carl-Gustav Esseen. A note on Fourier-Stieltjes transforms and absolutely continuous functions. *Math. Scand*, 2:153–157, 1954.

[FG15]   Carlos Fernandez-Granda. Super-resolution of point sources via convex programming. *arXiv preprint arXiv:1507.07034*, 2015.

[Gre09] Hayit Greenspan. Super-resolution in medical imaging. *The Computer Journal*, 52(1):43–63, 2009.

[HR55]   Edwin Hewitt and Herman Rubin. The maximum value of a Fourier-Stieltjes transform. *Math. Scand*, 3:97–102, 1955.

[HT29]   Einar Hille and J.D. Tamarkin. Remarks on a known example of a monotone continuous function. *American Mathematical Monthly*, 36(5):255–264, 1929.

[Kah70] Jean-Pierre Kahane. *Séries de Fourier Absolument Convergentes*. Springer-Verlag, Berlin-New York, 1970.

[Kah12] Jean-Pierre Kahane. Analyse et synthése harmoniques. *Histoire de Mathématiques,* Ed. Éc. Polytech., Palaiseau, pages 17–53, 2012.

[KLM04] Valery Khaidukov, Evgeny Landa, and Tijmen Jan Moser. Diffraction imaging by focusing-defocusing: An outlook on seismic superresolution. *Geophysics*, 69(6):1478–1490, 2004.

[KMM07] Aggelos K. Katsaggelos, Rafael Molina, and Javier Mateos. *Super Resolution of Images and Video*. Morgan & Claypool Publishers, 2007.

[Lan87]   Henry J. Landau. Maximum entropy and the moment problem. *Bulletin of the American Mathematical Society*, 16(1):47–77, 1987.

[Lin12]   Jari Lindberg. Mathematical concepts of optical superresolution. *Journal of Optics*, 14(8):083001, 2012.

[PK05]   K.G. Puschmann and F. Kneer. On super-resolution in astronomical imaging. *Astronomy & Astrophysics*, 436(1):373–378, 2005.

[PPK03]   Sung Cheol Park, Min Kyu Park, and Moon Gi Kang. Super-resolution image reconstruction: a technical overview. *Signal Processing Magazine, IEEE*, 20(3):21–36, 2003.

[Rie99]   Fred Rieke. *Spikes: Exploring the Neural code*. MIT press, 1999.

[Rud62]   Walter Rudin. *Fourier Analysis on Groups*. John Wiley & Sons, Inc., New York, 1962.

[Sal42]   Raphaël Salem. On singular monotonic functions of the Cantor type. *J. Math. Phys*, 21:69–82, 1942.

[Sle78]   David Slepian. Prolate spheroidal wave functions, Fourier analysis, and uncertainty - V: The discrete case. *Bell System Technical Journal*, 57(5):1371–1430, 1978.

[TBR15]   Gongguo Tang, Badri Narayan Bhaskar, and Benjamin Recht. Near minimax line spectral estimation. *IEEE Transactions on Information Theory*, 61(1):499–512, 2015.

[TBSR13]   Gongguo Tang, Badri Narayan Bhaskar, Parikshit Shah, and Benjamin Recht. Compressed sensing off the grid. *IEEE Trans. Information Theory*, 59(11):7465–7490, 2013.

[Zyg59]   Antoni Zygmund. *Trigonometric Series*, volume 1. Cambridge University Press, 2002 (1959).

NORBERT WIENER CENTER, DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MARYLAND, COLLEGE PARK, MD 20742, USA

E-mail address: `jjb@math.umd.edu`

NORBERT WIENER CENTER, DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MARYLAND, COLLEGE PARK, MD 20742, USA

E-mail address: `wl298@math.umd.edu`