ARTICLE IN PRESS

YACHA:1264

Appl. Comput. Harmon. Anal. ••• (••••) •••–•••

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Super-resolution by means of Beurling minimal extrapolation

John J. Benedetto *, Weilin Li *

Norbert Wiener Center, Department of Mathematics, University of Maryland, College Park, MD 20742, USA

A R T I C L E   I N F O

Article history:
Received 14 August 2016
Received in revised form 6 February 2018
Accepted 2 May 2018
Available online xxxx
Communicated by Jared Tanner

MSC:
43A25
46E27
46N10
94A12

Keywords:
Super-resolution
Minimal extrapolation
Total variation optimization

A B S T R A C T

We investigate the super-resolution capabilities of total variation minimization. Namely, given a finite set $\Lambda \subseteq \mathbb{Z}^d$ and spectral data $F = \widehat{\mu}\,|_\Lambda$, where μ is an unknown bounded Radon measure on the torus $\mathbb{T}^d$, the problem is to find the measures with smallest norm whose Fourier transforms agree with F on Λ. Our main theorem shows that solutions to the problem depend crucially on a set $\Gamma \subseteq \Lambda$, defined in terms of F and Λ. For example, when $\#\Gamma = 0$, the solutions are singular measures supported in the zero set of an analytic function, and when $\#\Gamma \geq 2$, the solutions are singular measures supported in the intersection of $\binom{\#\Gamma}{2}$ hyperplanes. 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 solutions. This theorem has implications to the possibility *and* impossibility of uniquely recovering μ from F on 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. Our concept of an *admissibility range* fundamentally connects Beurling's theory of minimal extrapolation [7,8] with Candès and Fernandez-Granda's work on super-resolution [12]. This connection is exploited to address situations where current algorithms fail to compute a numerical solution to the total variation minimization problem.

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 [37], super-resolution refers to the process of up-sampling an image onto a finer grid, which is a *spatial interpolation* procedure. In other situations [33], 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

* Corresponding authors.
E-mail addresses: jjb@math.umd.edu (J.J. Benedetto), wl298@math.umd.edu (W. Li).

https://doi.org/10.1016/j.acha.2018.05.002
1063-5203/

Please cite this article in press as: J.J. Benedetto, W. Li, Super-resolution by means of Beurling minimal extrapolation, Appl. Comput. Harmon. Anal. (2018), https://doi.org/10.1016/j.acha.2018.05.002

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 a focused spectral extrapolation version, which we simply refer to as the total variation (TV) minimization problem, see (TV) for a precise statement.

Although our point of view is mathematical, we do want to set the stage briefly by mentioning the following applications and scenarios, noting that these are just the tip of the iceberg, see, e.g., [39,31].

(a) In astronomy [38], 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, 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 [26], 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.
(c) In certain applications, an image is obtained by convolving the object with the *point spread function* of an optical lens. 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 [33]. 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.

1.2. Problem statement

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 $*$. The support of a measure μ is denoted by $\mathrm{supp}(\mu)$. 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 as

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

See [5] for further details.

We say that F is *spectral data on a set* $\Lambda \subseteq \mathbb{Z}^d$ if $F = \widehat{\mu}\,|_\Lambda$ for some $\mu \in M(\mathbb{T}^d)$. One method for producing an approximation of the unknown μ is to solve the *total variation minimization problem*: For a given finite subset $\Lambda \subseteq \mathbb{Z}^d$ and given spectral data F on Λ, find the solution(s) ν to

$$\inf_{\nu} \|\nu\| \quad \text{such that} \quad \nu \in M(\mathbb{T}^d) \quad \text{and} \quad F = \widehat{\nu} \quad \text{on } \Lambda. \tag{TV}$$

This is a *convex minimization problem* and we interpret a solution ν as a *simple* or *least complicated* high resolution extrapolation of F.

Remark 1.1 *(Image processing).*

(a) A primary objective is to determine if $\mu \in M(\mathbb{T}^d)$ is a solution to the super-resolution problem given $F = \widehat{\mu}\,|_\Lambda$. The current literature has focused on discrete $\mu \in M(\mathbb{T}^d)$, e.g., see [12,43,25]. However, in the context of image processing, μ is the unknown high resolution image, and a typical image cannot be approximated realistically by a discrete measure, although such approximation is possible mathematically in the weak-$*$ topology [5]. Hence, it is important to determine if non-discrete measures are solutions to the super-resolution problem. For this reason, we have formulated and shall study (TV) in an abstract way.

(b) Further, in the context of image processing, we can think of being given information that represents an image $\mu \in M(\mathbb{T}^d)$, and that this information is in the form,

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

for some $\psi\colon \mathbb{T}^d \to \mathbb{C}$. 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. This is the background for our formulation of (TV).

(c) A common approach for *discrete* image recovery in inverse problems in medical imaging is to minimize the ℓ^1 norm of the finite differences of the image, which promotes the recovery of piecewise constant images. A different but related approach is to study the recovery of *continuous* domain piecewise constant images. For example, [35] showed that the edge set of a piecewise constant image is uniquely identifiable from low-pass Fourier coefficients of the image when it coincides with the zero-set of a trigonometric polynomial, and [34] studied a practical approach for spectral extrapolation using low-rank Toeplitz matrix completion. Related works [36,24] studied the recovery of multi-dimensional singular measures supported on curves defined by zero-sets of analytic functions. Our paper studies a similar formulation, but our recovery method, TV minimization, is quite different from the aforementioned papers.

With regard to (TV), a fundamental question of uniqueness must be addressed. To see why this is so, if μ is not the unique solution, 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 *reconstruction of μ from F on Λ is possible* if and only if μ is the unique solution to (TV). Of course, it could be theoretically possible to reconstruct μ by other means.

1.3. Background

The mathematical theory for (TV) almost exclusively pertains to discrete measures satisfying a strong separation condition. If $\Lambda = \{-M, -M+1, \dots, M\}^d \subseteq \mathbb{Z}^d$, we say a discrete μ satisfies the *minimum separation condition with constant $C > 0$* if

$$\min_{x,y\in \mathrm{supp}(\mu),\ x\neq y} \|x-y\|_{\ell^\infty(\mathbb{T}^d)} \geq \frac{C}{M},$$

where $\|\cdot\|_{\ell^\infty(\mathbb{T}^d)}$ is the ℓ^∞ norm on $\mathbb{T}^d$.

Please cite this article in press as: J.J. Benedetto, W. Li, Super-resolution by means of Beurling minimal extrapolation, Appl. Comput. Harmon. Anal. (2018), https://doi.org/10.1016/j.acha.2018.05.002

Theorem 1.2 *(Candès and Fernandez-Granda, Theorems 1.2–1.3, [12], Fernandez-Granda, Theorem 2.2, [25]). Let F be spectral data on $\Lambda = \{-M, -M+1, \dots, M\}^d \subseteq \mathbb{Z}^d$, where $F = \widehat{\mu}\,|_\Lambda$ for some discrete $\mu \in M(\mathbb{T}^d)$. In any of the following cases, μ is the unique solution to (TV).*

(a) *$d = 1$, $M \geq 128$, and μ satisfies the minimum separation condition with $C = 2$.*
(b) *$d = 1$, $M \geq 128$, and μ is real-valued and satisfies the minimum separation condition with $C = 1.87$.*
(c) *$d = 1$, $M \geq 10^3$, and μ satisfies the minimum separation condition with $C = 1.26$.*
(d) *$d = 2$, $M \geq 512$, and μ is real-valued and satisfies the minimum separation condition for $C = 2.38$.*

To prove this theorem, the authors studied the following *dual problem*: Given spectral data F on a finite set $\Lambda \subseteq \mathbb{Z}^d$, find the solution(s) φ to

$$\max_{\varphi} \Big| \sum_{m \in \Lambda} \widehat{\varphi}(m) F(m) \Big| \quad \text{such that} \quad \|\varphi\|_\infty \leq 1 \quad \text{and} \quad \operatorname{supp}(\widehat{\varphi}) \subseteq \Lambda. \tag{TV$'$}$$

Here, $\| \cdot \|_\infty$ denotes the usual $L^\infty(\mathbb{T}^d)$ norm. We call its solutions *dual polynomials*, and we let $\mathcal{D}(F, \Lambda)$ denote the set of dual polynomials.

While this viewpoint provides a satisfactory theoretical result, the computational aspect requires overcoming additional challenges. For $d = 1$, the authors proposed a *semi-definite program* (SDP) formulation of (TV′), see [12, Corollary 4.1], which relies on a spectral factorization theorem [20, Theorem 4.24].

Algorithm 1.3. Suppose μ and Λ satisfy the assumptions in Theorem 1.2a.

(1) Compute a polynomial φ using the SDP reformulation of the dual problem.
(2) If $|\varphi| \not\equiv 1$, then $\operatorname{supp}(\mu) \subseteq S = \{x \in \mathbb{T} \colon |\varphi(x)| = 1\}$ consists of at most $2M$ elements.
(3) Locate S using a root finding method.
(4) Invert a linear system in order to solve for the coefficients of μ.

Candès and Fernandez-Granda pointed out that this method is incomplete due to the assumption in the second step of Algorithm 1.3. We emphasize that this can occur even if μ satisfies any of the first three conditions in Theorem 1.2.

Further, while the SDP reformation does not hold for $d \geq 2$, there are more involved iterative primal-dual algorithms that can be substituted for the first step of Algorithm 1.3, e.g., see [45,18]. Alternatively, one could compute the solutions to (TV) directly without dealing with (TV′), and this would avoid some of the difficulties mentioned here, e.g., see [10], which uses a conditional gradient method.

We mention some related work on super-resolution, most of which adopt similar hypotheses as in Theorem 1.2. For example, [43] considered the recovery of discrete measures with random coefficients and a randomly selected low-frequency sampling set Λ. Again, under similar separation hypotheses, several papers addressed the noise situation, e.g., see [11,9,3,21,42]. By considering different sampling schemes and assumptions on the discrete measure, it is possible to study other "super-resolution" problems [17,35,34,2]. Finally, many these works were inspired by ℓ^1 minimization [15], compressed sensing [13,19], and convex geometry [14].

1.4. Our approach

Several previous papers have relied on the following well-known relationship: $\mu \in M(\mathbb{T}^d)$ is the *unique* solution to (TV) given spectral data F on Λ if and only if $F = \widehat{\mu}\,|_\Lambda$ and there exists $\varphi \in \mathcal{D}(F, \Lambda)$ such that $\varphi = \operatorname{sign}(\mu)$ ν-a.e. and $|\varphi(x)| < 1$ for all $x \notin \operatorname{supp}(\nu)$, e.g., see [12,21]. However, this characterization has not yielded useful results except in very special cases of μ and Λ, such as the one considered in Theorem 1.2. The technical reason is the following. Even if we reduce the space $\mathcal{D}(F,\Lambda)$ by identifying functions with the same global phase, in general, there is still no unique dual polynomial. This forces one to use generic tools, whereas not much is known about the sign of arbitrary trigonometric polynomials.

In order to obtain results that pertain to a richer class of measures, such as the ones discussed in Remark 1.1, we do not work with the aforementioned relationship. Instead, we develop a general theory that does not impose any conditions on μ and Λ. In fact, we study (TV) from an abstract, functional analysis perspective. Our methods are inspired by Beurling's work on minimal extrapolation [7,8]. (For recent results on non-uniform sampling, going back to [7] and balayage, see [1].) We later connect our results with the Candès and Fernandez-Granda theory on super-resolution [12].

We first introduce additional notation and adopt Beurling's language, since referring to the solutions of (TV) can be ambiguous when working with several different measures. We say $\nu \in M(\mathbb{T}^d)$ is an *extrapolation* of F on Λ if $F = \widehat{\nu}\,|_\Lambda$; and ν is a *minimal extrapolation* of F on Λ if it is a solution to (TV). The minimum value attained in (TV) is denoted by

$$\varepsilon = \varepsilon(F,\Lambda) = \inf\{\|\nu\|\colon \nu \in M(\mathbb{T}^d) \text{ and } F = \widehat{\nu} \text{ on } \Lambda\}.$$

While ε is in general unknown, we shall see that, for many important applications, it can be deduced. To this end, let

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

denote the set of all minimal extrapolations. Thus, for any $\nu \in \mathcal{E}(F,\Lambda)$, we have $\varepsilon(F,\Lambda) = \|\nu\|$. In order to understand the behavior of the minimal extrapolations, we study the set of dual polynomials, $\mathcal{D}(F,\Lambda)$. Both $\mathcal{E}(F,\Lambda)$ and $\mathcal{D}(F,\Lambda)$ are non-empty, see Proposition 2.1. It turns out that some of these dual polynomials can be characterized by the set,

$$\Gamma = \Gamma(F,\Lambda) = \{m \in \Lambda\colon |F(m)| = \varepsilon(F,\Lambda)\}.$$

This connection is made in Proposition 2.2.

The following is our main theorem.

Theorem 1.4. *Let F be spectral data on a finite set $\Lambda \subseteq \mathbb{Z}^d$.*

(a) *Suppose $\Gamma = \emptyset$. For any $\varphi \in \mathcal{D}(F,\Lambda)$, the closed set $S = \{x \in \mathbb{T}^d\colon |\varphi(x)| = 1\}$ has d-dimensional Lebesgue measure zero, and each minimal extrapolation of F on Λ is a singular measure supported in S.*
In particular, if $d = 1$, then S is a finite number of points and so each minimal extrapolation is a discrete measure supported in S.

(b) *Suppose $\#\Gamma \geq 2$. For each distinct pair $m,n \in \Gamma$, define $\alpha_{m,n} \in \mathbb{R}/\mathbb{Z}$ by $e^{2\pi i\alpha_{m,n}} = F(m)/F(n)$ and define the closed set,*

$$S = \bigcap_{\substack{m,n\in\Gamma \\ m\neq n}} \{x \in \mathbb{T}^d\colon x\cdot(m-n) + \alpha_{m,n} \in \mathbb{Z}\}, \tag{1.1}$$

which is an intersection of $\binom{\#\Gamma}{2}$ periodic hyperplanes. Then, each minimal extrapolation of F on Λ is a singular measure supported in S.
In particular, if $d = 1$, then S is a finite number of points and so each minimal extrapolation is a discrete measure supported in S.

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 $\mathbb{T}^d$.

The following topics encapsulate the contributions of this paper.

(a) *Qualitative behavior of minimal extrapolations.* It is interesting to note that Γ provides significant information about the minimal extrapolations. Theorem 1.4 shows that, regardless of the dimension, when $\#\Gamma = 0$ or $\#\Gamma \geq 2$, the minimal extrapolations are always singular measures. The $\#\Gamma = 1$ case is pathological. Proposition 2.8 shows that this scenario is connected with the existence of positive absolutely continuous minimal extrapolations. Example 3.2 shows that there can exist both *uncountably* many discrete, as well as positive absolutely continuous, minimal extrapolations. Hence, this behavior is a fundamental feature of (TV) and is not an artifact of our analysis.

(b) *Computational consequences.* One important aspect of Theorem 1.4 is its relationship with Algorithm 1.3 and similar variations. Proposition 2.6 shows that the algorithm fails precisely when $\varepsilon = \|F\|_{\ell^\infty(\Lambda)}$ and $\Gamma \neq \emptyset$. Since we are given the values of F on Λ, this immediately tells us what ε and Γ are. If $\#\Gamma \geq 2$, then the theorem is applicable and the minimal extrapolations are singular measures supported in the set defined in (1.1), which can be explicitly computed. We show how to apply our theorem to compute pertinent analytical examples in Section 3. Hence, our theorem is applicable even when Algorithm 1.3 and similar variations fail.

(c) *Super-resolution of singular measures.* When $d \geq 2$, Theorem 1.4 suggests that some singular continuous measures could be solutions to the super-resolution problem. Somewhat surprisingly, this is indeed the case because Example 3.8 provides a singular continuous minimal extrapolation when $\#\Gamma \geq 2$. This also demonstrates that the conclusion of Theorem 1.4b is optimal. Proposition 2.12 shows that surface measures corresponding to the zero set of trigonometric polynomials are also minimal extrapolations.

(d) *Impossibility of super-resolution.* Theorem 1.4 does not require structural assumptions on the finite subset $\Lambda \subseteq \mathbb{Z}^d$ or on the measure $\mu \in M(\mathbb{T}^d)$ that generates the spectral data F on Λ. Since the theorem also describes the support set of the minimal extrapolations, it is useful for determining whether it is possible for a given μ to be a minimal extrapolation. To illustrate this point, in Example 3.6, we provide a simple proof that shows, in general, a minimal separation condition is necessary to super-resolve a discrete measure.

(e) *Discrete-continuous correspondence.* Total variation minimization can be viewed as a continuous analogue of the basis pursuit algorithm [15]. Let $\mathcal{F}_N \in \mathbb{C}^{N\times N}$ be the DFT matrix, $x \in \mathbb{C}^N$ be an unknown vector, and $\Lambda \subseteq \{0, 1, \ldots, N-1\}$. The goal is to recover x from $y = \mathcal{F}_N x\,|_\Lambda$ by solving

$$\min_{\widetilde{x}\in\mathbb{C}^N} \|\widetilde{x}\|_{\ell^1} \quad \text{such that} \quad y = \mathcal{F}_N \widetilde{x} \quad \text{on } \Lambda. \tag{BP}$$

This technique was further studied and popularized by compressed sensing [13,19].
By considering the case that $\mu = \sum_{n=1}^N x_n \delta_{n/N}$, it is straightforward to see that (TV) is a generalization of (BP). From this point of view, Theorem 1.4 is a discrete-continuous correspondence result. Indeed, if $\#\Gamma \geq 2$, and either $d = 1$ or the vectors $\{p_j\}$ exist, then the minimal extrapolations are necessarily discrete measures whose support lie on a lattice; as we just discussed, such measures can be identified with vectors that are solutions to the discrete problem.

(f) *Mathematical connections.* Our results are closely related to Beurling's work on minimal extrapolation, and we make this relationship precise in Section 2.6. Proposition 2.4 connects our results to the Candès and Fernandez-Granda theory [12] by introducing a concept called an admissibility range for ε. This connection is exploited to prove Proposition 2.6, which in turn, shows that our theorem is applicable to situations where Algorithm 1.3 fails.

1.5. Outline

Section 2 contains the basis of our mathematical theory. In Section 2.1, we prove well-known results about (TV) and (TV′) using only basic functional analysis, and we characterize the set Γ. In Section 2.2, we prove the main result, Theorem 1.4. Section 2.3 contains the material that fully connects the Beurling and Candès and Fernandez-Granda theories, as well as the relationship between Theorem 1.4 and Algorithm 1.3. Section 2.4 includes a basic uniqueness result, and the aforementioned non-uniqueness behavior for the case $\#\Gamma = 1$, Proposition 2.8. Section 2.5 examines whether it is possible to relate the minimal extrapolations of μ with those of $T\mu$ for various operators T. Section 2.6 discusses the relationship between this paper with Beurling's theory of minimal extrapolation. Section 3 contains all of the examples that we have previously mentioned, including the minimal extrapolations of discrete measures and of singular continuous measures.

2. Mathematical theory

2.1. Preliminary results

This subsection contains basic facts about (TV) and (TV′). In contrast to previous approaches relying on convex analysis, e.g., see [21], our proofs only use basic functional analysis, which also illustrates its important role.

Proposition 2.1 contains the main functional analysis results. 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 [5, 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

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

(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

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

Proposition 2.1a shows that (TV) is well-posed, by using standard functional analysis arguments. 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 the minimal extrapolations only depend on F on Λ, and not the unknown measure μ that generates F, we consider the subspace

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

Proposition 2.1b 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 2.1c demonstrates that

Please cite this article in press as: J.J. Benedetto, W. Li, Super-resolution by means of Beurling minimal extrapolation, Appl. Comput. Harmon. Anal. (2018), https://doi.org/10.1016/j.acha.2018.05.002

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

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 the unknown $\mu \in M(\mathbb{T}^d)$ 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 2.1d, e show that we have the stronger statement,

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

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

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

See [5, Theorem 5.3.2, page 242, and Theorem 5.3.5, page 244] for further details. Recall that the support of $\mu \in M(\mathbb{T}^d)$, $\operatorname{supp}(\mu)$, is the complement of all open sets $A \subseteq \mathbb{T}^d$ such that $\mu(A) = 0$.

Proposition 2.1. *Let F be spectral data on a finite set $\Lambda \subseteq \mathbb{Z}^d$, where $F = \widehat{\mu}\,|_\Lambda$ for some $\mu \in M(\mathbb{T}^d)$.*

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

Proof. (a) By definition of ε, there exists a sequence $\{\nu_j\}$ such that $\|\nu_j\| \to \varepsilon$ and $\widehat{\nu_j} = F$ on Λ. 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 each $m \in \Lambda$, we have

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

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, $\|\nu_t\| \geq \varepsilon$. By the triangle inequality, we have $\|\nu_t\| \leq (1-t)\|\nu_0\| + t\|\nu_1\| = \varepsilon$. 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, $\|f_j\|_\infty \leq 1$, 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 $|a_{j,m}| \leq \|f_j\|_\infty \leq 1$. Then, for any $x, y \in \mathbb{T}^d$, we have

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

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

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

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 $\|f\|_\infty \leq 1$, 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,

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

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

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 $\|\sigma\| \geq \varepsilon$. Combining these facts, we have

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

This proves the lower bound.

(e) We know that $\|L_\mu\| = \varepsilon$. By definition, there exists $\{f_j\} \subseteq U$ such that $|L_\mu(f_j)| \geq \varepsilon - 1/j$. 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 $|L_\mu(f)| \leq \|f\|_\infty \varepsilon \leq \varepsilon$. For the reverse inequality, as $k \to \infty$,

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

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

(f) By the previous part and a simple calculation,

$$\varepsilon = \max_{f\in U} |\langle f, \mu\rangle| = \max_{f\in U} \Big|\sum_{m\in\Lambda} \widehat{f}(m)\overline{F(m)}\Big| = \max_{f\in U} \Big|\sum_{m\in\Lambda} \widehat{f}(m)F(m)\Big|.$$

Hence, there exists a $\varphi \in U$ that attains this maximum, which by definition, is a dual polynomial. $\square$

Proposition 2.2a, b establish *uniform* statements: Even if there are several minimal extrapolations, they must all be supported in a common set and they must have similar sign patterns. Proposition 2.2c characterizes the case that $|\varphi| \equiv 1$ in terms of the set Γ. Intuitively, $\#\Gamma$ represents the number of "bad" dual polynomials. While it is desirable to have $\Gamma = \emptyset$, perhaps surprisingly, Proposition 2.2d shows that we can make strong statements even when $\#\Gamma$ is large.

Proposition 2.2. *Let F be spectral data on a finite set $\Lambda \subseteq \mathbb{Z}^d$, where $F = \widehat{\mu}\,|_\Lambda$ for some $\mu \in M(\mathbb{T}^d)$.*

(a) *For any $\varphi \in \mathcal{D}(F,\Lambda)$, we have*

$$\forall \nu \in \mathcal{E}(F,\Lambda), \quad \mathit{supp}(\nu) \subseteq \{x \in \mathbb{T}^d \colon |\varphi(x)| = 1\}.$$

(b) *There exists $\varphi \in \mathcal{D}(F,\Lambda)$ such that*

$$\forall \nu \in \mathcal{E}(F,\Lambda), \quad \varphi = \mathit{sign}(\nu) \quad \nu\text{-a.e.}$$

(c) *There exists $\varphi \in \mathcal{D}(F,\Lambda)$ with $|\varphi| \equiv 1$ if and only if $\Gamma \neq \emptyset$.*

(d) *For each $m \in \mathbb{Z}^d$, define $\alpha_m \in \mathbb{R}/\mathbb{Z}$ by the formula $e^{-2\pi i \alpha_m} F(m) = |F(m)|$. If $m \in \Gamma$, then*

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

Proof. (a) Using that $\varphi \in U(\mathbb{T}^d;\Lambda)$ and Proposition 2.1e, for all $\nu \in \mathcal{E}$, we have

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

Since $\|\varphi\|_\infty \le 1$ and $|\langle \varphi,\nu\rangle| = \|\nu\|$, there exists $\theta \in \mathbb{R}$ such that $\varphi = e^{2\pi i\theta}\mathrm{sign}(\nu)$ ν-a.e. Using that $|\varphi| = |\mathrm{sign}(\nu)| = 1$ ν-a.e. and that ν is a Radon measure,

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

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

(b) From the previous part, we know there exists a $f \in \mathcal{D}(F,\Lambda)$ such that $|\langle f,\mu\rangle| = \varepsilon$. Then, there exists $\theta \in \mathbb{R}$ such that $e^{2\pi i\theta}\langle f,\mu\rangle = \varepsilon$. Define the function $\varphi = e^{2\pi i\theta} f$, which is also a dual polynomial. Repeating the same argument from the previous part, we see that for all $\nu \in \mathcal{E}$, $\varphi = \mathrm{sign}(\nu)$ ν-a.e.

(c) If $\varphi \in \mathcal{D}(F,\Lambda)$ and $|\varphi| \equiv 1$, then $\varphi = e^{2\pi i\theta}e^{2\pi i m\cdot x}$ for some $\theta \in \mathbb{R}$ and $m \in \Lambda$. Then, we have $\varepsilon = |\langle \varphi,\mu\rangle| = |F(m)|$ which shows that $m \in \Gamma$.
Conversely, let $m \in \Gamma$. Then we readily check that $\varphi(x) = e^{2\pi i m\cdot x}$ is a dual polynomial and $|\varphi| \equiv 1$.

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

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

This shows $\mathrm{sign}(\nu)(x) = e^{2\pi i\alpha_m} e^{2\pi i m\cdot x}$ ν-a.e. $\square$

Please cite this article in press as: J.J. Benedetto, W. Li, Super-resolution by means of Beurling minimal extrapolation, Appl. Comput. Harmon. Anal. (2018), https://doi.org/10.1016/j.acha.2018.05.002

2.2. An analogue of Beurling's theorem

We are ready to prove our main theorem.

Proof of Theorem 1.4. (a) Let $\varphi \in \mathcal{D}(F,\Lambda)$. By Proposition 2.2a, c, $|\varphi| \not\equiv 1$ because $\Gamma = \emptyset$ and each minimal extrapolation is supported in the closed set

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

Consider the function

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

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 2.2d, such that

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

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 $e^{2\pi i\alpha_{m,n}} = F(m)/F(n)$ because

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

Suppose $\{p_1,\dots,p_d\}$ satisfies the hypothesis. By the support assertion that we just proved, there exists $\beta = (\beta_1,\beta_2,\dots,\beta_d) \in \mathbb{T}^d$ 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, $p_1, p_2, \dots, p_d$, 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 Fig. 2.1 for an illustration.

Please cite this article in press as: J.J. Benedetto, W. Li, Super-resolution by means of Beurling minimal extrapolation, Appl. Comput. Harmon. Anal. (2018), https://doi.org/10.1016/j.acha.2018.05.002

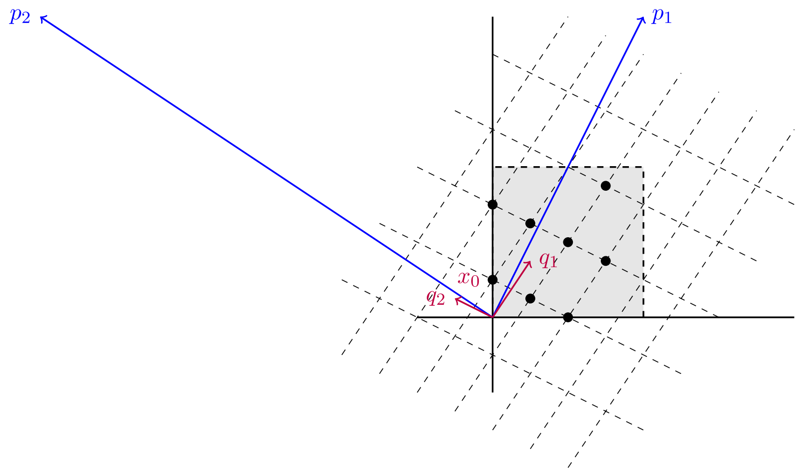

Fig. 2.1. An illustration of Theorem 1.4b, where $d = 2$, $p_1 = (1, 2)$, $p_2 = (-3, 2)$, $\beta_1 = 1/2$, $\beta_2 = -1/2$, $q_1 = (1/4, 3/8)$, and $q_2 = (-1/4, 1/8)$. The family of hyperplanes are the dashed lines, the lattice S is the black dots, and the shaded region is $[0, 1)^2$.

For a rigorous proof of this observation, first note that

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

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, $q_1, q_2, \ldots, q_d$. $\square$

Remark 2.3. Example 3.8 provides an example where $\#\Gamma = 3$ and there exists a singular continuous minimal extrapolation. This demonstrates that the conclusion of Theorem 1.4b is optimal.

2.3. The admissibility range

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 in general, ε is unknown. However, in many important situations, it is possible to deduce the value of ε. We say $[A, B] \subseteq \mathbb{R}^+$ is an *admissibility range* for ε provided that $0 \le A \le \varepsilon \le B$. The following proposition shows that we have $A = \|F\|_{\ell^\infty(\Lambda)}$ and $B = \|\mu\|$ as an admissibility range, and B can be improved in certain situations. While the proof is elementary and is a direct consequence of Hölder's inequality, we shall use this proposition throughout the remainder of this paper; and this also demonstrates its centrality and importance in our theory.

Proposition 2.4. *Let F be spectral data on a finite set $\Lambda \subseteq \mathbb{Z}^d$, where $F = \widehat{\mu}\,|_\Lambda$ for some $\mu \in M(\mathbb{T}^d)$. We have the lower and upper bounds,*

$$\|F\|_{\ell^\infty(\Lambda)} \le \varepsilon \le \|\mu\|. \tag{2.1}$$

Further, if there exists an extrapolation $\nu \in M(\mathbb{T}^d)$ and $\|\nu\| < \|\mu\|$, then

$$\|F\|_{\ell^\infty(\Lambda)} \le \varepsilon \le \|\nu\| < \|\mu\|. \tag{2.2}$$

Please cite this article in press as: J.J. Benedetto, W. Li, Super-resolution by means of Beurling minimal extrapolation, Appl. Comput. Harmon. Anal. (2018), https://doi.org/10.1016/j.acha.2018.05.002

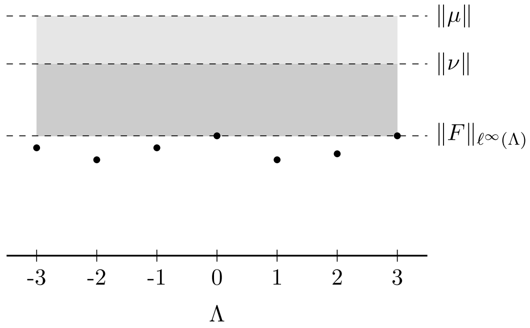

Fig. 2.2. The black points represent the values of $|F|$ on Λ. The union of the light and dark gray regions is the admissibility range $[\|F\|_{\ell^\infty(\Lambda)}, \|\mu\|]$, whereas the dark gray region is an improved admissibility range $[\|F\|_{\ell^\infty(\Lambda)}, \|\nu\|]$.

Proof. To see the lower bound for ε in (2.1) and (2.2), let σ be a minimal extrapolation of F on Λ. Then,

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

The upper bounds, $\varepsilon \le \|\mu\|$ and $\varepsilon \le \|\nu\|$ in (2.1) and (2.2), follow by definition of ε. $\square$

Remark 2.5. By tightening the admissibility range for ε, we can deduce information about the minimal extrapolations. The simplest case is when $\varepsilon = \|F\|_{\ell^\infty(\Lambda)} = \|\mu\|$, see Examples 3.2 and 3.3. The next simplest case is when there exists an extrapolation ν, such that $\varepsilon = \|F\|_{\ell^\infty(\Lambda)} = \|\nu\|$, see Example 3.4. A more complicated case is when $\|F\|_{\ell^\infty(\Lambda)} < \varepsilon$, see Example 3.5. See Fig. 2.2 for an illustration.

Candès and Fernandez-Granda [12] focused entirely on the case that $\|\mu\| = \varepsilon$ because this is a necessary condition to recover μ using (TV). In contrast to their result, our theory is better suited for situations when $\|F\|_{\ell^\infty(\Lambda)} = \varepsilon$. This is because Theorem 1.4b is strongest when $|F(m)| = \varepsilon$ for many points $m \in \Lambda$, i.e., when $\#\Gamma$ is large. Hence, our theory is better suited for deducing the impossibility of reconstruction using (TV).

Algorithm 1.3 fails precisely when it computes a dual polynomial φ such that $|\varphi| \equiv 1$. The following proposition shows that this occurs precisely when $\varepsilon = \|F\|_{\ell^\infty}$. Since we are given F on Λ, we then immediately get the exact value of ε.

Proposition 2.6. *Let F be spectral data on a finite set $\Lambda \subseteq \mathbb{Z}^d$, where $F = \widehat{\mu}\,|_\Lambda$ for some $\mu \in M(\mathbb{T}^d)$. Suppose there exists $\varphi \in \mathcal{D}(F,\Lambda)$ such that $|\varphi| \equiv 1$. Then, $\Gamma \ne \emptyset$ and $\varepsilon = \|F\|_{\ell^\infty(\Lambda)}$.*

Proof. By Proposition 2.2c, we must have $\Gamma \ne \emptyset$. Let $m \in \Gamma$ and by definition of Γ, we have $|F(m)| = \varepsilon$. This combined with the lower bound in Proposition 2.4 proves that $\varepsilon = \|F\|_{\ell^\infty(\Lambda)}$. $\square$

2.4. On uniqueness and non-uniqueness of minimal extrapolations

Theorem 1.4 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. The following proposition is well-known, see, e.g., [17] for a similar result.

Proposition 2.7. *Let F be spectral data on a finite set $\Lambda = \{m_1, m_2, \ldots, m_J\} \subseteq \mathbb{Z}^d$, where $F = \widehat{\mu}\,|_\Lambda$ for some $\mu \in M(\mathbb{T}^d)$. Suppose there exists a finite set, $\{x_k\}_{k=1}^K \subseteq \mathbb{T}^d$, such that each minimal extrapolation of F on Λ is supported in this set. Suppose that the matrix,*

Please cite this article in press as: J.J. Benedetto, W. Li, Super-resolution by means of Beurling minimal extrapolation, Appl. Comput. Harmon. Anal. (2018), https://doi.org/10.1016/j.acha.2018.05.002

$$E(m_1,\dots,m_J;x_1,\dots,x_K)=\begin{pmatrix} e^{-2\pi i m_1\cdot x_1} & e^{-2\pi i m_1\cdot x_2} & \dots & e^{-2\pi i m_1\cdot x_K}\\ e^{-2\pi i m_2\cdot x_1} & e^{-2\pi i m_2\cdot x_2} & \dots & e^{-2\pi i m_2\cdot x_K}\\ \vdots & \vdots & & \vdots\\ e^{-2\pi i m_J\cdot x_1} & e^{-2\pi i m_J\cdot x_2} & \dots & e^{-2\pi i m_J\cdot x_K}\end{pmatrix}, \tag{2.3}$$

has full column rank (this can only occur if $J \geq K$). Then, there is a unique minimal extrapolation of F on Λ.

Proof. Let ν be the difference of any two minimal extrapolations of F on Λ. Then ν is also supported in $\{x_k\}_{k=1}^K$ 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 $j=1,2,\dots,J$, which is equivalent to the linear system $Ea=0$, where $a=(a_1,\dots,a_K)\in\mathbb{C}^K$. By assumption, E has full column rank. This implies $a=0$. □

We address the situation when $\#\Gamma=1$, i.e., the missing case of Theorem 1.4. A measure $\mu\in M(\mathbb{T}^d)$ is *positive* if $\mu(A)\geq 0$ for all Borel sets $A\subseteq\mathbb{T}^d$. A sequence $\{a_m\}_{m\in\mathbb{Z}^d}$ is *positive-definite* if for all sequences $\{b_m\}_{m\in\mathbb{Z}^d}$ of finite support, we have

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

Bochner's theorem, that is true for all locally compact abelian groups [40], asserts in our case of $\mathbb{T}^d$ and $\mathbb{Z}^d$ that a sequence is positive-definite if and only if it is the sequence of Fourier coefficients of a positive measure. In this special case of $\mathbb{T}^d$ and $\mathbb{Z}^d$, Bochner's theorem is called Herglotz' theorem. Herglotz proved it in 1911 for $d=1$, and natural modifications to the proof in [30, page 39] yield the statement in higher dimensions.

Proposition 2.8. *Let F be spectral data on a finite set $\Lambda\subseteq\mathbb{Z}^d$. Suppose there exists $n\in\Lambda$ such that $\{F(m+n)\}_{m\in\Lambda-n}$ extends to a positive-definite sequence on $\mathbb{Z}^d$. Then, $n\in\Gamma$, and each positive-definite extension of $\{F(m+n)\}_{m\in\Lambda-n}$ corresponds to a positive measure ν, such that $e^{2\pi i n\cdot x}\nu(x)$ is a minimal extrapolation of F on Λ.*

Proof. Extend $\{F(m+n)\}_{m\in\Lambda-n}$ to a positive-definite sequence on $\mathbb{Z}^d$. By Herglotz' theorem, there exists a positive measure $\nu\in M(\mathbb{T}^d)$ such that $\widehat{\nu}(m)=F(m+n)$ for all $m\in\Lambda-n$. Then, $\sigma(x)=e^{2\pi i n\cdot x}\nu(x)$ is an extrapolation of F on Λ, which implies

$$\|\nu\|=\|\sigma\|\geq\varepsilon.$$

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

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

where the last inequality follows by Proposition 2.4. This shows that $|F(n)|=\|\nu\|=\varepsilon$, which proves that $n\in\Gamma$ and $M_n\nu$ is a minimal extrapolation of F on Λ. □

Remark 2.9. Beurling [8] and Esseen [22] essentially proved the analogue of Proposition 2.8 for the special case that $n=0$, and for $\mathbb{R}$ instead of $\mathbb{T}^d$. Proposition 2.8 generalizes their result to handle situations when $0\notin\Lambda$. This is important because from the viewpoint of Proposition 2.13c, the super-resolution problem is invariant under simultaneous translations of F and Λ, which means that $0\in\mathbb{Z}^d$ is no more special than any other point $n\in\mathbb{Z}^d$.

Remark 2.10. Proposition 2.8 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 3.2, $\#\Gamma = 1$, and there exist uncountably many discrete and positive absolutely continuous minimal extrapolations.

Remark 2.11. Suppose that F 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 F to the positive-definite sequence $\{a_m\}_{m\in\mathbb{Z}}$ 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 [4, Theorems 3.6.3 and 3.6.6]. MEM is related to spectral estimation methods [16], the maximum likelihood method [16], and moment problems [32].

Finally, we show that the surface measure of the zero set of a trigonometric polynomial is a minimal extrapolation. If $A, B \subseteq \mathbb{Z}^d$, let $A - B = \{a - b \colon a \in A,\ b \in B\}$.

Proposition 2.12. *Let $\Lambda \subseteq \mathbb{Z}^d$ be a finite set and $\varphi \in C(\mathbb{T}^d; \Lambda)$ such that $S = \{x \in \mathbb{T}^d \colon \varphi(x) = 0\}$ is non-trivial. Let F be spectral data on $\Lambda - \Lambda$, where $F = \widehat{\sigma}\,|_{\Lambda-\Lambda}$ and σ is the surface measure of S. Then, $\Gamma \neq \emptyset$, σ is a positive minimal extrapolation of F on $\Lambda - \Lambda$, and there exists $\Phi \in \mathcal{D}(F, \Lambda - \Lambda)$ such that $S = \{x \in \mathbb{T}^d \colon |\Phi(x)| = 1\}$.*

Proof. Of course, σ is an extrapolation of F on $\Lambda - \Lambda$, so we have $\|\sigma\| \geq \varepsilon$. To prove the reverse inequality, we observe that σ is a positive measure and by Proposition 2.4,

$$\|\sigma\| = |\widehat{\sigma}(0)| = |F(0)| \leq \|F\|_{\ell^\infty(\Lambda-\Lambda)} \leq \varepsilon(F, \Lambda - \Lambda).$$

This also shows that $0 \in \Gamma$.

Since S is invariant if we multiply φ by a constant, without loss of generality, assume that $\|\varphi\|_\infty \leq 1$. Define the function $\Phi = 1 - |\varphi|^2$, and note that $\Phi \in U(\mathbb{T}^d; \Lambda - \Lambda)$ and $S = \{x \in \mathbb{T}^d \colon \Phi(x) = 1\}$. Using these facts, we have $|\langle \Phi, \sigma \rangle| = \|\sigma\| = \varepsilon$. This shows that Φ is a dual polynomial. $\square$

2.5. 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 numbers. We are also interested in the operations that are well-behaved under the Fourier transform on $\mathbb{T}^d$, namely, translation, modulation, convolution, and product of measures. For any $y \in \mathbb{R}^d$, let M_y be the modulation operator defined by $M_y f(x) = e^{2\pi i y\cdot x} f(x)$ and let T_y be the translation operator $T_y f(x) = f(x - y)$.

Proposition 2.13. *Let F be spectral data on a finite set $\Lambda \subseteq \mathbb{Z}^d$. Let $a \in \mathbb{C}$ be non-zero, $n \in \mathbb{Z}^d$, and $y \in \mathbb{R}^d$.*

(a) *Multiplication by constants is bijective: $a\varepsilon(F, \Lambda) = \varepsilon(aF, \Lambda)$, and $\nu \in \mathcal{E}(F, \Lambda)$ if and only if $a\nu \in \mathcal{E}(aF, \Lambda)$.*
(b) *Translation is bijective: $\varepsilon(\mu, \Lambda) = \varepsilon(M_y F, \Lambda)$, and $\nu \in \mathcal{E}(F, \Lambda)$ if and only if $T_y\nu \in \mathcal{E}(M_y F, \Lambda)$.*
(c) *Minimal extrapolation is invariant under simultaneous shifts of F and Λ: $\varepsilon(F, \Lambda) = \varepsilon(T_n F, \Lambda + n)$, and $\mathcal{E}(\mu, \Lambda) = \mathcal{E}(T_n F, \Lambda + n)$.*
(d) *The product of minimal extrapolations is a minimal extrapolation for the product: For $j = 1, 2$, let F_j be spectral data on a finite set $\Lambda_j \subseteq \mathbb{Z}^{d_j}$ and let $\nu_j \in \mathcal{E}(F_j, \Lambda_j)$. Then $\varepsilon(F_1, \Lambda_1)\varepsilon(F_2, \Lambda_2) = \varepsilon(F_1 \otimes F_2, \Lambda_1 \times \Lambda_2)$, and $\nu_1 \times \nu_2 \in \mathcal{E}(F_1 \otimes F_2, \Lambda_1 \times \Lambda_2)$.*

Please cite this article in press as: J.J. Benedetto, W. Li, Super-resolution by means of Beurling minimal extrapolation, Appl. Comput. Harmon. Anal. (2018), https://doi.org/10.1016/j.acha.2018.05.002

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

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

(c) If $f \in U(\mathbb{T}^d;\Lambda)$, then $M_nf \in U(\mathbb{T}^d;\Lambda+n)$. We have,

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

Using Proposition 2.1e, we see that $\varepsilon(F,\Lambda) = \varepsilon(T_nF,\Lambda+n)$.

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

(d) For convenience, let $\mu = \mu_1\times\mu_2$, $\nu = \nu_1\times\nu_2$, $\Lambda = \Lambda_1\times\Lambda_2$, $F = F_1\otimes F_2$, $\varepsilon_j = \varepsilon(F_j,\Lambda_j)$, and $\varepsilon = \varepsilon(F,\Lambda)$. Since ν is an extrapolation of F on Λ, by Proposition 2.4, we have

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

To see the reverse inequality, by Proposition 2.1f, there exist $\varphi_j \in \mathcal{D}(F_j,\Lambda)$, for $j = 1,2$. Let $\varphi = \varphi_1\otimes\varphi_2$ and observe that $\varphi \in U(\mathbb{T}^d;\Lambda)$. By Proposition 2.1e, we have

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

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

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 2.14.

Proposition 2.14.

(a) *For $j = 1,2$, there exist spectral data F_j on a finite subset $\Lambda \subseteq \mathbb{Z}$ and $\nu_j \in \mathcal{E}(F_j,\Lambda)$, such that $\nu_1+\nu_2 \notin \mathcal{E}(F_1+F_2,\Lambda)$.*

(b) *There exist spectral data F on a finite subset $\Lambda \subseteq \mathbb{Z}$, $\nu \in \mathcal{E}(F,\Lambda)$, and $n \in \mathbb{Z}$, such that $M_n\nu \notin \mathcal{E}(T_nF,\Lambda)$.*

(c) *For $j = 1,2$, there exist spectral data F_j on a finite subset $\Lambda \subseteq \mathbb{Z}$, and $\nu_j \in \mathcal{E}(F_j,\Lambda)$, such that $\nu_1 * \nu_2 \notin \mathcal{E}(F_1F_2,\Lambda)$.*

Proof. (a) Let $\mu_1 = \delta_0+\delta_{1/2}$, $\mu_2 = -\delta_0-\delta_{1/2}$, $\Lambda = \{-1,0,1\}$, and $F_j = \widehat{\mu_j}\,|_\Lambda$. By Example 3.2, we have $\nu_1 = \delta_0+\delta_{1/2} \in \mathcal{E}(F_1,\Lambda)$, and $\nu_2 = -\delta_{1/4}-\delta_{3/4} \in \mathcal{E}(F_2,\Lambda)$. Then, $\mu_1+\mu_2 = 0$, and so $\varepsilon(F_1+F_2,\Lambda) = 0$. However, $\nu_1+\nu_2 \notin \mathcal{E}(F_1+F_2,\Lambda)$ because $\|\nu_1+\nu_2\| = \|\delta_0-\delta_{1/4}+\delta_{1/2}-\delta_{3/4}\| > 0$.

(b) Let $\mu = \delta_0+\delta_{1/2}$, $\Lambda = \{-1,0,1\}$, $n = -1$, and $F = \widehat{\mu}\,|_\Lambda$. By Example 3.2, we have $\nu = \delta_{1/4} + \delta_{3/2} \in \mathcal{E}(F,\Lambda)$. However, $M_{-1}\mu = \delta_0-\delta_{1/2}$, and by Example 3.3, $\mathcal{E}(T_1F,\Lambda) = \{\delta_0-\delta_{1/2}\}$. Thus, $M_{-1}\nu \notin \mathcal{E}(T_1F,\Lambda)$.

(c) Let $\mu_1 = \delta_0+\delta_{1/2}$, $\mu_2 = \delta_0-\delta_{1/2}$, $\Lambda = \{-1,0,1\}$, and $F_j = \widehat{\mu_j}\,|_\Lambda$. Then $F_1F_2 = 0$ on Λ, which implies $\varepsilon(F_1F_2,\Lambda) = 0$. By Examples 3.2 and 3.3, we have $\nu_1 = \mu_1 \in \mathcal{E}(F_1,\Lambda)$ and $\nu_2 = \mu_2 \in \mathcal{E}(F_2,\Lambda)$. However, $\nu_1*\nu_2 \notin \mathcal{E}(F_1F_2,\Lambda)$ because $\|\nu_1*\nu_2\| = \|\delta_0-\delta_1\| > 0$. □

Please cite this article in press as: J.J. Benedetto, W. Li, Super-resolution by means of Beurling minimal extrapolation, Appl. Comput. Harmon. Anal. (2018), https://doi.org/10.1016/j.acha.2018.05.002

2.6. Connection to Beurling's theorem

We explain why Theorem 1.4 is an adaptation to the torus and a generalization to higher dimensions of Beurling's theorem. Let $M_b(\mathbb{R})$ be the space of complex Radon measures on $\mathbb{R}$ with finite total variation norm. Since this is the only part of this paper that deals with measures on $\mathbb{R}$, we slightly abuse notation, and we denote the total variation norm on $\mathbb{R}$ by $\|\cdot\|$ and the Fourier transform of $\mu \in M_b(\mathbb{R})$ by $\widehat{\mu}\colon \mathbb{R} \to \mathbb{C}$.

Suppose we are given the set $\Lambda = [-\lambda, \lambda]$ and the spectral data $G = \widehat{\mu}\,|_\Lambda$ for some $\mu \in M_b(\mathbb{R})$. Beurling studied the total variation problem on $\mathbb{R}$ of finding solution(s) ν to

$$\inf_\nu \|\nu\| \quad \text{such that} \quad \nu \in M_b(\mathbb{R}) \quad \text{and} \quad G = \widehat{\nu} \quad \text{on } \Lambda. \tag{2.4}$$

The solutions to this problem are called minimal extrapolations of G on Λ. He defined the quantities

$$\begin{aligned} m &= \inf\{\|\nu\|\colon \nu \in M_b(\mathbb{R}) \text{ and } G = \widehat{\nu} \text{ on } \Lambda\}, \\ M &= \{\nu \in M_b(\mathbb{R})\colon G = \widehat{\nu} \text{ on } \Lambda \text{ and } \|\nu\| = m\}, \\ \Lambda_m &= \{\gamma \in \Lambda\colon |G(\gamma)| = m\}. \end{aligned}$$

Theorem 2.15 *(Beurling, Theorem 2, page 362, [8]). Let G be spectral data on $\Lambda = [-\lambda, \lambda]$, where $G = \widehat{\mu}\,|_\Lambda$ for some $\mu \in M_b(\mathbb{R})$.*

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

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

is the unique minimal extrapolation of G on 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 G on Λ_m.

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

Remark 2.16. It is difficult to deduce information about the minimal extrapolations when $\#\Lambda_m = 1$ because they might not be unique and there may exist positive absolutely continuous minimal extrapolations, e.g., see [8,22,23,27] for specific examples and related work.

Remark 2.17. Beurling relied heavily on complex analysis and did not provide a higher dimensional version of Theorem 2.15. In contrast, we avoided the use of complex analysis in our proofs of Theorem 1.4 and the propositions leading to it. The extension to higher dimensions is not trivial, since as we saw, geometry plays a major role in $d \geq 2$. There are two important advantages of working with $\mathbb{T}^d$ as opposed to $\mathbb{R}$ from an application point of view. First, it is reasonable to assume that measures encountered in applications are compactly supported, and thus, their supports can be normalized to be the unit cube. Second, it is not clear how to solve (2.4) numerically, whereas there are algorithms for solving (TV).

Please cite this article in press as: J.J. Benedetto, W. Li, Super-resolution by means of Beurling minimal extrapolation, Appl. Comput. Harmon. Anal. (2018), https://doi.org/10.1016/j.acha.2018.05.002

3. Examples

3.1. 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 1.4, the minimal extrapolations of a non-discrete measure are sometimes discrete measures, so they appear naturally in our analysis.

As discussed in Section 1.4, examples of μ that have minimal extrapolations supported in a lattice can be interpreted in the context of deterministic compressed sensing. Examples 3.2–3.5 and Example 3.8 can be written in the context of compressed sensing.

Remark 3.1. In view of Proposition 2.13a, 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 3.2. Let $\Lambda = \{-1, 0, 1\}$ and $F = \widehat{\mu}\,|_\Lambda$, where $\mu = \delta_0 + \delta_{1/2}$. We have $F(0) = 2$, and $F(\pm 1) = 0$. Clearly $\|F\|_{\ell^\infty(\Lambda)} = \|\mu\| = 2$. By Proposition 2.4, $\varepsilon = \|F\|_{\ell^\infty(\Lambda)} = \|\mu\| = 2$, which implies μ is a minimum extrapolation of F on Λ.

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,

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

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

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

Consider the real-valued function

$$\begin{aligned} f_{N,c}(x) &= 2 + \sum_{n=-N}^{-2} (a_{N,c})_n e^{2\pi i n x} + \sum_{n=2}^{N} (a_{N,c})_n e^{2\pi i n x} \\ &= 2 + c \sum_{n=-N}^{-2} \Big(1 - \frac{|n|}{N+1}\Big) e^{2\pi i n x} + c \sum_{n=2}^{N} \Big(1 - \frac{|n|}{N+1}\Big) e^{2\pi i n x}. \end{aligned}$$

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

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

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

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

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 3.3. Let $\Lambda = \{-1, 0, 1\}$ and $F = \widehat{\mu}\,|_\Lambda$, where $\mu = \delta_0 - \delta_{1/2}$. We have $F(0) = 0$, and $F(\pm 1) = 2$. Further, we have $\|F\|_{\ell^\infty(\Lambda)} = \|\mu\| = 2$, so that by Proposition 2.4, we have $\varepsilon = \|F\|_{\ell^\infty(\Lambda)} = \|\mu\| = 2$. Thus, μ is a minimal extrapolation of F on Λ.

Consequently, $\Gamma = \{-1, 1\}$. By Theorem 1.4b, there exists $\alpha_{-1,1} \in \mathbb{R}/\mathbb{Z}$ satisfying

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

and the minimal extrapolations are supported in the set $\{x \in \mathbb{T}\colon 2x \in \mathbb{Z}\} = \{0, 1/2\}$. 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 (2.3) 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 2.7, μ is the unique minimal extrapolation of F on Λ. Thus, reconstruction of μ from F on Λ is possible.

Example 3.4. Let $\Lambda = \{-1, 0, 1\}$ and $F = \widehat{\mu}\,|_\Lambda$, where $\mu = \delta_0 - \delta_{1/4}$. We have $F(\pm 1) = 1 \pm i = \sqrt{2}e^{\pm \pi i/4}$, and $F(0) = 0$. Note that $\|F\|_{\ell^\infty(\Lambda)} = \sqrt{2} < 2 = \|\mu\|$, 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 of F on Λ. By Proposition 2.4, $\varepsilon = \sqrt{2}$, which implies ν is a minimal extrapolation of F on Λ. This also implies $\mu \notin \mathcal{E}(F, \Lambda)$, and so reconstruction of μ from F on Λ using (TV) is impossible.

We claim that ν is the unique minimal extrapolation. The matrix E from (2.3) is

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

which we observe to have full column rank. By Proposition 2.7, we conclude that ν is the unique minimal extrapolation of F on Λ. Thus, reconstruction of ν from F on Λ is possible.

We explain the derivation of ν. We guess that $\varepsilon = \sqrt{2}$ and see what Theorem 1.4b implies. Under this assumption that $\varepsilon = \sqrt{2}$, we have $\Gamma = \{-1, 1\}$. By Theorem 1.4b, there exists $\alpha_{-1,1} \in \mathbb{R}/\mathbb{Z}$ satisfying

Please cite this article in press as: J.J. Benedetto, W. Li, Super-resolution by means of Beurling minimal extrapolation, Appl. Comput. Harmon. Anal. (2018), https://doi.org/10.1016/j.acha.2018.05.002

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

and the minimal extrapolations are supported in $\{x \in \mathbb{T}\colon 2x + 1/4 \in \mathbb{Z}\} = \{3/8, 7/8\}$. Hence, if $\varepsilon = \sqrt{2}$, then every $\sigma \in \mathcal{E}(F, \Lambda)$ 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 $F = \widehat{\sigma}$ 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 of F on Λ.

Example 3.5. Let $\Lambda = \{-1, 0, 1\}$ and $F = \widehat{\mu}\,|_\Lambda$, where $\mu = \delta_0 + e^{\pi i/3}\delta_{1/3}$. We have $F(-1) = 0$, $F(0) = 1 + e^{\pi i/3} = \sqrt{3}\, e^{\pi i/6}$, and $F(1) = 1 + e^{-\pi i/3} = \sqrt{3}\, e^{-\pi i/6}$.

Suppose, for the purpose of obtaining a contradiction, that $\varepsilon = \|F\|_{\ell^\infty(\Lambda)} = \sqrt{3}$. Then $\Gamma = \{0, 1\}$. By Theorem 1.4b, there is $\alpha_{0,1} \in \mathbb{R}/\mathbb{Z}$ such that

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

and each $\nu \in \mathcal{E}(F, \Lambda)$ 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, $F \neq \widehat{\nu}$ on Λ, which is a contradiction.

Thus, $\varepsilon > \sqrt{3}$, i.e., $\Gamma = \emptyset$. Therefore, Theorem 1.4a applies, and so there is a finite set S such that $\mathrm{supp}(\nu) \subseteq S$ for each $\nu \in \mathcal{E}(F, \Lambda)$. In particular, each $\nu \in \mathcal{E}(F, \Lambda)$ is discrete. Hence, we have to solve the optimization problem given in Proposition 2.1e, which is

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

This optimization problem can be written as a semi-definite program, see [12, Corollary 4.1, page 936]. After obtaining numerical approximations to the optimizers of this problem, we guess that 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 $|ae^{2\pi i x} + b + ce^{-2\pi i x}| \leq 1$ for all $x \in \mathbb{T}$. Thus, $\varepsilon = 2$ and μ is a minimal extrapolation of F on Λ. Since $|ae^{2\pi i x} + b + ce^{-2\pi i x}| = 1$ precisely on $S = \{0, 1/3\}$, by Theorem 1.4a, the minimal extrapolations are supported in S.

The matrix E from (2.3) 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, then by Proposition 2.7, μ is the unique minimal extrapolation of F on Λ. Thus, reconstruction of μ from F on 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, without any assumptions on the coefficients of the discrete measure, a minimum separation condition is necessary to super-resolve a sum of two Dirac measures, such as in [12].

Example 3.6. Let $\Lambda \subseteq \mathbb{Z}^d$ be any finite subset and let $F_y = \widehat{\mu_y}\,|_\Lambda$, where $\mu_y = \delta_0 - \delta_y$ for some non-zero $y \in \mathbb{T}^d$. We claim that if y is sufficiently small depending on Λ, then μ_y is not a minimal extrapolation of

Please cite this article in press as: J.J. Benedetto, W. Li, Super-resolution by means of Beurling minimal extrapolation, Appl. Comput. Harmon. Anal. (2018), https://doi.org/10.1016/j.acha.2018.05.002

F_y on Λ. 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

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

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

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

We let D_Λ be the generalized Dirichlet kernel, defined by the formula

$$D_\Lambda(x) = \sum_{m\in\Lambda} e^{2\pi i m\cdot x}.$$

Then we have,

$$\|\nu_y\| = \int_{\mathbb{T}^d} \Big| \sum_{m\in\Lambda} F_y(m)e^{2\pi i m\cdot x}\Big|\ dx\ = \int_{\mathbb{T}^d} |D_\Lambda(x) - D_\Lambda(x-y)|\ dx.$$

By a fundamental theorem of calculus argument and Bernstein's inequality for trigonometric polynomials, we have the upper bound

$$|D_\Lambda(x) - D_\Lambda(x-y)| \le |y|\big\||\nabla D_\Lambda|\big\|_\infty \le 2\pi\sqrt{d}|y| \max_{m\in\Lambda} |m| \|D_\Lambda\|_\infty = 2\pi\sqrt{d}|y| \max_{m\in\Lambda} |m|(\#\Lambda),$$

where $|\cdot|$ denotes the Euclidean norm and ∇ is the gradient. Consequently, if $|y|$ is small enough so that

$$|y| < \frac{1}{\pi\sqrt{d}\max_{m\in\Lambda}|m|(\#\Lambda)},$$

then $\|\nu_y\| < 2 = \|\mu_y\|$. In this case, μ_y is not a minimal extrapolation of F_y on Λ. Note that this argument does not contradict Proposition 2.4 because $\|F_y\|_{\ell^\infty(\Lambda)} \to 0$ as $y \to 0$. Thus, for y sufficiently small, reconstruction of μ_y from F on Λ using (TV) is impossible.

Remark 3.7. The authors of [21, Corollary 1] showed that when $\Lambda = \{-M, -M+1, \dots, M\}$, there exists a real measure μ with minimum separation $1/(2M)$ that cannot be recovered using (TV), given its Fourier coefficients on Λ. For this particular case, their result is sharper than the one in Example 3.6 because they used a special fact about trigonometric polynomials due to Turán [44]. In contrast, our result does not require any assumptions on Λ or whether μ is real or complex, and it holds in all dimensions.

3.2. Singular continuous measures

The following example is an analogue of Example 3.2 for higher dimensions.

Example 3.8. Let $\Lambda = \{-1, 0, 1\}^2 \setminus \{(1,-1), (-1,1)\}$ and $F = \widehat{\mu}\ |_\Lambda$, where $\mu = \delta_{(0,0)} + \delta_{(1/2,1/2)}$. Then, $F(m) = 1 + e^{-\pi i(m_1+m_2)}$, and, in particular, $F(1,1) = F(-1,-1) = F(0,0) = 2$ and $F(\pm1, 0) = F(0,\pm1) = 0$. We deduce that $\varepsilon = \|\mu\| = \|F\|_{\ell^\infty(\Lambda)} = 2$ from Proposition 2.4, and so μ is a minimal extrapolation of F on Λ.

Please cite this article in press as: J.J. Benedetto, W. Li, Super-resolution by means of Beurling minimal extrapolation, Appl. Comput. Harmon. Anal. (2018), https://doi.org/10.1016/j.acha.2018.05.002

Further, $\Gamma = \{(0,0),(1,1),(-1,-1)\}$, and so $\#\Gamma = 3$. According to the definition of $\alpha_{m,n}$ in Theorem 1.4b, set $\alpha_{(-1,-1),(0,0)} = \alpha_{(0,0),(1,1)} = \alpha_{(-1,-1),(1,1)} = 0$. By the conclusion of Theorem 1.4b, 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

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

It follows that the minimal extrapolations are supported in

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

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

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

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 $\widehat{\nu_{y,K}} = F$ 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 σ is a minimal extrapolation of F on Λ. In particular, S is the smallest set that contains the support of all the minimal extrapolations.

Since μ is not the unique minimal extrapolation, reconstruction of μ from F on Λ using (TV) is impossible.

Example 3.9. 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

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

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 [46, 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.

Let $\Lambda \subseteq \mathbb{Z}$ be any set containing 0, and let $F = \widehat{\sigma_q}\,|_\Lambda$. Since $F(0) = \|\sigma_q\| = 1$, we immediately see that $\varepsilon = 1$ and σ_q is a minimal extrapolation of F on Λ. Again, we cannot determine whether σ_q is the unique minimal extrapolation because Theorem 1.4 cannot handle the case $\#\Gamma = 1$, see Remark 2.10. For related examples that can be analyzed in terms of minimal extrapolations, see [28,41,29].

Example 3.10. Let $\sigma_A, \sigma_B \in M(\mathbb{T}^d)$ be the surface measures of the Borel sets $A = \{x \in \mathbb{T}^2 \colon x_2 = 0\}$ and $B = \{x \in \mathbb{T}^2 \colon x_2 = 1/2\}$, respectively. Let $\Lambda = \{-2, -1, \dots, 1\}^2$ and $F = \widehat{\mu}\,|_\Lambda$, where $\mu = \sigma_A + \sigma_B$. Then,

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

We immediately see that $\varepsilon = \|F\|_{\ell^\infty(\Lambda)} = \|\mu\| = 2$, which implies μ is a minimal extrapolation of F on Λ. Then, $\Gamma = \{(0, 0), (0, 2), (0, -2)\}$, and, by Theorem 1.4b, the minimal extrapolations are supported in $\{x \in \mathbb{T}^2 \colon x_2 = 0\} \cup \{x \in \mathbb{T}^2 \colon x_2 = 1/2\}$. Determining whether μ is the unique minimal extrapolation is beyond the theory we have developed herein, and we shall examine this uniqueness problem in [6].

Acknowledgments

The first named author gratefully acknowledges the support of ARO Grant W911NF-16-1-0008, ARO Grant W911NF-17-1-0014, and DTRA Grant 1-13-1-0015. The second named author gratefully acknowledges the support of the Ann G. Wylie Dissertation Fellowship and DTRA Grant 1-13-1-0015. Both authors appreciate the salient critique of the three reviewers, and, in fact, have incorporated all of their suggestions.

References

[1] Enrico Au-Yeung, John J. Benedetto, Generalized Fourier frames in terms of balayage, J. Fourier Anal. Appl. 21 (3) (2015) 472–508.
[2] Céline Aubel, David Stotz, Helmut Bölcskei, A theory of super-resolution from short-time Fourier transform measurements, J. Fourier Anal. Appl. (2017) 1–63.
[3] Jean-Marc Azaïs, Yohann De Castro, Fabrice Gamboa, Spike detection from inaccurate samplings, Appl. Comput. Harmon. Anal. 38 (2) (2015) 177–195.
[4] John J. Benedetto, Harmonic Analysis and Applications, CRC Press Inc., 1996.
[5] John J. Benedetto, Wojciech Czaja, Integration and Modern Analysis, Birkhäuser Boston, Inc., Boston, 2009.
[6] John J. Benedetto, Weilin Li, Super-resolution of singular continuous measures on $\mathbb{T}^d$, 2018, in preparation.
[7] Arne Beurling, Balayage of Fourier–Stieltjes transforms, in: The Collected Works of Arne Beurling, vol. 2, 1989, pp. 341–350.
[8] Arne Beurling, Interpolation for an interval in $\mathbb{R}^1$, in: The Collected Works of Arne Beurling, vol. 2, 1989, pp. 351–365.

[9] Badri Narayan Bhaskar, Gongguo Tang, Benjamin Recht, Atomic norm denoising with applications to line spectral estimation, IEEE Trans. Signal Process. 61 (23) (2013) 5987–5999.
[10] Nicholas Boyd, Geoffrey Schiebinger, Benjamin Recht, The alternating descent conditional gradient method for sparse inverse problems, SIAM J. Optim. 27 (2) (2017) 616–639.
[11] Emmanuel J. Candès, Carlos Fernandez-Granda, Super-resolution from noisy data, J. Fourier Anal. Appl. 19 (6) (2013) 1229–1254.
[12] Emmanuel J. Candès, Carlos Fernandez-Granda, Towards a mathematical theory of super-resolution, Comm. Pure Appl. Math. 67 (6) (2014) 906–956.
[13] Emmanuel J. Candès, Justin K. Romberg, Terence Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inform. Theory 52 (2) (2006) 489–509.
[14] Venkat Chandrasekaran, Benjamin Recht, Pablo A. Parrilo, Alan S. Willsky, The convex geometry of linear inverse problems, Found. Comput. Math. 12 (6) (2012) 805–849.
[15] Scott S. Chen, David L. Donoho, Michael A. Saunders, Atomic decomposition by basis pursuit, SIAM Rev. 43 (1) (2001) 129–159.
[16] Donald G. Childers, Modern Spectrum Analysis, IEEE Press, Piscataway, New Jersey, 1978.
[17] Yohann De Castro, Fabrice Gamboa, Exact reconstruction using Beurling minimal extrapolation, J. Math. Anal. Appl. 395 (1) (2012) 336–354.
[18] Yohann De Castro, Fabrice Gamboa, Didier Henrion, J-B. Lasserre, Exact solutions to super resolution on semi-algebraic domains in higher dimensions, IEEE Trans. Inform. Theory 63 (1) (2017) 621–630.
[19] David L. Donoho, Compressed sensing, IEEE Trans. Inform. Theory 52 (4) (2006) 1289–1306.
[20] Bogdan Dumitrescu, Positive Trigonometric Polynomials and Signal Processing Applications, Springer, Dordrecht, 2007.
[21] Vincent Duval, Gabriel Peyré, Exact support recovery for sparse spikes deconvolution, Found. Comput. Math. 15 (5) (2015) 1315–1355.
[22] Carl-Gustav Esseen, Fourier analysis of distribution functions. A mathematical study of the Laplace–Gaussian law, Acta Math. 77 (1) (1945) 1–125.
[23] Carl-Gustav Esseen, A note on Fourier–Stieltjes transforms and absolutely continuous functions, Math. Scand. 2 (1954) 153–157.
[24] Mitra Fatemi, Arash Amini, Loic Baboulaz, Martin Vetterli, Shapes from pixels, IEEE Trans. Image Process. 25 (3) (2016) 1193–1206.
[25] Carlos Fernandez-Granda, Super-resolution of point sources via convex programming, Inform. Inference (2016) 1–53.
[26] Hayit Greenspan, Super-resolution in medical imaging, Comput. J. 52 (1) (2009) 43–63.
[27] Edwin Hewitt, Herman Rubin, The maximum value of a Fourier–Stieltjes transform, Math. Scand. 3 (1955) 97–102.
[28] Einar Hille, J.D. Tamarkin, Remarks on a known example of a monotone continuous function, Amer. Math. Monthly 36 (5) (1929) 255–264.
[29] Jean-Pierre Kahane, Raphaël Salem, Ensembles Parfaits et Séries Trigonométriques, Hermann, Paris, 1963.
[30] Yitzhak Katznelson, An Introduction to Harmonic Analysis, Cambridge University Press, 2004.
[31] Valery Khaidukov, Evgeny Landa, Tijmen Jan Moser, Diffraction imaging by focusing-defocusing: an outlook on seismic superresolution, Geophysics 69 (6) (2004) 1478–1490.
[32] Henry J. Landau, Maximum entropy and the moment problem, Bull. Amer. Math. Soc. 16 (1) (1987) 47–77.
[33] Jari Lindberg, Mathematical concepts of optical superresolution, J. Opt. 14 (8) (2012) 083001.
[34] Greg Ongie, Sampurna Biswas, Mathews Jacob, Convex recovery of continuous domain piecewise constant images from non-uniform Fourier samples, IEEE Trans. Signal Process. 66 (1) (2017) 236–250.
[35] Greg Ongie, Mathews Jacob, Off-the-grid recovery of piecewise constant images from few Fourier samples, SIAM J. Imaging Sci. 9 (3) (2016) 1004–1041.
[36] Hanjie Pan, Thierry Blu, Pier Luigi Dragotti, Sampling curves with finite rate of innovation, IEEE Trans. Signal Process. 62 (2) (2014) 458–471.
[37] Sung C. Park, Min K. Park, Moon G. Kang, Super-resolution image reconstruction: a technical overview, IEEE Signal Process. Mag. 20 (3) (2003) 21–36.
[38] K.G. Puschmann, F. Kneer, On super-resolution in astronomical imaging, Astron. Astrophys. 436 (1) (2005) 373–378.
[39] Fred Rieke, Spikes: Exploring the Neural Code, MIT press, 1999.
[40] Walter Rudin, Fourier Analysis on Groups, John Wiley & Sons, Inc., New York, 1962.
[41] Raphaël Salem, On singular monotonic functions of the Cantor type, J. Math. Phys. 21 (1942) 69–82.
[42] Gongguo Tang, Badri N. Bhaskar, Benjamin Recht, Near minimax line spectral estimation, IEEE Trans. Inform. Theory 61 (1) (2015) 499–512.
[43] Gongguo Tang, Badri N. Bhaskar, Parikshit Shah, Benjamin Recht, Compressed sensing off the grid, IEEE Trans. Inform. Theory 59 (11) (2013) 7465–7490.
[44] Paul Turán, On rational polynomials, Acta Univ. Szeged., Sect. Sci. (1946) 106–113.
[45] Weiyu Xu, Jian-Feng Cai, Kumar V. Mishra, Myung Cho, Anton Kruger, Precise semidefinite programming formulation of atomic norm minimization for recovering d-dimensional ($D \geq 2$) off-the-grid frequencies, in: Information Theory and Applications Workshop, 2014, IEEE, 2014, pp. 1–4.
[46] Antoni Zygmund, Trigonometric Series, vol. 1, Cambridge University Press, 1959, p. 2002.

Please cite this article in press as: J.J. Benedetto, W. Li, Super-resolution by means of Beurling minimal extrapolation, Appl. Comput. Harmon. Anal. (2018), https://doi.org/10.1016/j.acha.2018.05.002