

# Sampling the Fourier transform along radial lines

Charles Dossal \*

Vincent Duval<sup>†</sup>

Clarice Poon <sup>‡§</sup>

December, 2016; Revised May, 2017

## Abstract

This article considers the use of total variation minimization for the recovery of a superposition of point sources from samples of its Fourier transform along radial lines. We present a numerical algorithm for the computation of solutions to this infinite dimensional problem. The theoretical results of this paper make precise the link between the sampling operator and the recoverability of the point sources.

## 1 Introduction

Over the last decades, the use of regularizers for inverse problems has largely shifted from the use of Tikhonov regularization to  $\ell^1$  minimization, following the pioneering works of Donoho and Tibshirani [11, 37]. Using these approaches, one is generally led to search for a sparse signal on a predefined grid and to solve a finite-dimensional problem. Furthermore, the introduction of compressed sensing [14, 10] has in particular triggered an intense amount of research into the notion of sparse recovery. Although compressed sensing has largely been a finite dimensional theory, there has, in recent years, been several works studying the impact of sparsity for *infinite-dimensional* inverse problems. One aspect of this is to work in a gridless setting, replacing the discrete  $\ell^1$  norm with its continuous counterpart, the total variation of Radon measures. This approach was proposed independently by several authors [9, 6, 13, 4], and substantial mathematical insight was provided in [9].

In this article, we continue this line of investigation and explore the use of total variation minimization for recovering the parameters of a superposition of point sources when restricted to sampling along radial lines of its Fourier transform. Our analysis reveals that the full total variation minimization problem can be solved by considering a sequence of univariate minimization problems. Utilising this observation, we describe an algorithm for solving the total variation minimization problem by solving a sequence of univariate semi-definite programs. Our approach is infinite dimensional in the sense that it allows for the recovery of the point sources without resorting to computations on a discrete grid. On the theoretical level, we provide sufficient conditions on the number of radial lines and the number of samples along these radial lines to guarantee exact reconstruction. Our main results show that in dimension  $d$ , one can recover the parameters of a superposition of  $M$  point sources by sampling its Fourier transform along  $d + 1$  radial lines. Furthermore, the number of samples we require along each line is, up to log factors, linear with  $M$ .

**Motivation** The problem of parameter estimation for a superpositions of point sources is rooted in applications such as astronomy, NMR (nuclear magnetic resonance) spectroscopy [18, 41, 28, 40, 19] and microscopy [27, 34]. In these applications, the signal of interest can often be modelled as point sources and limitations in the hardware mean that one is required to resolve fine details from low frequency data. Convex programming techniques for parameter estimation from low frequency measurements date back to the 1980's, with works in NMR spectroscopy [24] and in seismic prospecting [31]. Furthermore, since the publication [9], this subject has been a topic of intense research within the mathematical community [8, 36, 2, 3, 1, 15, 5]. However, these works have focussed on the case where one samples the Fourier transform at Cartesian grid points. On the other hand, physical constraints can sometimes restrict observations to certain angular directions [32, 33], and in the case of NMR spectroscopy, one is required to sample along continuous trajectories such as radial lines.

---

\*IMB, Université de Bordeaux ([charles.dossal@math.u-bordeaux.fr](mailto:charles.dossal@math.u-bordeaux.fr))

<sup>†</sup>MOKAPLAN, INRIA Paris, CEREMADE, Université Paris-Dauphine ([vincent.duval@inria.fr](mailto:vincent.duval@inria.fr))

<sup>‡</sup>DAMTP, Centre for Mathematical Sciences, University of Cambridge ([C.M.H.S.Poon@maths.cam.ac.uk](mailto:C.M.H.S.Poon@maths.cam.ac.uk))

<sup>§</sup>corresponding author

In this article, we consider this problem under the additional constraint that one can only sample along radial lines in the Fourier domain.

## 1.1 Problem statement

### 1.1.1 Acquisition model

Let  $d \in \mathbb{N}$  with  $d \geq 2$  and let  $X = \overline{B}(0, 1/2) \subset \mathbb{R}^d$  be the centered closed ball with radius  $1/2$ . We denote by  $\mathbb{S}^{d-1}$  the sphere embedded in  $\mathbb{R}^d$  and by  $\mathcal{M}(X)$  the space of Radon measures with support in  $X$ . For  $\theta \in \mathbb{S}^{d-1}$ , let  $\ell_\theta := \{t\theta : t \in \mathbb{R}\}$  be the radial line directed by  $\theta$  and  $P_{\ell_\theta}$  the orthogonal projection onto  $\ell_\theta$ . Given  $x \in X$ , let  $\delta_x$  denote the Dirac measure at  $x$ . The Fourier transform of  $\mu \in \mathcal{M}(X)$  at  $\xi \in \mathbb{R}^d$  is defined by

$$\mathcal{F}\mu(\xi) = \int_{\mathbb{R}^d} e^{-i2\pi\langle \xi, x \rangle} \mu(dx).$$

Throughout this paper, given a finite set  $E$ ,  $|E|$  denotes its cardinality.

In this article, we are interested in the recovery of a discrete measure  $\mu_0 = \sum_{j=1}^M a_j \delta_{x_j}$  where  $\{a_j\}_{j=1}^M \subset \mathbb{C}$  and  $\{x_j\}_{j=1}^M \subset X$ , given  $T$  samples of its Fourier transform along  $L$  radial lines. More precisely, for  $N \in \mathbb{N}$ , let  $\Gamma \subset \{-N, \dots, N\}$  with  $T = |\Gamma|$ , and let  $\Theta \subset \mathbb{S}^{d-1}$  with  $L = |\Theta|$ . Then, the given observation is a vector  $y_0 := \Phi\mu_0 \in \mathbb{C}^{T \times L}$ , where

$$\Phi : \mathcal{M}(X) \rightarrow \mathbb{C}^{T \times L}, \quad \Phi\mu := (\mathcal{F}\mu(k\theta))_{(k,\theta) \in \Gamma \times \Theta}. \quad (1.1)$$

Note that the Fourier transform of  $\mu_0$  at  $\xi \in \mathbb{R}^d$  is equal to  $\mathcal{F}\mu_0(\xi) = \sum_{j=1}^M a_j e^{i2\pi\langle \xi, x_j \rangle}$ .

### 1.1.2 Total variation minimization

Given the observation  $y_0 = \Phi\mu_0 \in \mathbb{C}^{T \times L}$  described above, in this article, we consider the solutions to the minimization problem

$$\min_{\mu \in \mathcal{M}(X)} \{\|\mu\|_{TV} : \Phi\mu = y_0\} \quad (\mathcal{P}_0(y_0))$$

where the total variation norm  $\|\cdot\|_{TV}$  is defined by

$$\|\mu\|_{TV} = \sup \{\operatorname{Re} \langle f, \mu \rangle_{\mathcal{C}(X), \mathcal{M}(X)} : f \in \mathcal{C}(X), \|f\|_\infty \leq 1\},$$

where  $\mathcal{C}(X) = \mathcal{C}(X, \mathbb{C})$  is the space of complex-valued continuous functions defined on  $X$ , and  $\langle f, \mu \rangle_{\mathcal{C}(X), \mathcal{M}(X)} := \int_{\mathbb{R}^d} \bar{f} d\mu$ . In the case of a discrete measure  $\mu_0 = \sum_{j=1}^M a_j \delta_{x_j}$ , its total variation amounts to  $\|\mu_0\|_{TV} = \sum_{j=1}^M |a_j|$ . So, in the discrete setting  $(\mathcal{P}_0(y_0))$  is simply an  $\ell^1$  minimization problem.

Our contributions in this article are two-fold: First, by considering the dual problem of  $(\mathcal{P}_0(y_0))$ , we observe that in certain cases, one can decompose our multi-variate minimization problem into a sequence of univariate minimization problems. This naturally leads to semi-definite programming approach for the computation of minimizers of  $(\mathcal{P}_0(y_0))$ . Secondly, by analysing the conditions under which the proposed method successfully recovers solutions of  $(\mathcal{P}_0(y_0))$ , we are led to a theoretical understanding of the solutions of  $(\mathcal{P}_0(y_0))$ . In particular, our main theorems will address the following questions:

- (Q1) How many radial lines should one sample along?
- (Q2) How many samples should one observe along each line?

**Overview** In Section 2, we derive our numerical algorithm, and in doing so, reveal some key properties which we would analyse for the derivation of the recoverability conditions for  $(\mathcal{P}_0(y_0))$ . In Section 3, we present the main theoretical results, which address (Q1) and (Q2) on how one should sample to guarantee the uniqueness of solutions of  $(\mathcal{P}_0(y_0))$ , and also when these solutions can be computed via our numerical algorithm. Numerical results are presented in Section 4. Although this paper is devoted to the study of the noiseless problem  $(\mathcal{P}_0(y_0))$ , we describe in Section 5 how our algorithm can be extended to handle the noisy setting. Section 6 gives an overview of other related works and also possible future directions for the present work. Finally, the proofs of our main results are presented in Section 7.

## 2 The numerical algorithm

In this section, we describe a gridless computational approach to  $(\mathcal{P}_0(y_0))$ . One of the key tools of this section was discovered in [9], where the authors showed that a univariate total variation minimization problem (with Fourier sampling at grid points) can be formulated as a semi-definite programming problem (SDP). The SDP approach allows for the computation of solutions to an infinite dimensional optimization problem without resorting to computations on a discrete grid. Utilising this link with SDP, we describe how, in certain cases,  $(\mathcal{P}_0(y_0))$  can be solved via a sequence of univariate SDP's. We begin this section by recalling some facts about the dual formulation of  $(\mathcal{P}_0(y_0))$ .

### 2.1 The dual formulation

Given  $y_0 = \Phi\mu_0$  with  $\mu_0 \in \mathcal{M}(X)$ , the dual problem<sup>1</sup> of  $(\mathcal{P}_0(y_0))$  is

$$\sup \left\{ \operatorname{Re} \langle q, y_0 \rangle : q \in \mathbb{C}^{T \times L} \quad \|\Phi^* q\|_\infty \leq 1 \right\}, \quad (\mathcal{D}_0(y_0))$$

where given  $q, y \in \mathbb{C}^{T \times L}$ ,  $\langle q, y \rangle := \sum_{(k,\theta) \in \Gamma \times \Theta} y_{k,\theta} \bar{q}_{k,\theta}$ . Observe that when endowed with the weak-\* topology, the dual of the space  $\mathcal{M}(X)$  is  $\mathcal{C}(X)$ , and the adjoint operator  $\Phi^*$  of  $\Phi$  is given by

$$\Phi^* : \mathbb{C}^{T \times L} \rightarrow \mathcal{C}(X), \quad q \in \mathbb{C}^{T \times L} \mapsto \sum_{(k,\theta) \in \Gamma \times \Theta} q_{\theta,k} e^{i2\pi k \langle \theta, \cdot \rangle}. \quad (2.1)$$

Moreover, one can prove the existence of solutions to both  $(\mathcal{P}_0(y_0))$  and  $(\mathcal{D}_0(y_0))$ , and that strong duality holds, that is  $(\mathcal{P}_0(y_0)) = (\mathcal{D}_0(y_0))$ . The primal and dual problems are related such that  $\mu \in \mathcal{M}(X)$  solves  $(\mathcal{P}_0(y_0))$  and  $q \in \mathbb{C}^{T \times L}$  solves  $(\mathcal{D}_0(y_0))$  if and only if

$$\Phi\mu = y_0, \quad \text{and} \quad \Phi^* q \in \{f \in \mathcal{C}(X) : \|f\|_\infty \leq 1, \quad \langle f, \mu \rangle = \|\mu\|_{TV}\}. \quad (2.2)$$

For a discrete measure  $\mu_0 = \sum_{j=1}^M a_j \delta_{x_j}$ , if there exists  $q \in \mathbb{C}^{T \times L}$  such that

$$\Phi^* q(x_j) = \operatorname{sgn}(a_j) := \frac{a_j}{|a_j|}, \quad \forall j \in \{1, \dots, M\}, \quad \|\Phi^* q\|_\infty \leq 1, \quad (2.3)$$

and the extremal points of  $\Phi^* q$  form a finite set  $E := \{x : |\Phi^* q(x)| = 1\}$  such that the map  $b \in \mathbb{C}^{|E|} \mapsto \Phi(\sum_{x \in E} b_x \delta_x)$  is injective, then  $\mu_0$  is the unique solution of  $(\mathcal{P}_0(y_0))$ . This result is essentially proved in [13, Lem. 1.1] where  $\operatorname{sgn}(a_j)$  are real numbers, note however, that a similar result for complex numbers is also proved in [9, Prop. A.1], and our conditions in fact imply the conditions of [9, Prop. A.1]. Since finding a vector  $q$  which satisfies (2.3) guarantees that  $\mu_0$  is a solution, we shall call  $\Phi^* q$  a dual certificate.

### 2.2 Splitting the dual problem

For each  $\theta \in \Theta$ , let  $y_\theta := (\mathcal{F}\mu_0(k\theta))_{k \in \Gamma}$ , and let  $\Theta' \subseteq \Theta$  be any subset of cardinality  $L' \leq L$ . Instead of  $(\mathcal{D}_0(y_0))$ , let us consider the following optimization problem:

$$\begin{aligned} & \sup \left\{ \frac{1}{L'} \sum_{\theta \in \Theta'} \operatorname{Re} \langle c_\theta, y_\theta \rangle : \forall \theta \in \Theta', \quad c_\theta \in \mathbb{C}^T, \quad \text{and} \quad \sup_{x \in X} \left| \sum_{k \in \Gamma} c_{\theta,k} e^{i2\pi k \langle \theta, x \rangle} \right| \leq 1 \right\} \\ &= \frac{1}{L'} \sum_{\theta \in \Theta'} \sup \left\{ \operatorname{Re} \langle c_\theta, y_\theta \rangle : \quad c_\theta \in \mathbb{C}^T, \quad \text{and} \quad \sup_{x \in X} \left| \sum_{k \in \Gamma} c_{\theta,k} e^{i2\pi k \langle \theta, x \rangle} \right| \leq 1 \right\}. \end{aligned} \quad (\tilde{\mathcal{D}}_0(y_0))$$

Given any family  $(c_\theta)_{\theta \in \Theta'}$  admissible for  $(\tilde{\mathcal{D}}_0(y_0))$ , we may construct  $q \in \mathbb{C}^{L \times T}$  with

$$q_{\theta,k} = \begin{cases} \frac{c_{\theta,k}}{L'} & \text{for } \theta \in \Theta', \\ 0 & \text{for } \theta \in \Theta \setminus \Theta'. \end{cases} \quad (2.4)$$

and we see that  $\|\Phi^* q\|_\infty \leq 1$  and  $q$  is admissible for  $(\mathcal{D}_0(y_0))$ , with  $\operatorname{Re} \langle q, y_0 \rangle = \frac{1}{L'} \sum_{\theta \in \Theta'} \operatorname{Re} \langle c_\theta, y_\theta \rangle$ . As a result  $(\tilde{\mathcal{D}}_0(y_0)) \leq (\mathcal{D}_0(y_0)) \leq (\mathcal{P}_0(y_0))$ .

---

<sup>1</sup>For all aspects of convex optimization, duality, subdifferential... we shall regard a complex valued measure  $m$  as a real vector-valued measure. We rely on the duality between  $\mathcal{C}(X, \mathbb{R}^2)$  and  $\mathcal{M}(X, \mathbb{R}^2)$  and we apply the results in [16] for real locally convex vector spaces.

Now, suppose that for each  $\theta \in \Theta'$ , there exists  $c_\theta \in \mathbb{C}^T$  such that

$$p_\theta : t \mapsto \sum_{k \in \Gamma} c_{\theta,k} e^{i2\pi k t} \text{ satisfies } \|p_\theta\|_\infty \leq 1, \quad \text{and } p_\theta(\langle \theta, x_j \rangle) = \text{sgn}(a_j), \quad \forall j \in \{1, \dots, M\}. \quad (2.5)$$

Then,  $q$  defined as in (2.4) is admissible for  $(\mathcal{D}_0(y_0))$  and satisfies  $(\Phi^* q)(x_j) = \text{sgn}(a_j)$  for all  $j \in \{1, \dots, M\}$ .

As a result,

$$\frac{1}{L'} \sum_{\theta \in \Theta'} \operatorname{Re} \langle c_\theta, y_\theta \rangle = \operatorname{Re} \langle q, y_0 \rangle = \langle \Phi^* q, \mu_0 \rangle = \sum_{j=1}^M |a_j| = \|\mu_0\|_{TV},$$

and we deduce that  $(\tilde{\mathcal{D}}_0(y_0)) = (\mathcal{D}_0(y_0)) = (\mathcal{P}_0(y_0))$  and that  $\mu_0$  is optimal for  $(\mathcal{P}_0(y_0))$ . Incidentally, each  $c_\theta$  maximizes the corresponding summand in  $(\tilde{\mathcal{D}}_0(y_0))$ .

To see why this observation is useful, we recall an observation from [9] concerning the solution to the dual problem in the univariate case: A univariate trigonometric polynomial satisfies  $\left\| \sum_{j \in \Gamma} c_j e^{i2\pi j \cdot} \right\|_\infty \leq 1$  with  $\Gamma \subset \{-N, \dots, N\}$  if and only if there exists a Hermitian matrix  $Q \in \mathbb{C}^{N \times N}$  such that

$$\begin{bmatrix} Q & c \\ c^* & 1 \end{bmatrix} \succeq 0, \quad \sum_{i=1}^{N-j} Q_{i,i+j} = \begin{cases} 1 & j=0 \\ 0 & j=1, 2, \dots, N-1 \end{cases}, \quad c_{\Gamma^c} = 0,$$

where  $c_{\Gamma^c}$  is the restriction of  $c$  to coefficients indexed by  $\Gamma^c$ . So, the dual problem

$$\sup_{c \in \mathbb{C}^N} \left\{ \operatorname{Re} \langle c, y \rangle : p = \sum_{j \in \Gamma} c_j e^{i2\pi j \cdot}, \|p\|_\infty \leq 1 \right\}$$

is equivalent to

$$\sup_{Q \in \mathbb{C}^{N \times N}, c \in \mathbb{C}^N} \left\{ \operatorname{Re} \langle c, y \rangle : \begin{bmatrix} Q & c \\ c^* & 1 \end{bmatrix} \succeq 0, \quad \sum_{i=1}^{N-j} Q_{i,i+j} = \begin{cases} 1 & j=0 \\ 0 & j=1, \dots, N-1 \end{cases}, \quad c_{\Gamma^c} = 0 \right\}.$$

Crucially, the latter equation can be solved using semi-definite programming (SDP).

From this observation, it is clear that since  $(\tilde{\mathcal{D}}_0(y_0))$  is formulated in terms of  $L'$  univariate trigonometric polynomials,  $(\tilde{\mathcal{D}}_0(y_0))$  can be solved via a sequence of univariate SDPs. More precisely, each summand in  $(\tilde{\mathcal{D}}_0(y_0))$  can be rewritten as

$$\sup_{Q \in \mathbb{C}^{N \times N}, c \in \mathbb{C}^N} \left\{ \operatorname{Re} \langle c, y_\theta \rangle : \begin{bmatrix} Q & c \\ c^* & 1 \end{bmatrix} \succeq 0, \quad \sum_{i=1}^{N-j} Q_{i,i+j} = \begin{cases} 1 & j=0 \\ 0 & j=1, \dots, N-1 \end{cases}, \quad c_{\Gamma^c} = 0 \right\}.$$

Under the assumption (2.5), for each  $\theta \in \Theta'$ ,  $(\langle x_j, \theta \rangle)_{j=1}^M$  is contained in the extremal points of the trigonometric polynomial  $p_\theta$ , which we denote by  $\mathcal{T}_\theta$ . In particular, for all  $x \in \{x_j : j = 1, \dots, M\}$  and all  $\theta \in \Theta'$ , there exists  $t \in \mathcal{T}_\theta$  such that  $\langle x, \theta \rangle = t$ , i.e.  $x \in t\theta + \ell_\theta^\perp$ . So,

$$\tilde{\Delta}_{\Theta'} := \bigcap_{\theta \in \Theta'} \bigcup_{t \in \mathcal{T}_\theta} \left( t\theta + \ell_\theta^\perp \right) \supset \{x_j : j = 1, \dots, M\}. \quad (2.6)$$

In the following, we write  $\tilde{\Delta}_{\Theta'} = \{\tilde{x}_j : j = 1, \dots, \tilde{M}\}$  provided  $\tilde{\Delta}_{\Theta'}$  is finite. In that case, assuming that the operator

$$A_{\Theta, \Gamma, \tilde{\Delta}_{\Theta'}} : a \in \mathbb{C}^{\tilde{M}} \mapsto \left( \sum_{j=1}^{\tilde{M}} a_j e^{-i2\pi k \langle \theta, \tilde{x}_j \rangle} \right)_{\theta \in \Theta, k \in \Gamma} \in \mathbb{C}^{T \times L} \quad \text{is injective,} \quad (2.7)$$

the solution  $(\tilde{a}_j)_{j=1}^{\tilde{M}}$  to the linear system  $A_{\Theta, \Gamma, \tilde{\Delta}_{\Theta'}} \tilde{a} = y_0$  will satisfy  $\sum_{j=1}^{\tilde{M}} \tilde{a}_j \delta_{\tilde{x}_j} = \mu_0$ .

This discussion suggests the following reconstruction procedure:

**Algorithm SPLITCERT:**

Let  $L' \leq L$  be some parameter chosen by the user.

1. For each  $\theta \in \Theta$ , via SDP, compute the maximizer  $c_\theta$  of

$$\sup \left\{ \operatorname{Re} \langle y_\theta, c \rangle : \begin{bmatrix} Q & c \\ c^* & 1 \end{bmatrix} \succeq 0, \quad \sum_{i=1}^{N-j} Q_{i,i+j} = \begin{cases} 1 & j=0 \\ 0 & j=1, \dots, N-1 \end{cases}, \quad c \Gamma^c = 0 \right\}.$$

Let  $\mathcal{T}_\theta$  be the extremal values of the trigonometric polynomial

$$p_\theta = \sum_{k \in \Gamma} c_{\theta,k} e^{2i\pi k \cdot}.$$

The values are found by constructing the companion matrix of the associated algebraic polynomial<sup>2</sup> and finding its eigenvalues with absolute value within a threshold of 1 (in our experiments, we choose a threshold of  $10^{-4}$ ).

2. For each subset  $\Theta' \subset \Theta$  of cardinality  $L'$ , let  $\tilde{\Delta}_{\Theta'}$  be as defined in (2.6), and  $A_{\Theta, \Gamma, \tilde{\Delta}_{\Theta'}}$  be as defined in (2.7). If  $\tilde{\Delta}_{\Theta'}$  is a discrete set  $\{\tilde{x}_j : j = 1, \dots, \tilde{M}\}$  and  $A_{\Theta, \Gamma, \tilde{\Delta}_{\Theta'}}$  is injective, then recover  $a_{\Theta'}$  as the solution to the linear system  $A_{\Theta, \Gamma, \tilde{\Delta}_{\Theta'}} a_{\Theta'} = y_0$ , and let

$$\mu_{\Theta'} = \sum_{j=1}^{\tilde{M}} a_{\Theta',j} \delta_{\tilde{x}_j}.$$

If  $\operatorname{sgn}(a_{\Theta'})_j = \frac{1}{|\Theta'|} \sum_{\theta \in \Theta'} p_\theta(\langle x_j, \theta \rangle)$  for all  $j \in \{1, \dots, \tilde{M}\}$ , then return  $\mu_{\Theta'}$ .

If the algorithm finishes step 2 without having returned any measure, it means that  $(\mathcal{P}_0(y_0))$  cannot be solved using the proposed splitting approach (with this choice of  $L'$ ). The rest of this paper is devoted to showing that the splitting approach succeeds in many practical cases. Note that when  $L' < L$ , the above algorithm may return several measures, however, the last assertion in Step 2 above ensures that each one of these measures is a minimizer of  $(\mathcal{P}_0(y_0))$ . Furthermore, by collecting the assumptions in the discussion above, the algorithm SPLITCERT returns precisely  $\mu_0$  provided that the following two conditions are satisfied:

- (A1) There exists  $\Theta' \subset \Theta$  of cardinality  $L'$  such that for each  $\theta \in \Theta'$ , there exists  $c_\theta \in \mathbb{C}^T$  such that the trigonometric polynomial  $p_\theta = \tilde{\Phi}^* c_\theta$  satisfies

$$\|p_\theta\|_\infty \leq 1, \quad p_\theta(\langle \theta, x_j \rangle) = \operatorname{sgn}(a_j), \quad \forall j = 1, \dots, M.$$

- (A2) For the set  $\Theta'$  defined in (A1),  $\tilde{\Delta}_{\Theta'}$  defined in (2.6),  $A_{\Theta, \Gamma, \tilde{\Delta}_{\Theta'}}$  is injective.

As mentioned in Section 2.1, a discrete measure  $\mu_0 = \sum_{j=1}^M a_j \delta_{x_j}$  with  $\Phi \mu = y_0$  is a solution to  $(\mathcal{P}_0(y_0))$  if there exists a dual certificate  $q \in \mathbb{C}^{T \times L}$  such that  $\Phi^* q(x_j) = \operatorname{sgn}(a_j)$  and  $\|\Phi^* q\|_\infty \leq 1$ . Since Algorithm SPLITCERT describes how to construct such dual certificate, we have in fact derived the following statement:

**Proposition 2.1.** *If (A1) and (A2) hold, then  $\mu_0$  is the unique solution to  $(\mathcal{P}_0(y_0))$  and is recovered by Algorithm SPLITCERT.*

**Remark 2.1** Note that by choosing  $L' < L$ , it is easier for the conditions (A1) and (A2) to be satisfied, however, there is an increased computational cost in Step 2. Although we do not theoretically study the case  $L' < L$ , we present a numerical example in Section 4 to demonstrate one situation where it may be advantageous to let  $L' < L$ .

In the next section, we present some results describing how one may choose  $\Theta$  and  $\Gamma$  such that (A1) and (A2) are satisfied with  $L' = L = d + 1$ .

### 3 Theoretical results

Before presenting our main results which provide conditions under which (A1) and (A2) are satisfied, let us recall some of the existing results on total variation minimization.

<sup>2</sup>As in [9], we associate to each trigonometric polynomial  $t \mapsto \sum_{-N \leq k \leq N} c_k e^{2i\pi kt}$  the polynomial  $1 - X^{2N} \sum_{-N \leq k, j \leq N} c_k \bar{c}_j X^{k-j}$ . Its roots on the unit circle yield the points where the trigonometric polynomial reaches 1 in modulus.

**Existing results on total variation minimization** A significant mathematical breakthrough in the understanding of the total variation minimization problem is [9], where the authors derived precise conditions under which one can exactly solve the parameter estimation problem for  $M$  point sources. For the purpose of analysing the problem of recovering point sources from samples of the Fourier transform at grid points, the key notion introduced in [9] is the minimum separation distance.

**Definition 3.1.** Let  $\mathbb{T}$  denote the one-dimensional torus. Given any discrete set  $\Delta \subset \mathbb{T}$ , let the minimum separation distance of  $\Delta$  be defined by

$$\nu_{\min}(\Delta) = \min_{t, t' \in \Delta, t \neq t'} |t - t'|_{\mathbb{T}}.$$

where  $|t - t'|_{\mathbb{T}}$  is the canonical distance on the torus between  $t$  and  $t'$ .

Their main result shows that one is guaranteed exact recovery provided that the sampling range is inversely proportional to the minimum separation distance between the positions of the point sources. More precisely, it suffices to sample Fourier coefficients with frequency no greater than  $f_c$  provided that the positions of the point sources are separated by a distance of at least  $\mathcal{O}(1/f_c)$ . Explicit constants are given in the one and two variate setting, see [9] for further details and [17] for improved constants. Moreover, on the practical side, the authors presented a numerical algorithm based on semi-definite programming which allows for the computation of the minimizer without resorting to discrete grids.

The theoretical result of [9] is extended to the probabilistic framework in [35] where it is shown that one can in fact guarantee exact recovery with high probability by subsampling at random the Fourier coefficients with frequency no greater than  $f_c$ , at a rate which is up to log factors linear with sparsity. Several other variants have been proposed, replacing the Dirichlet kernel with more general kernels such as the Cauchy or Gaussian kernel [36, 2], or extending the framework to spherical domains [3]. The robustness to noise of such methods is investigated in [8, 1, 15, 5].

In the remainder of this section, we present our main results, which can be seen as the analogy of the results of [9] and [35] in the case of sampling the Fourier transform along radial lines.

### 3.1 Main result I

The key notion in our setting is the minimum separation distance between the *projected* positions. In particular, if one is restricted to sampling along the directions specified by some set  $\Theta \subset \mathbb{S}^{d-1}$ , then along each line, one should observe the Fourier samples up to frequency  $N$ , which is inversely proportional to  $\inf_{\theta \in \Theta} \nu_{\min}(\Delta_\theta)$  where  $\Delta_\theta = \{\langle \theta, x_j \rangle\}_{j=1}^M$ . Now, as we show in Lemma 7.2, there are in fact only a finite number of directions  $\Theta$  for which this value is zero, so it is finite for almost every choice of  $\Theta$ . However, it may become arbitrarily small depending on the choice of the set  $\Theta$ . Understanding this minimum separation distance forms a substantial part of the analysis in this work, and in our first result, we make precise the dependence between the minimum separation distance of the projected positions of the underlying point sources and the sampling range. Furthermore, by allowing for random sampling along radial lines, provided that we sample from a ‘good’ range of angles based on some a-priori knowledge on the signal, up to log factors, one can guarantee exact recovery with high probability using total variation minimization from  $\mathcal{O}((d+1)M)$  samples.

**Theorem 3.2.** Let  $\mu_0 = \sum_{j=1}^M a_j \delta_{x_j}$  where  $\{x_j\}_{j=1}^M \subset X$  consists of fixed distinct points. Let  $S \subset \mathbb{S}^{d-1}$  be a set of non-zero measure such that

$$\nu_{\min} := \inf_{\theta \in S} \nu_{\min}(\Delta_\theta) > 0,$$

where  $\Delta_\theta = \{\langle \theta, x_j \rangle\}_{j=1}^M$ . Let  $\Theta$  be a set of  $d+1$  distinct elements drawn uniformly at random from  $S$  and let  $N = \lceil 2/\nu_{\min} \rceil$ . Then, the following holds:

1. If  $\Gamma = \{-N, \dots, N\}$ , then  $\mu_0$  is the unique solution to  $(\mathcal{P}_0(y_0))$  and it can be recovered by Algorithm SPLITCERT.
2. If  $\Gamma$  consists of  $m$  indices drawn uniformly at random from  $\{-N, \dots, N\}$ , where

$$m \gtrsim \max\{\log^2(N/\delta), M \log(M/\delta) \log(N/\delta)\},$$

and  $\{\text{sgn}(a_j)\}_{j=1}^M$  are drawn i.i.d. from the uniform distribution on the complex unit circle, then with probability exceeding  $1 - (d+1)\delta$ ,  $\mu_0$  is the unique solution to  $(\mathcal{P}_0(y_0))$  and it can be recovered by Algorithm SPLITCERT.

**Remark 3.1** As in [35], we need to assume in 2 that  $\{\text{sgn}(a_j)\}_{j=1}^M$  are drawn i.i.d. from the uniform distribution on the complex unit circle. It is likely that this is simply an artefact of proof techniques rather than an actual requirement.

### 3.2 Main result II: Recovery with fixed sampling directions

The previous result considered the recovery of some fixed signal from Fourier samples along a random selection of radial lines. In this subsection, we consider the case where the radial lines which we sample along are fixed, but the positions of the point sources are distributed uniformly in space.

The following result describes when the algorithm presented in this article succeeds with  $L = L' = d + 1$ . Under this setting, the sampling range along each line is  $\mathcal{O}(M^2)$ . Furthermore, for exact recovery, it suffices to randomly draw  $\mathcal{O}(M \log^2 M)$  Fourier samples along the  $d+1$  prescribed radial lines.

**Theorem 3.3.** *Let  $\mu_0 = \sum_{j=1}^M a_j \delta_{x_j}$ , where the positions  $(x_j)_{j=1}^M \in X$  are i.i.d. random variables with uniform law in  $X$ . Let  $\delta > 0$  and let  $N \geq \frac{4(d+1)}{\delta \sqrt{\pi(2d-1)}} M(M-1)$  and let  $\Theta \subset \mathbb{S}^{d-1}$  be a set of  $d+1$  fixed distinct angles, of which any subset of cardinality  $d$  is linearly independent.*

1. If  $\Gamma = \{-N, \dots, N\}$ , then with probability at least  $1 - (d+1)\delta$ ,  $\mu_0$  is the unique solution of  $(\mathcal{P}_0(y_0))$  and can be recovered by Algorithm SPLITCERT.

2. If  $\Gamma$  consists of  $m$  indices chosen uniformly at random from  $\{-N, \dots, N\}$  with

$$m \gtrsim M \log^2(M/\delta),$$

and  $\{\text{sgn}(a_j)\}_{j=1}^M$  are drawn i.i.d. from the uniform distribution on the complex unit circle, then, with probability at least  $1 - (d+1)\delta$ ,  $\mu_0$  is the unique solution of  $(\mathcal{P}_0(y_0))$  and can be recovered by Algorithm SPLITCERT.

**Comparison with Cartesian grid sampling** Let us comment further on Theorem 3.3 in the bivariate case. In [9], it was proved that if  $\mu_0 = \sum_{j=1}^M a_j \delta_{x_j}$ , where  $\Delta := \{x_j : j = 1, \dots, M\} \subset \mathbb{T}^2$  and  $\nu_{\min}(\Delta) \geq 2.38/N$ , then  $\mu_0$  can be exactly recovered by solving  $(\mathcal{P}_0(y_0))$  with the following sampling operator:

$$\Phi_{\text{Cartesian}} : \mu \mapsto \{\mathcal{F}\mu(k) : k \in \mathbb{Z}^2 : |k|_\infty \leq N\}.$$

Now, if the positions in  $\Delta$  are chosen uniformly at random, one can show (see Lemmas 7.5 and 7.6) that  $\nu_{\min}(\Delta) \asymp 1/M$  with high probability, while the *projected minimum separation distance* is of the order  $1/M^2$ . Therefore, one can recover  $\mu_0$  from  $M^2$  samples by either sampling the  $M^2$  Fourier coefficients of lowest frequencies, or according to Theorem 3.3, by sampling  $M^2$  Fourier coefficients of lowest frequencies along 3 distinct directions. Thus, when restricted of sampling along 3 radial lines, one attains the same sampling bounds as in the Cartesian grid sampling case. One can make a similar comparison in the second case of Theorem 3.3 – our result guarantees exact recovery with high probability by drawing at random  $\mathcal{O}(M \log^2(M))$  coefficients from  $\{-CM^2, \dots, CM^2\}$  along 3 radial lines, while the probabilistic result [12, 35] in the case of sampling the Fourier coefficients on a Cartesian grid requires  $\mathcal{O}(M \log^2(M))$  coefficients drawn at random from the grid  $\{k \in \mathbb{Z}^2 : |k| \leq C'M\}$ .

## 4 Numerical results

### 4.1 Verification of Theorem 3.2

Theorem 3.2 states that given any discrete measure  $\mu_0$ , one can reconstruct exactly by sampling frequencies in  $\{-N, \dots, N\}$  along 3 radial lines, whose angles are chosen at random from a range  $S$  for which the minimum projected separation distance is sufficiently large with respect to  $S$ . In Figure 1, we let

$$S = \{(\sin(\pi t), \cos(\pi t)) : |t - 1/2| \leq 1/K\} \tag{4.1}$$

for some  $K \in \mathbb{N}$  and demonstrate the reconstruction of  $M$  diracs which are well separated relative to  $S$ . For their reconstruction, we sample their Fourier frequencies along 3 directions which are chosen at random from  $S$ . Letting  $N = \lceil 2/\nu_{\min} \rceil$  where

$$\nu_{\min} = \min_{\theta \in S} \min_{j \neq k} |\langle x_j - x_k, \theta \rangle|_{\mathbb{T}},$$

we present the dual certificates constructed in the cases where one samples  $\{-N, \dots, N\}$  along each line, and where one chooses 50% of the samples at random from  $\{-N, \dots, N\}$ . In all cases, the positions and amplitudes are constructed with error smaller than  $10^{-5}$ . Throughout this section, the error between the reconstructed positions  $\hat{x}$  and the true positions  $x$  is  $\text{Err}_{\text{pos}} = \|\hat{x} - x\|_{\ell^2}$  and the error between the reconstructed amplitudes  $\hat{a}$  and the true amplitudes  $a$  is  $\text{Err}_{\text{amp}} = \|\hat{a} - a\|_{\ell^2}/\|a\|_{\ell^2}$ .

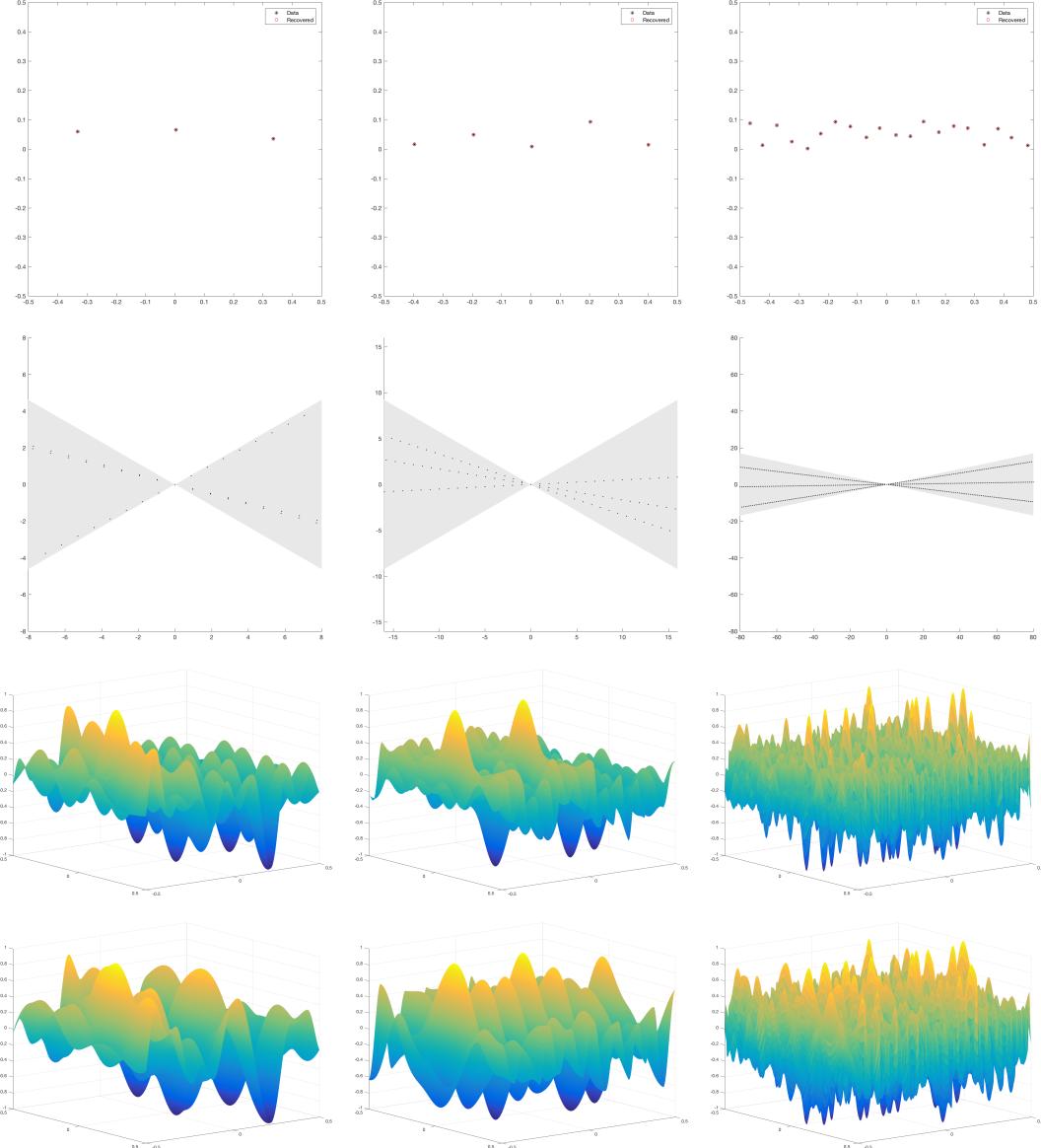


Figure 1: This figure shows the dual polynomials associated with the reconstruction of  $M$  diracs using the splitting construction introduced in this article. The far left, middle and far right columns correspond respectively to  $M = 3$ ,  $M = 5$  and  $M = 20$ . In the case of  $M = 3$  and  $M = 5$ , we let  $K = 6$  in the directions set  $S$  described in (4.1). For this choice,  $\nu_{\min} = 0.2181$  and  $\nu_{\min} = 0.1309$  respectively. In the case of  $M = 20$ , the set  $S$  is chosen with  $K = 15$  and  $\nu_{\min} = 0.0253$ . The angles spanned by  $S$  are indicated by the gray regions in the second row. In each case, the Fourier samples chosen along each line are from the frequencies  $\{-N, \dots, N\}$  where  $N = \lceil 2/\nu_{\min} \rceil$ . The indices of these Fourier samples are indicated by the black dots on the figures in the second row. The third row and last row show respectively the dual certificates constructed when one samples from  $\{-N, \dots, N\}$ , and the dual certificates constructed when one samples 50% of the frequencies in  $\{-N, \dots, N\}$ , along the three directions shown in the second row. In all cases, the original positions and amplitudes of the diracs are reconstructed with error at most  $10^{-5}$ .

## 4.2 Assessment of our algorithm

As discussed in Section 2.1, the existence of a dual certificate is closely related to the existence and uniqueness of solutions to  $(\mathcal{P}_0(y_0))$ . This article presented one possible construction of a dual certificate by combining a sequence of univariate dual certificates. Moreover, our algorithm is guaranteed

to provide the solutions of  $(\mathcal{P}_0(y_0))$  only when a splitting certificate exists and our main theorems provide a theoretical analysis of this. However, it is conceivable that there exists a dual certificate satisfying (2.3) even when the construction of a splitting certificate is not possible. To provide some understanding of the sharpness of our result, we shall compare numerically against the existence of another construction of a dual certificate, known as the vanishing derivatives precertificate [15].

One generic way of constructing a dual certificate is to consider the *vanishing derivatives precertificate*, which is defined as:

$$\eta_V := \Phi^* q_V \quad \text{where } q_V = \operatorname{argmin}\{|q|^2 : \forall i, (\Phi^* q)(x_i) = \operatorname{sgn}(a_i), d(\Phi^* q)(x_i) = 0\},$$

provided the problem is feasible (which is the case if  $\mu_0$  is a solution to  $(\mathcal{P}_0(y_0))$ ). Note that  $\eta_V$  is a dual certificate for  $\mu_0$  if and only if  $\|\eta_V\|_\infty \leq 1$ . In this case, we say that  $\eta_V$  is *nondegenerate*, and it is in fact equal to the *minimal norm certificate*, that is  $\eta_0 := \Phi^* q_0$  where  $q_0$  is the solution to  $(\mathcal{D}_0(y_0))$  with minimal  $\ell^2$ -norm.

The minimal norm certificate is of interest because it governs the support stability of problem  $(\mathcal{P}_0(y_0))$ , that is, the properties of the support of the solutions to

$$\inf_{\mu \in \mathcal{M}(X)} \lambda \|\mu\|_{TV} + \frac{1}{2} \|\Phi\mu - y_0 - w\|^2 \quad (\mathcal{P}_\lambda(y_0 + w))$$

where  $w \in \mathbb{C}^{T \times L}$  is some small noise added to the observation  $y_0$  and  $\lambda > 0$  is small. Here, support stability refers to the property that the method recovers the same number of spikes as the original measure  $\mu_0$ , and that their amplitudes and locations are close to those of  $\mu_0$ . We refer the reader to [15] for more detail on the connection between the support stability and the minimal norm certificate (although [15] is written for a convolution, the analysis carries over without major difficulty to more general operators).

The vanishing derivatives precertificate  $\eta_V$  can be seen as a proxy to understanding the stability properties of  $(\mathcal{P}_\lambda(y_0 + w))$  and is attractive as an analytical tool since it can be computed by simply solving a linear system. In particular,  $q_V$  is computed by applying the Moore-Penrose pseudo-inverse of  $\begin{pmatrix} \Phi_{x_0} & \Phi_{x_0}^{(1)} & \dots & \Phi_{x_0}^{(d)} \end{pmatrix}^*$  to  $(\operatorname{sgn}(a), \mathbf{0}_{dM})^T$ , where  $\mathbf{0}_{dM}$  consist of  $dM$  zeros,

$$\begin{aligned} \Phi_{x_0}^* &= \begin{pmatrix} e^{2i\pi\langle k_1\theta_1, x_1 \rangle} & \dots & e^{2i\pi\langle k_T\theta_L, x_1 \rangle} \\ \vdots & & \vdots \\ e^{2i\pi\langle k_1\theta_1, x_N \rangle} & \dots & e^{2i\pi\langle k_T\theta_L, x_N \rangle} \end{pmatrix}, \\ \Phi_{x_0}^{(\ell)*} &= 2i\pi \begin{pmatrix} \langle k_1\theta_1, v_\ell \rangle e^{2i\pi\langle k_1\theta_1, x_1 \rangle} & \dots & \langle k_T\theta_L, v_\ell \rangle e^{2i\pi\langle k_T\theta_L, x_1 \rangle} \\ \vdots & & \vdots \\ \langle k_1\theta_1, v_\ell \rangle e^{2i\pi\langle k_1\theta_1, x_N \rangle} & \dots & \langle k_T\theta_L, v_\ell \rangle e^{2i\pi\langle k_T\theta_L, x_N \rangle} \end{pmatrix}, \end{aligned}$$

and  $(v_1, \dots, v_d)$  is an orthonormal basis of  $\mathbb{R}^d$ . Note that the minimal norm certificate and vanishing derivatives precertificate can be constructed only when the underlying measure  $\mu_0$  is known, so they should be seen as analytic tools to understanding when one can recover  $\mu_0$  in a stable manner.

Numerically, in the case of sampling the Fourier transform on a grid,  $\eta_V$  has been numerically observed to be nondegenerate whenever the positions of the point sources are  $\mathcal{O}(1/f_c)$  apart, where  $f_c$  is the range of the sampled Fourier coefficients. This suggests that  $\eta_V$  yields an accurate understanding of stability and recoverability. Furthermore, we remark that theoretical results on conditions under which  $\eta_V$  is nondegenerate have been derived in the case of weighted Fourier samples [21] and convolutional sampling operators [2]. Therefore, in what follows, we shall compute the corresponding vanishing derivatives precertificate and regard its nondegeneracy as a means of checking whether there are cases of recoverability where our splitting certificate cannot be constructed.

**Example** In Figure 2, we present the experimental results where one samples along the directions

$$\{(\sin(\pi t), \cos(\pi t)) : t \in \{1/2 - 1/7, 1/2, 1/2 + 1/7\}\}.$$

Since the directions are closer to the vertical axis, as suggested by Theorem 3.3, we shall consider the recovery of point sources whose positions are drawn from a distribution which favours concentration along the horizontal axis. For each  $M$ , we generate 200 signals at random, whose amplitudes are drawn at random from the range  $[-55, 55]$  and whose positions are

$$\{\alpha_j(\cos(\beta_j), \sin(\beta_j)) : j = 1, \dots, M\},$$

where  $\beta_j \sim \mathcal{N}(0, 0.005)$  is drawn from a normal distribution with mean 0 and variance 0.005, and  $\alpha_j$  is drawn from the uniform distribution on  $[-1/2, 1/2]$ . Along each of the radial lines, the Fourier samples drawn are either those indexed by  $\{-N, \dots, N\}$  or 30% of those indexed by  $\{-N, \dots, N\}$ . The graphs show the fraction of these 200 signals which are exactly reconstructed (by exact, we mean that the reconstruction error of the positions and amplitudes are at most  $10^{-4}$ ). For comparison, we also compute the vanishing derivatives precertificate in each case and record the percentage of the signals for which the vanishing derivatives precertificate is nondegenerate. As mentioned in Section 4.2, nondegeneracy of the vanishing derivatives precertificate is generally an indication that one can stably recover a measure via TV minimization. Furthermore, note that the construction of the vanishing derivatives precertificate requires knowledge of the original measure, and thus can be seen as the ‘ground truth’ on which signals can be recovered. The similarity between the fraction of successful recovery via our splitting certificate and the fraction of signals for which the vanishing derivatives precertificate is nondegenerate suggests that the reconstruction via the SDP reconstruction procedure is close to that of the true TV minimization problem when the underlying measure can be reconstructed via TV minimization.

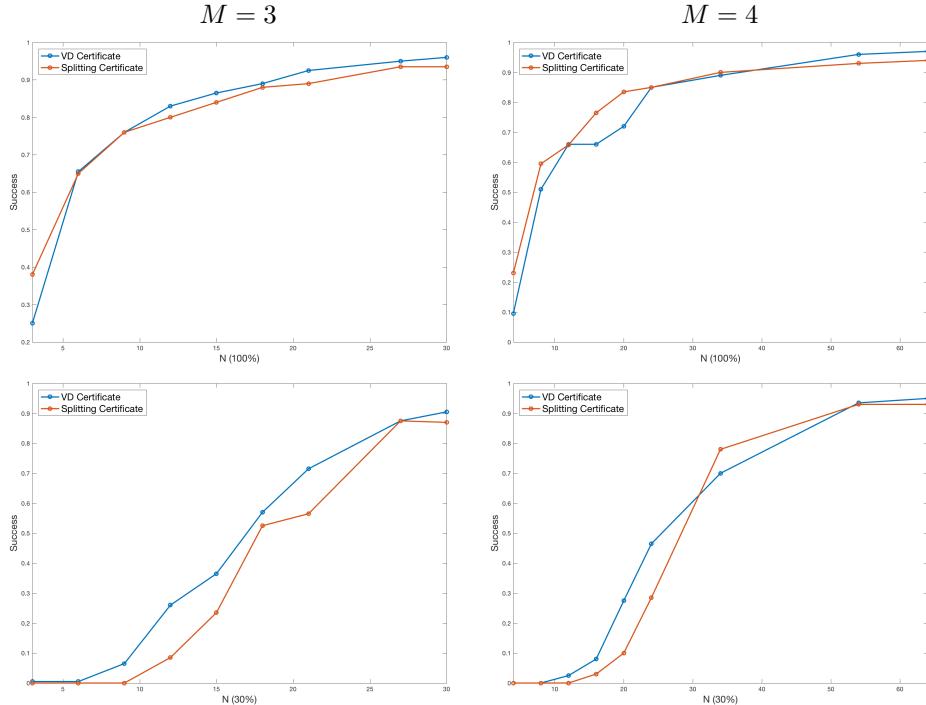


Figure 2: Fraction of the 200 signals which are successfully recovered via the splitting dual approach, and the fraction for which the vanishing derivatives precertificate is nondegenerate. The top row corresponds to sampling  $\{-N, \dots, N\}$  along 3 directions and the bottom row corresponds to subsampling 30% of  $\{-N, \dots, N\}$  along 3 directions

### 4.3 Verification of Theorem 3.3

In this experiment, we observe Fourier samples drawn along 3 fixed radial lines, directed by

$$\{(\sin(\pi t), \cos(\pi t)) : t \in \{0, 1/3, 2/3\}\}.$$

For each  $M$ , we consider the reconstruction of 200 signals whose positions are randomly chosen in the domain  $B(0, 1/2)$  and whose amplitudes are randomly chosen in  $[-55, 55]$ , in both instances, with the uniform distribution. Figure 3 shows the reconstruction results for the case where the Fourier samples observed along each line are indexed by  $\{-N, \dots, N\}$ . Figure 4 shows the reconstruction results for the case where we observe along each line 30% of the Fourier samples indexed by  $\{-N, \dots, N\}$ . For comparison, we show the fraction of signals for which the corresponding vanishing derivatives precertificate is nondegenerate.

As suggested by Theorem 3.3, one is required to sample at least  $M^2$  Fourier frequencies along each line to guarantee recovery with high probability. Notice that the fraction of successful recovery

via the spitting approach with  $L' = 3$  is lower than the fraction of signals for which the vanishing derivatives precertificate is nondegenerate, whereas, choosing  $L' = 2$  results in reconstruction rates which are roughly in line with the nondegeneracy of the vanishing derivatives precertificate. To see why choosing  $L' = 2$  yields better results, first note that it is easier for condition (A1) to be satisfied when  $L'$  is smaller – since choosing  $L' = 2$  means that we allow for the nonexistence of a dual certificate along one of the 3 chosen directions, one can choose a slightly lower sampling range. Furthermore, in the case where we have only two angles for which (A1) holds, the set  $\tilde{\Delta}$  for which (A2) is required to hold consists of up to  $M^2$  points. However, since each line consists of  $\mathcal{O}(M^2)$  sampling points, one can in fact show that the operator  $A_{\Theta, \Gamma, \tilde{\Delta}}$  is injective almost surely when the positions are drawn at random. The closeness of the recovery rates for this splitting certificate with the ‘ground truth’ recovery rates as approximated by the vanishing derivatives precertificate suggest that further analysis of this splitting certificate may lead to a deeper understanding of the solutions of  $(\mathcal{P}_0(y_0))$ .

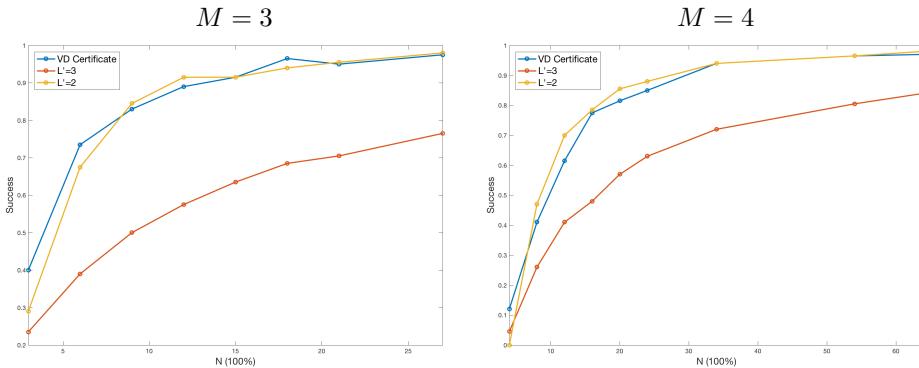


Figure 3: Fraction of the 200 signals which are successfully recovered via the splitting dual approach, and the fraction for which the vanishing derivatives precertificate is nondegenerate. Here, we sample all Fourier coefficients indexed by  $\{-N, \dots, N\}$  along angles  $\{0, \pi/3, 2\pi/3\}$ .

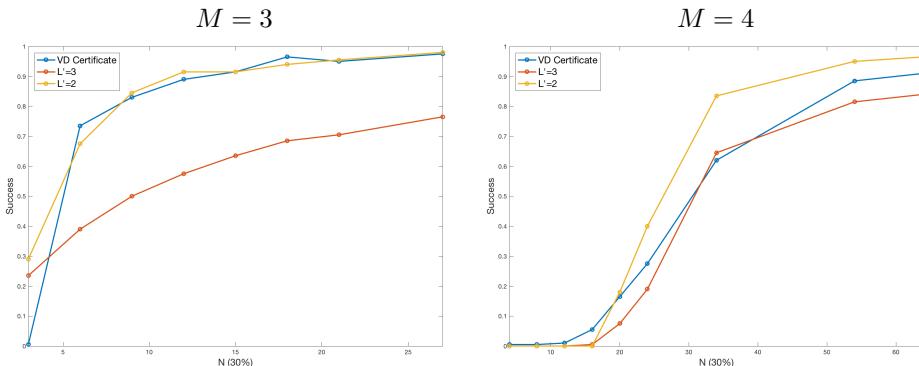


Figure 4: Fraction of the 200 signals which are successfully recovered via the splitting dual approach, and the fraction for which the vanishing derivatives precertificate is nondegenerate. Here, we sample 30% of the Fourier coefficients in  $\{-N, \dots, N\}$  along angles  $\{0, \pi/3, 2\pi/3\}$ .

## 5 Extension to the noisy setting

In this section, we explain how algorithm SPLITCERT can be extended for the recovery of  $\mu_0 = \sum_{j=1}^M a_j \delta_{x_j}$ , given noisy measurements  $y = y_0 + w$  where  $y_0 = \Phi \mu_0$  and  $w \in \mathbb{C}^{T \times L}$  is such that  $\|w\|_2 \leq \delta$  for some small  $\delta > 0$ . However, let us stress the fact that the proposed method does not coincide with the solution of  $(\mathcal{P}_\lambda(y_0 + w))$  for  $\lambda > 0$ .

First note that the dual problem of the robust TV minimization problem  $(\mathcal{P}_\lambda(y_0 + w))$  is

$$\sup \left\{ \operatorname{Re} \langle y_0, q \rangle - \frac{\lambda}{2} \|q\|_2^2 : \|\Phi^* q\|_\infty \leq 1 \right\}.$$

This naturally suggests that we can extend Algorithm SPLITCERT to the noisy setting by carrying out the following modifications:

- (I) Replace each SDP problem in Step 1 by its robust version:

$$\sup \left\{ \operatorname{Re} \langle y_\theta, c \rangle - \lambda \|c\|_2^2 : \begin{bmatrix} Q & c \\ c^* & 1 \end{bmatrix} \succeq 0, \quad \sum_{i=1}^{N-j} Q_{i,i+j} = \begin{cases} 1 & j = 0 \\ 0 & j = 1, \dots, N-1 \end{cases}, \quad c_{\Gamma^c} = 0 \right\}. \quad (5.1)$$

- (II) In Step 2, let

$$\tilde{\Delta}_{\Theta'} := \bigcap_{\theta \in \Theta'} \bigcup_{t \in \mathcal{T}_\theta} \mathcal{N}_\lambda(t\theta + \ell_\theta^\perp). \quad (5.2)$$

where for the first  $d$  directions in  $\Theta'$ , we let  $\mathcal{N}_\lambda(H) = H$  and for the remaining directions,

$$\mathcal{N}_\lambda(H) = \{x : \operatorname{dist}(x, H) \leq \lambda\}.$$

Note that in (II), the intersection of the hyperplanes directed by the first  $d$  directions consist of  $M^d$  points (see the proof of Lemma 7.2), and intersection of this finite point set with the remaining hyperplanes ensure that we recover only the points which are within a small neighbourhood of the hyperplanes prescribed by the remaining directions.

A detailed analysis of this reconstruction procedure is beyond the scope of this paper. Let us mention, however, that

- the problem (5.1) corresponds to the dual of the one-dimensional BLASSO problem. The study in [15] (see also the discussion in Section 4.2) ensures the support recovery at low noise for such problems, provided a Non-Degenerate Source Condition holds (which is empirically the case if the projection of all the spikes onto  $\ell_\theta$  are sufficiently separated). As a result, for small values of  $\lambda$  and  $\|w\|_2$ , the saturations of  $p_\theta$  should be exactly the (slightly shifted) projections  $\langle \theta, x_j \rangle$ , and nothing else ;
- in Figure 5, we present some numerical examples to demonstrate that the above modifications make Algorithm SPLITCERT robust to small perturbations in the Fourier measurements.

In Figure 5, we show the reconstruction of  $M$  point sources when given noisy samples along three directions. The samples which we observe are

$$y_{\text{observed}} = y_{\text{true}} + 0.15 \frac{n}{\|n\|},$$

where  $n$  is a vector whose entries are i.i.d. normal Gaussian, and  $y_{\text{true}} = \Phi \mu_0$  with

$$\Theta = \{(\sin(\pi t), \cos(\pi t)) : t \in \{0, 1/3, 2/3\}\}, \quad \Gamma = \{-N, \dots, N\}, \quad N = \lceil 1/\nu_{\min} \rceil.$$

In our examples, the error on the recovered positions ( $\text{Err}_{\text{pos}}$ ) and relative error on the recovered amplitudes ( $\text{Err}_{\text{amp}}$ ) are all below 0.05.

**Remark 5.1** If one is interested in the true BLASSO  $(\mathcal{P}_\lambda(y_0 + w))$ , the splitting approach exposed in this paper still provides some information in the sense that it constructs a dual certificate for  $(\mathcal{P}_0(y_0))$ . In the study of convex variational problems, it has been shown that the existence of a dual certificate  $v$  implies robustness to noise corrupted measurements, see [7] for stability estimates with respect to the Bregman distance and [8] for stability estimates with respect to the  $L^2$  norm. In particular, these estimates depend on the decay of  $|v|$  away from 1 outside the support of  $\mu_0$ . One of the contributions of this paper is the construction of a dual certificate in the case radial lines sampling. We therefore expect that by careful analysis of our splitting certificate, one can apply the results of [7] and [8] to understand the robustness properties of  $(\mathcal{P}_\lambda(y_0 + w))$ .

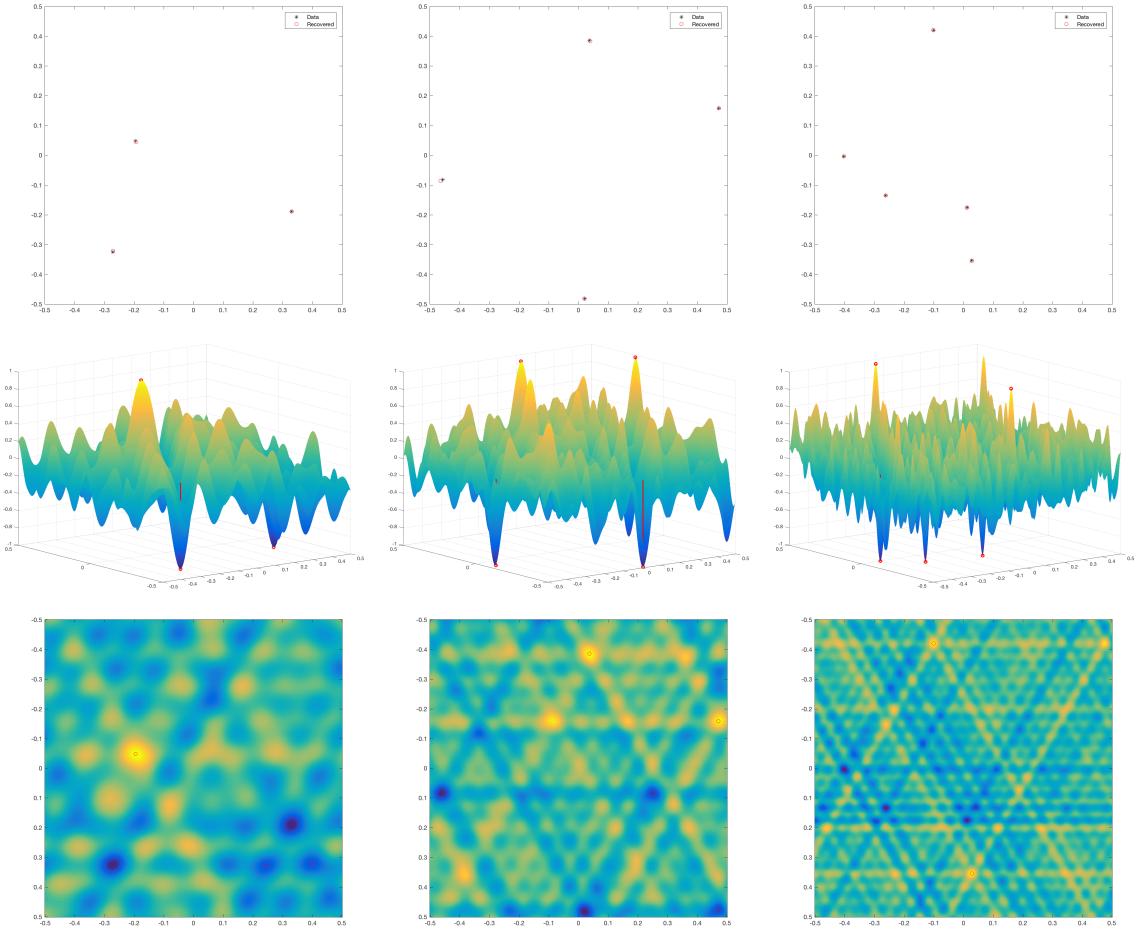


Figure 5: Reconstruction of point sources when the Fourier measurements have a relative error of 0.15. The top row shows the original and reconstructed positions, the middle row shows the corresponding dual certificates (with the true positions shown in red) and the bottom row shows a 2-dimensional view of the dual certificates. Left column:  $M = 3$ ,  $\nu_{\min} = 0.1208$ ,  $\text{Err}_{\text{pos}} = 0.0061$ ,  $\text{Err}_{\text{amp}} = 0.0359$ . Middle column:  $M = 4$ ,  $\nu_{\min} = 0.0704$ ,  $\text{Err}_{\text{pos}} = 0.0098$ ,  $\text{Err}_{\text{amp}} = 0.0280$ . Right column:  $M = 5$ ,  $\nu_{\min} = 0.0395$ ,  $\text{Err}_{\text{pos}} = 0.0021$ ,  $\text{Err}_{\text{amp}} = 0.0261$ .

## 6 Other related works and further comments

### 6.1 Prony methods

The question of how many projections are required to recover  $M$  sources has been of interest since the introduction of the Radon transform by J. Radon in 1917 [29]. In particular, it is long known [30] that  $M$  points are completely determined if its projection along  $M + 1$  lines are known. Since the early 2000's there has been numerous works describing how one can compute the parameters related to  $M$  point sources using Prony's method, when sampling its Fourier transform along  $M + 1$  (or fewer) radial lines. In [23], this problem was formulated in the framework of signals with a finite rate of innovation [39] and it was shown, using Prony's method (also referred to as the annihilating filter method), that one can exactly recover the parameters of  $M$  point sources from  $2M$  samples along  $M + 1$  lines. In [25], it is shown that one can recover the parameters of  $M$  point sources from  $2M$  samples along 3 lines, which are adaptively chosen. This result is similar to that of [26], where the authors present an algorithm to sample only along sufficiently many lines. In these works, the algorithms rely on the fact that the multi-dimensional parameter estimation problem can be projected onto a series of one-dimensional problems, under which one can simply apply the Prony's method in the univariate setting.

## 6.2 Links to the Radon transform

Here, we provide an informal discussion of the links between the results of this article and the Radon transform.

The Radon transform of a measure  $\mu_0$ , which consists of the collection of the projections of the measure  $\mu_0$  onto the line directed by  $\theta \in \mathbb{S}^{d-1}$ . That is, for  $\theta \in \mathbb{S}^{d-1}$ , we observe the measure  $P_{\ell_\theta \sharp} \mu \in \mathcal{M}(\ell_\theta)$ , where  $P_{\ell_\theta \sharp} \mu$  is the pushforward of  $\mu$  by  $P_{\ell_\theta}$ . So, given any bounded continuous function  $f : \ell_\theta \rightarrow \mathbb{R}$ ,

$$\int_{\ell_\theta} f(x') dP_{\ell_\theta \sharp} \mu(x') = \int_{\mathbb{R}^d} f(P_{\ell_\theta}(x)) d\mu(x). \quad (6.1)$$

For example, in the particular case where  $\mu$  has density  $\rho \in L^1(\mathbb{R}^d)$  with respect to the  $d$ -dimensional Lebesgue measure  $\mathcal{L}^d$ , then  $P_{\ell_\theta \sharp} \mu$  has density

$$\rho_\theta(s) = \int_{\ell_\theta^\perp} \rho(s\theta + v) d\mathcal{L}^{d-1}(v).$$

with respect to the Lebesgue measure  $\mathcal{L}^1$  on  $\ell_\theta$ , where  $\ell_\theta^\perp$  denotes the orthogonal space to  $\ell_\theta$ .

The Fourier slice theorem states that the slice of  $\mathcal{F}\mu$  in the direction  $\theta$  is the one-dimensional Fourier transform of  $P_{\ell_\theta \sharp} \mu$ , so, for all  $t \in \mathbb{R}$ ,

$$\mathcal{F}\mu(t\theta) = \int_{\mathbb{R}^d} e^{-2i\pi \langle x, t\theta \rangle} d\mu(x) = \int_{\mathbb{R}^d} e^{-2i\pi \langle P_{\ell_\theta}(x), t\theta \rangle} d\mu(x) = \int_{\ell_\theta} e^{-2i\pi t \langle x', \theta \rangle} dP_{\ell_\theta \sharp} \mu(x').$$

Consequently, given the Radon transform of  $\mu_0 \in \mathcal{M}(X)$  along a finite number of directions  $\Theta$ , by letting  $\eta_\theta = P_{\ell_\theta \sharp} \mu_0$ ,

$$\min_{\mu \in \mathcal{M}(X)} \|\mu\|_{TV} \text{ subject to } P_{\ell_\theta \sharp} \mu = \eta_\theta, \quad \forall \theta \in \Theta,$$

is equivalent to

$$\min_{\mu \in \mathcal{M}(X)} \|\mu\|_{TV} \text{ subject to } (\mathcal{F}\mu)(k\theta) = (\mathcal{F}\mu_0)(k\theta), \quad \forall \theta \in \Theta, \quad \forall k \in \mathbb{Z}$$

(since  $X = \overline{B}(0, 1/2)$ , there is no aliasing artifact due to the 1-periodization of  $P_{\ell_\theta \sharp} \mu_0$ ). We finally remark that in practice, the presence of white noise will typically render the higher Fourier coefficients unusable. Therefore, one is led to consider the above minimization problem with  $|k| \leq N$  for some  $N \in \mathbb{N}$ . This is precisely the problem studied in this paper and has been considered for applications such as electron tomography [20].

## 6.3 Generalizations to shift invariant linear operators

Suppose that  $\Phi$  is a Fourier sampling operator, let  $\mathcal{S}(X)$  denote the Schwartz space and let  $\mathcal{L} : \mathcal{S}(X) \rightarrow \mathcal{M}(X)$  be any differential operator. Let

$$\mathcal{M}_{\mathcal{L}}(X) = \left\{ f \in \mathcal{S}(X) : \|\mathcal{L}f\|_{TV} = \sup_{\varphi \in C_c^\infty(X) : \|\varphi\|_\infty=1} \operatorname{Re} \langle \mathcal{L}f, \varphi \rangle < \infty \right\}.$$

We now consider the generalized total variation minimization problem, given measurements  $y_0 = \Phi f_0$  for some  $f_0 \in \mathcal{M}_{\mathcal{L}}(X)$ :

$$\inf_{f \in \mathcal{M}_{\mathcal{L}}} \|\mathcal{L}f\|_{TV} \text{ subject to } \Phi(f) = y_0. \quad (6.2)$$

This problem is a specialized form of the generalized total variation minimization problems recently considered in [38], where  $\mathcal{L}$  is a linear shift invariant operator, which is associated with a finite dimensional null space  $\mathcal{N}(\mathcal{L})$  and a Green's function  $\rho \in \mathcal{M}_{\mathcal{L}}$  such that  $\mathcal{L}\rho = \delta$ , the Dirac measure. Examples of differential operators for (6.2) include the Laplacian and the fractional Laplacian, and the associated Green's functions are polyharmonic splines [22].

The main result of [38] establishes that, under mild conditions on the sampling operator  $\Phi$  of rank  $N$ , any solution to (6.2) is necessarily of the form

$$f(x) = \sum_{k=1}^K \alpha_k \rho(x - x_k) + \sum_{n=1}^{N_0} \beta_n g_n(x),$$

where  $\{g_n : n = 1, \dots, N_0\}$  is a basis for  $\mathcal{N}(\mathcal{L})$ ,  $\{x_k\}_{k=1}^K \subset \mathbb{R}^d$ ,  $\alpha \in \mathbb{R}^K$  and  $\beta \in \mathbb{R}^{N_0}$ , with  $K \leq N - N_0$ . In particular, any solution to (6.2) satisfies  $\sum_{j=1}^K a_j \delta_{x_j}$ . Now, returning to the case where  $\Phi$  is a Fourier operator, given  $f_0$  such that  $\mathcal{L}f_0 = \sum_{j=1}^M a_j \delta_{x_j}$ , suppose that the following hold.

1. There exists a trigonometric polynomial  $p = \Phi^* q$  such that

$$p(x_j) = \text{sgn}(a_j), \quad \|p\|_\infty \leq 1,$$

and the extremal points of  $p$  form a discrete set  $\{\tilde{x}_j\}_{j=1}^{\tilde{M}}$  such that

$$c \in \mathbb{C}^{\tilde{M}} \mapsto \Phi \left( \sum_{j=1}^{\tilde{M}} c_j \delta_{\tilde{x}_j} \right) \quad \text{is injective.}$$

2.  $\Phi$  is injective on  $\mathcal{N}(\mathcal{L})$ , the null space of  $\mathcal{L}$ .

Then, one can show that  $f_0$  is the unique minimizer to (6.2). Therefore, to understand the recoverability conditions on the sampling operator  $\Phi$ , one simply needs to understand the construction of the appropriate interpolating polynomial, and the behaviour of  $\Phi$  on  $\mathcal{N}_{\mathcal{L}}$ . Therefore, for the Fourier operator restricted to radial lines, the techniques developed in this paper should be readily extendible for the analysis of (6.2).

## 7 Proofs

### 7.1 Preliminary results

We first require a definition.

**Definition 7.1.** Let  $\mu \in \mathcal{M}(X)$ . Given  $\theta \in \mathbb{S}^{d-1}$ , let the one-dimensional subspace  $\ell_\theta = \{t\theta : t \in \mathbb{R}\}$  denote its corresponding radial line. We say that  $\theta \in \mathbb{S}^{d-1}$  is a projecting direction for  $\mu$  if the projection  $P_{\ell_\theta} : X \rightarrow \ell_\theta$ ,  $x \mapsto \langle x, \theta \rangle \theta$  is injective on  $\text{Supp}(\mu)$ .

Observe that in order for (A1) to make sense, for any values of  $\text{sgn}(a_j)$ , it is necessarily the case that each element  $\theta$  in  $\Theta$  is a projecting direction for  $\mu_0$ . Note also that if the extremal points of each  $p_\theta$  in (A1) is precisely  $\{x_1, \dots, x_M\}$ , then the set  $\tilde{\Delta}$  in (2.6) can be written as

$$\tilde{\Delta} = \bigcap_{\theta \in \Theta} \bigcup_{j=1}^M (x_j + \ell_\theta^\perp). \quad (7.1)$$

The following lemmas describe conditions under which this set is precisely the original positions  $\{x_j : j = 1, \dots, M\}$ .

**Lemma 7.2.** (i) For almost every finite choice of  $\Theta \subset \mathbb{S}^{d-1}$ , each element  $\theta$  in  $\Theta$  is a projecting direction for  $\mu_0$ ,  
(ii) Let  $\tilde{\Delta}$  be as defined in (7.1). If  $|\Theta| = d$ , then  $\tilde{\Delta}$  is a discrete set consisting of at most  $M^d$  elements. Moreover, for almost every choice of  $\Theta \subset \mathbb{S}^{d-1}$  with  $|\Theta| \geq d+1$ ,  $\tilde{\Delta} = \{x_1, \dots, x_M\}$ .

*Proof.* We first consider (i). A direction  $\theta \in \mathbb{S}^{d-1}$  is non-projecting for  $\mu_0$  if and only if there exists  $x_j \neq x_k$  such that  $x_j - x_k \in \ell_\theta^\perp$  or equivalently,  $\theta \in \ell_{x_j - x_k}^\perp$ . So,  $\theta \in \mathbb{S}^{d-1}$  is a non-projecting direction if and only if it belongs to one of the  $M(M-1)/2$  hyperplanes  $\{\ell_{x_j - x_k}^\perp\}_{1 \leq j < k \leq M}$ . However, since  $\mathbb{S}^{d-1} \cap \left( \bigcup_{1 \leq j < k \leq M} \ell_{x_j - x_k}^\perp \right)$  is of measure 0 (with respect to the Hausdorff measure on  $\mathbb{S}^{d-1}$ ), (i) is satisfied for almost every choice of  $\Theta$  from  $\mathbb{S}^{d-1}$ .

For (ii), observe that for almost every choice of  $d$  unit vectors  $\Theta' := \{\theta_1, \dots, \theta_d\}$ ,  $\Theta'$  spans  $\mathbb{R}^d$ . Then  $\bigcap_{k=1}^d \bigcup_{j=1}^M (x_j + \ell_{\theta_k}^\perp)$  consists of at most  $M^d$  points. Indeed, let  $z \in \mathbb{R}^d$ . Then  $z \in \bigcap_{k=1}^d \bigcup_{j=1}^M (x_j + \ell_{\theta_k}^\perp)$  if and only if for  $1 \leq k \leq d$ , there exists  $j_k \in \{1, \dots, M\}$  such that  $z \in x_{j_k} + \ell_{\theta_k}^\perp$ . That is equivalent to  $\langle \theta_k, z \rangle = \langle \theta_k, x_{j_k} \rangle$  for  $1 \leq k \leq d$ , or equivalently

$$(\theta_1 \quad \theta_2 \quad \cdots \quad \theta_d)^T z = b \quad (7.2)$$

for some  $b \in \mathbb{R}^d$  such that  $b_k := \langle \theta_k, x_{j_k} \rangle$ . There are at most  $M^d$  possible choices of  $b$ , and for each  $b$ , there is a unique solution  $z$  to the system (7.2).

Now, let us denote by  $Z$  the set of all such  $z$  (so that  $|Z| \leq M^d$ ). So,  $\Theta' \cup \{\theta_{d+1}\}$  satisfies (ii) provided that there does not exist  $z \in Z \setminus \{x_1, \dots, x_M\}$  such that  $z \in \bigcup_{i=1}^M (x_i + \ell_{\theta_{d+1}}^\perp)$ . This is equivalent to choosing  $\theta_{d+1}$  such that

$$\theta_{d+1} \in \mathbb{S}^{d-1} \setminus H, \quad H := \bigcup \left\{ \ell_{x_j - z}^\perp : 1 \leq j \leq M, z \in Z \setminus \{x_1, \dots, x_M\} \right\}.$$

Observe that  $H$  is the union of finitely many hyperplanes and so, just as in (i),  $H \cap \mathbb{S}^{d-1}$  is of measure 0 and almost every choice of  $\theta_{d+1}$  will ensure that  $\Theta' \cup \{\theta_{d+1}\}$  satisfies (ii). In particular, if  $|\Theta| = d + 1$ , then (ii) is satisfied for almost every choice of  $\Theta$ . The conclusion thus follows by observing that  $\{x_1, \dots, x_M\} \subset \tilde{\Delta}$  for any choice of  $\Theta$ .

□

The next lemma deals with the converse situation where the set of directions  $\Theta$  is fixed but the positions  $\{x_1, \dots, x_M\}$  are random.

**Lemma 7.3.** *Let  $\Theta \subset \mathbb{S}^{d-1}$  be a set of cardinality  $L \geq d + 1$ . Assume that any  $d$  elements of  $\Theta$  are linearly independent. Then for almost every choice of  $M$  points  $\{x_1, \dots, x_M\}$  in  $X$ ,  $\tilde{\Delta} = \{x_1, \dots, x_M\}$ .*

**Remark 7.1** In the case of  $d = 2$ , any 2 distinct elements of  $\mathbb{S}^1$  will be linearly independent. In general, if we let  $0 < t_1 < t_2 < \dots < t_L$ , then any  $d$  columns of the matrix

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_L \\ \vdots & \vdots & & \vdots \\ t_1^{d-1} & t_2^{d-1} & \cdots & t_L^{d-1} \end{pmatrix}$$

forms a Vandermonde matrix and hence forms a linearly independent set. So, an example of valid set  $\Theta$  for Lemma 7.3 is the set  $\{u_j : j = 1, \dots, L\}$  with

$$u_j := v_j / \|v_j\|_2 \quad v_j := (1, t_j, t_j^2, \dots, t_j^L).$$

*Proof.* Suppose we are given  $n$  points  $\{x_1, \dots, x_n\}$  such that

$$\bigcap_{k=1}^L \bigcup_{j=1}^n (x_j + \ell_{\theta_k}^\perp) = \{x_1, \dots, x_n\}. \quad (7.3)$$

Note that this is automatically true if  $n = 1$ . Let  $x_{n+1}$  be a random variable in  $X$ . Let  $P_{n+1}$  be the probability that

$$\bigcap_{k=1}^L \bigcup_{j=1}^{n+1} (x_j + \ell_{\theta_k}^\perp) = \{x_1, \dots, x_{n+1}\}.$$

Suppose that there exists

$$z \in \bigcap_{k=1}^L \bigcup_{j=1}^{n+1} (x_j + \ell_{\theta_k}^\perp) \setminus \{x_1, \dots, x_{n+1}\}. \quad (7.4)$$

Then, there exists  $\theta \in \Theta$  such that  $\langle z, \theta \rangle = \langle x_{n+1}, \theta \rangle$ , otherwise, we would violate assumption (7.3). Also, since  $z \neq x_{n+1}$ , there exists  $\Theta' \subset \Theta$  of cardinality at least  $L - d + 1$  such that

$$z \in \bigcap_{\theta \in \Theta'} \bigcup_{j=1}^n (x_j + \ell_\theta^\perp). \quad (7.5)$$

Indeed, if this was not the case, then there exists  $\{\theta_{k_1}, \dots, \theta_{k_d}\} \subset \Theta$  such that

$$\langle z - x_{n+1}, \theta_{j_k} \rangle = 0, \quad k = 1, \dots, d.$$

Since any  $d$  elements of  $\Theta$  form a basis of  $\mathbb{R}^d$ , this would imply that  $z = x_{n+1}$  and thus contradicting (7.4).

Let  $\Theta' \subseteq \Theta$  be the largest subset for which (7.5) holds. Note that  $|\Theta'| = d' \in [2, d]$ . Then,

$$\langle z - x_{n+1}, \theta \rangle = 0 \quad \forall \theta \in \Upsilon := \Theta \setminus \Theta'. \quad (7.6)$$

From (7.5), we see that  $z$  belongs to the intersection of  $d'$  hyperplanes. In particular, it belongs to a subspace  $V$  of dimension  $d - d' \leq d - 2$ . Also, from (7.6), we have that  $x_{n+1} - z \in W^\perp$ , where  $W = \text{span}(\theta \in \Upsilon)$ . Note that  $\Upsilon$  contains  $L - d' \geq d + 1 - d'$  elements. So,  $W$  is a subspace of dimension at least  $d + 1 - d'$  and  $W^\perp$  is of dimension at most  $d' - 1$ . So,  $x_{n+1} \in V + W^\perp$  is contained in a subspace of dimension at most  $d - 1$ , which is a set of measure 0 in  $X$ . Also, since there are finitely many combinations of sets of cardinality  $d' \in [2, d]$  in  $\Theta$ , and the union of finitely many sets of zero measure is also of zero measure, (7.4) holds with probability 0. Therefore,  $P_{n+1} = 1$  and by applying the chain rule for probabilities, the assertion of this lemma holds. □

## 7.2 The minimum separation distance and the sampling range

Solutions to  $(\mathcal{P}_0(y_0))$  in the case  $d = 1$  have been extensively studied [13, 9]. These works established conditions based on the *minimum separation distance*, under which one can construct one-variate trigonometric polynomials which interpolate the sign pattern related to the underlying measure.

**Lemma 7.4.** [35] Let  $\Delta = \{t_j : j = 1, \dots, M\} \subset \mathbb{T}$  and suppose that  $\nu_{\min}(\Delta) \geq 2/N$ .

(i) Let  $\{s_j\}_{j=1}^M \subset \{x \in \mathbb{C} : |x| = 1\}$ . If  $\Gamma = \{-N, \dots, N\}$ , then the operator  $\Psi : \mathbb{C}^M \rightarrow \mathbb{C}^m$  with  $\Psi a = \left( \sum_{j=1}^M a_j e^{-i2\pi k t_j} \right)_{k \in \Gamma}$  is injective and there exists  $q = \sum_{j \in \Omega} \alpha_j e^{i2\pi j \cdot}$  such that

$$q(t_j) = s_j, \quad j = 1, \dots, M \quad \text{and} \quad |q(t)| < 1, \quad \forall t \notin \Delta.$$

(ii) Let  $\{s_j\}_{j=1}^M$  be drawn i.i.d. from the uniform distribution on the complex unit circle. If  $\Gamma \subseteq \{-N, \dots, N\}$  consists of  $m$  indices chosen uniformly at random with

$$m \gtrsim \max\{\log^2(N/\delta), M \log(M/\delta) \log(N/\delta)\},$$

then the conclusions of (i) hold with probability at least  $1 - \delta$ .

As a consequence of Lemma 7.4, Lemma 7.2 and Proposition 2.1, we can prove the first of our main results.

*Proof of Theorem 3.2.* By Proposition 2.1, it suffices to show that conditions (A1) and (A2) are satisfied. We first consider the case where  $\Gamma = \{-N, \dots, N\}$ . By the assumption on the minimal distance  $\nu_{\min}$  and by Lemma (7.4), condition (A1) holds for all  $\theta \in \Theta$ . Furthermore, since for each  $\theta \in \Theta$ , the extremal values of  $p_\theta$  are precisely  $\{\langle x, \theta \rangle : \theta \in \Theta\}$ , by Lemma 7.2, we have that

$$\bigcap_{\theta \in \Theta} \bigcup_{j=1}^M (x_j + \ell_\theta^\perp) = \{x_1, \dots, x_M\} \tag{7.7}$$

holds with probability 1. So, the injectivity requirement of (A2) is satisfied if there exists  $\theta \in \Theta$  such that the operator  $A_\theta : \mathbb{C}^M \rightarrow \mathbb{C}^\Gamma$  defined by

$$A_\theta : \mathbb{C}^M \rightarrow \mathbb{C}^\Gamma, \quad a \in \mathbb{C}^M \mapsto \left( \sum_{j=1}^M a_j e^{-i2\pi k \langle \theta, x_j \rangle} \right)_{k \in \Gamma}.$$

is injective. However, this is true since  $\Gamma = \{-N, \dots, N\}$  contains  $M$  consecutive integers, and standard results on Vandermonde matrices imply that  $A_\theta$  is injective.

In the case where  $\Gamma$  consists of  $m$  indices chosen at random from  $\{-N, \dots, N\}$ , by applying Lemma 7.4 and the union bound, (A1) holds for all  $\theta \in \Theta$  with probability at least  $1 - (d+1)\delta$ . By Lemma 7.2, (A2) holds with probability 1 and by Lemma 7.4, given any  $\theta \in \Theta$ ,  $A_\theta$  is injective with probability least  $1 - \delta$ . So, (A2) holds with probability at least  $1 - \delta$ .  $\square$

## 7.3 Estimation of the minimum separation distance

Let  $Z = \{z_i : i = 1, \dots, M\}$  be i.i.d. random variables on the  $d$ -dimensional torus  $\mathbb{T}^d = (\mathbb{R}/\mathbb{Z})^d$  with density law  $f$ . Given  $z, z' \in \mathbb{T}^d$ , we denote by  $|z - z'|_{\mathbb{T}^d}$  the canonical (obtained from the Euclidean norm) distance on  $\mathbb{T}^d$ . Let  $E_M^\delta = \{\nu_{\min}(Z) \geq \delta\}$ , where, as before,  $\nu_{\min}(Z) := \min_{z, z' \in Z, z \neq z'} |z - z'|_{\mathbb{T}^d}$  and let  $V_d$  denote the volume of the unit ball of  $\mathbb{R}^d$ .

The following two lemmas show that with high probability  $\nu_{\min} \asymp M^{-2/d}$ , with Lemma 7.5 showing that  $\nu_{\min} \gtrsim M^{-2/d}$  and Lemma 7.6 showing that  $\nu_{\min} \lesssim M^{-2/d}$ .

**Lemma 7.5.** For any  $\rho > 0$  with  $\delta = \left( \frac{2\rho}{M(M-1)V_d \|f\|_2^2} \right)^{1/d} < 1$ , we have

$$\mathbb{P}(E_M^\delta) \geq 1 - \rho. \tag{7.8}$$

*Proof.* We first remark that

$$\begin{aligned} \mathbb{P}(E_2^\delta) &= 1 - \mathbb{P}(|z_1 - z_2|_{\mathbb{T}^d} < \delta) = 1 - \iint_{\mathbb{T}^d \times \mathbb{T}^d} \chi_{|t_1 - t_2|_{\mathbb{T}^d} < \delta} f_{z_1}(t_1) f_{z_2}(t_2) dt_1 dt_2 \\ &= 1 - \iint_{\mathbb{T}^d \times \mathbb{T}^d} \chi_{|s|_{\mathbb{T}^d} < \delta} f_{z_1}(t) f_{z_2}(t+s) dt ds \\ &= 1 - \int_{\mathbb{T}^d} \chi_{|s|_{\mathbb{T}^d} < \delta} \left( \int_{\mathbb{T}^d} f_{z_2}(t+s) f_{z_1}(t) dt \right) ds \geq 1 - \|f\|_2^2 \delta^d V_d. \end{aligned} \tag{7.9}$$

Then we observe that

$$\begin{aligned}\mathbb{P}(E_M^\delta) &= 1 - \mathbb{P}(\exists(i, j), i \leq M, j \leq M, i \neq j \text{ such that } |z_i - z_j|_{\mathbb{T}^d} < \delta) \\ &\geq 1 - \sum_{1 \leq i < j \leq M} \mathbb{P}(|z_i - z_j|_{\mathbb{T}^d} < \delta) \geq 1 - \frac{M(M-1)}{2} \mathbb{P}(|z_1 - z_2|_{\mathbb{T}^d} < \delta) \\ &\geq 1 - \frac{M(M-1)}{2} \|f\|_2^2 \delta^d V_d = 1 - \rho.\end{aligned}$$

□

**Lemma 7.6.** For any  $t > 0$ , defining  $\delta = 2 \left( \frac{t}{(M^2 - M - 1)V_d} \right)^{1/d}$ , we have

$$\mathbb{P}(E_M^\delta) \leq e^{-t} \quad (7.10)$$

*Proof.* We use now the expression of  $\mathbb{P}(E_M^\delta)$  as a product of conditional probabilities :

$$\mathbb{P}(E_M^\delta) = \prod_{j=3}^M \mathbb{P}(E_k^\delta | E_{k-1}^\delta) \times \mathbb{P}(E_2^\delta) \quad (7.11)$$

and recall from (7.9) that

$$\mathbb{P}(E_2^\delta) = 1 - V_d \delta^d \leq e^{-V_d \delta^d}. \quad (7.12)$$

We observe then that if  $E_{k-1}^\delta$  is satisfied, all the balls centered in  $z_i$ , ( $i \leq k-1$ ) with radius  $\frac{\delta}{2}$  are disjoint, hence the volume of the union of balls centered in  $z_i$  for  $i \leq k-1$  with radius  $\delta$  is at least  $\frac{(k-1)V_d \delta^d}{2^d}$ . This implies that

$$\mathbb{P}(E_k^\delta | E_{k-1}^\delta) \leq \max \left\{ 0, 1 - \frac{(k-1)V_d \delta^d}{2^d} \right\} \leq e^{-\frac{(k-1)V_d \delta^d}{2^d}}. \quad (7.13)$$

So,

$$\mathbb{P}(E_M^\delta) \leq \prod_{k=3}^M e^{-\frac{(k-1)V_d \delta^d}{2^d}} \leq e^{-\frac{(M^2 - M - 1)V_d \delta^d}{2^d}} = e^{-t}. \quad (7.14)$$

□

## 7.4 Proof of Theorems 3.3

Using Lemma 7.5 we can prove Theorem 3.3:

*Proof of Theorem 3.3.* By Proposition 2.1, it suffices to show that conditions (A1) and (A2) are satisfied. For the first part, we shall apply Lemma 7.5 to show that for our  $N$ ,  $\nu_{\min} \geq 2/N$  with probability larger than  $1 - (d+1)\delta$ . Hence, we need to consider the laws of the projections of points  $x_j$  on each directions  $\theta \in \Theta$ . Since the original density law of each  $x_j$  is uniform on the  $\ell_2$ -ball  $X$  of radius  $1/2$ , if we denote by  $f_{d,X}$  the density law of the projection  $(t_j)_{j=1}^d \mapsto (t_1, 0, \dots, 0)$ ,  $f_{d,X}(t)$  is the normalized  $\mathcal{H}^{d-1}$  measure of the intersection of the hyperplane  $\{(t, x) : x \in \mathbb{R}^{d-1}\}$  with the ball  $X$ , that is a  $(d-1)$ -dimensional ball of radius  $\sqrt{1/4 - t^2}$ . Therefore,

$$f_{d,X}(t) = \frac{V_{d-1}(\sqrt{1/4 - t^2})}{V_d(1/2)} = 2^d (1/4 - t^2)^{(d-1)/2} \frac{V_{d-1}(1)}{V_d(1)},$$

where  $V_d(r)$  denotes the volume of the  $d$ -dimensional ball of radius  $r$ . Hence,

$$\begin{aligned}\|f_{d,X}\|_2^2 &= 4 \left( \frac{V_{d-1}(1)}{V_d(1)} \right)^2 \int_0^1 (1 - t^2)^{d-1} dt = \frac{2V_{d-1}(1)^2 V_{2d-1}(1)}{V_d(1)^2 V_{2d-2}(1)} \\ &= \frac{2\pi^{-\frac{1}{2}} G(\frac{d}{2} + 1)^2 G(d)}{G(\frac{d+1}{2})^2 G(d + \frac{1}{2})} \leq \frac{2d+2}{\sqrt{\pi(2d-1)}},\end{aligned}$$

where  $G$  denotes the Gamma function. Observe that we should compute the density of the 1-periodized projections of the  $x_i$ 's, hence we should compute the 1-periodization of  $f_{d,X}$ . But the assumption on the domain ( $X = \overline{B}(0, 1/2)$ ) prevents aliasing effects, and the periodization of  $f_{d,X}$  coincides with its restriction on  $(-1/2, 1/2]$ .

Thus, we can lower bound the minimum separation distance along each projected direction. Then, by applying (i) of Lemma 7.4 and the union bound, when  $\Gamma = \{-N, \dots, N\}$ , we can conclude that (A1) holds with  $L' = L$  with probability at least  $1 - (d+1)\delta$ . For (A2), note that the fact that  $(x_j)_{j=1}^M$  are random and independent of the directions  $\Theta$  ensures, by Lemma 7.3, that we simply need to assert the injectivity condition in (A2) on  $\tilde{\Delta} = \{x_1, \dots, x_M\}$ . However, (A2) is true by standard results on Vandermonde matrices.

The second part of Theorem 3.3 is proved similarly, by applying Lemma 7.5, the union bound and (ii) of Lemma 7.4.  $\square$

## 8 Conclusion

There has been several theoretical works on the recovery of point sources (in an infinite dimensional setting) via convex optimization approaches. Previous works tend to consider sampling on some uniform grid, whereas, this article presents the analogous results in the case where we are restricted to sampling along radial lines. We describe how the semidefinite programming approach of [9] can be extended to compute the solutions of the total variation minimization problem in our radial lines setting. Although this work is concerned only with the recovery of discrete measures, our framework should be readily extendible for the study of generalized total variation minimization problems (as discussed in Section 6.3). For future work, one could consider the extension of our results to the recovery of more general functions. Such results would be of interest due to the practicality of radial lines sampling for applications such as magnetic resonance imaging.

## Acknowledgements

This work was partially supported by CNRS (Défi Imag'in de la Mission pour l'Interdisciplinarité, project CAVALIERI). The authors would like to thank Anders C. Hansen for reading an early draft of this work and for his constructive comments.

## References

- [1] J.-M. Azaïs, Y. De Castro, and F. Gamboa. Spike detection from inaccurate samplings. *Applied and Computational Harmonic Analysis*, 38(2):177–195, 2015.
- [2] T. Bendory, S. Dekel, and A. Feuer. Robust recovery of stream of pulses using convex optimization. *Journal of Mathematical Analysis and Applications*, 442(2):511–536, 2014.
- [3] T. Bendory, S. Dekel, and A. Feuer. Exact recovery of dirac ensembles from the projection onto spaces of spherical harmonics. *Constructive Approximation*, 42(2):183–207, 2015.
- [4] B. N. Bhaskar, G. Tang, and B. Recht. Atomic norm denoising with applications to line spectral estimation. *IEEE Transactions on Signal Processing*, 61(23):5987–5999, 2013.
- [5] C. Boyer, Y. De Castro, and J. Salmon. Adapting to unknown noise level in sparse deconvolution. Technical report, 2016.
- [6] K. Bredies and H. Pikkariainen. Inverse problems in spaces of measures. *ESAIM: Control, Optimisation and Calculus of Variations*, 19(1):190–218, 2013.
- [7] M. Burger and S. Osher. Convergence rates of convex variational regularization. *Inverse problems*, 20(5):1411, 2004.
- [8] E. J. Candès and C. Fernandez-Granda. Super-resolution from noisy data. *Journal of Fourier Analysis and Applications*, 19(6):1229–1254, 2013.
- [9] E. J. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, 2014.
- [10] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52(2):489–509, 2006.
- [11] S. Chen, D. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. *SIAM journal on scientific computing*, 20(1):33–61, 1999.

- [12] Y. Chi and Y. Chen. Compressive two-dimensional harmonic retrieval via atomic norm minimization. *Signal Processing, IEEE Transactions on*, 63(4):1030–1042, 2015.
- [13] Y. De Castro and F. Gamboa. Exact reconstruction using beurling minimal extrapolation. *Journal of Mathematical Analysis and Applications*, 395(1):336–354, 2012.
- [14] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.
- [15] V. Duval and G. Peyré. Exact support recovery for sparse spikes deconvolution. *Foundations of Computational Mathematics*, 15(5):1315–1355, 2015.
- [16] I. Ekeland and R. Temam. *Convex Analysis and Variational Problems*. Classics in Applied Mathematics. Siam, Philadelphia, 1999.
- [17] C. Fernandez-Granda. Super-resolution of point sources via convex programming. *Information and Inference*, page iaw005, 2016.
- [18] D. J. Holland, M. J. Bostock, L. F. Gladden, and D. Nietlispach. Fast multidimensional nmr spectroscopy using compressed sensing. *Angewandte Chemie International Edition*, 50(29):6548–6551, 2011.
- [19] K. Kazimierczuk and V. Y. Orehkov. Accelerated nmr spectroscopy by using compressed sensing. *Angewandte Chemie International Edition*, 50(24):5556–5559, 2011.
- [20] R. Leary, Z. Saghi, P. A. Midgley, and D. J. Holland. Compressed sensing electron tomography. *Ultramicroscopy*, 131:70–91, 2013.
- [21] Q. Li and G. Tang. Approximate support recovery of atomic line spectral estimation: A tale of resolution and precision. *arXiv preprint arXiv:1612.01459*, 2016.
- [22] W. R. Madych and S. Nelson. Polyharmonic cardinal splines. *Journal of Approximation Theory*, 60(2):141–156, 1990.
- [23] I. Maravic and M. Vetterli. A sampling theorem for the radon transform of finite complexity objects. In *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, volume 2, pages II–1197–II–1200, May 2002.
- [24] R. Newman. Maximization of entropy and minimization of area as criteria for nmr signal processing. *Journal of Magnetic Resonance (1969)*, 79(3):448–460, 1988.
- [25] G. Plonka and M. Wischerhoff. How many fourier samples are needed for real function reconstruction? *Journal of Applied Mathematics and Computing*, 42(1-2):117–137, 2013.
- [26] D. Potts and M. Tasche. Parameter estimation for multivariate exponential sums. *Electronic Transactions on Numerical Analysis*, 40:204–224, 2013.
- [27] K. Puschmann and F. Kneer. On super-resolution in astronomical imaging. *Astronomy & Astrophysics*, 436(1):373–378, 2005.
- [28] X. Qu, D. Guo, X. Cao, S. Cai, and Z. Chen. Reconstruction of self-sparse 2d nmr spectra from undersampled data in the indirect dimension. *Sensors*, 11(9):8888–8909, 2011.
- [29] J. Radon. 1.1 über die bestimmung von funktionen durch ihre integralwerte längs gewisser mannigfaltigkeiten. *Classic papers in modern diagnostic radiology*, 5, 2005.
- [30] A. Rényi. On projections of probability distributions. *Acta Mathematica Hungarica*, 3(3):131–142, 1952.
- [31] F. Santosa and W. W. Symes. Linear inversion of band-limited reflection seismograms. *SIAM Journal on Scientific and Statistical Computing*, 7(4):1307–1330, 1986.
- [32] P. N. Stewart, P. G. Tuthill, M. M. Hedman, P. D. Nicholson, and J. P. Lloyd. High-angular-resolution stellar imaging with occultations from the cassini spacecraft–i. observational technique. *Monthly Notices of the Royal Astronomical Society*, 433(3):2286–2293, 2013.
- [33] P. N. Stewart, P. G. Tuthill, P. D. Nicholson, M. M. Hedman, and J. P. Lloyd. High angular resolution stellar imaging with occultations from the cassini spacecraft–ii. kronocyclic tomography. *Monthly Notices of the Royal Astronomical Society*, 449(2):1760–1766, 2015.
- [34] V. Studer, J. Bobin, M. Chahid, H. S. Mousavi, E. Candès, and M. Dahan. Compressive fluorescence microscopy for biological and hyperspectral imaging. *Proceedings of the National Academy of Sciences*, 109(26):E1679–E1687, 2012.
- [35] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht. Compressed sensing off the grid. *Information Theory, IEEE Transactions on*, 59(11):7465–7490, 2013.

- [36] G. Tang and B. Recht. Atomic decomposition of mixtures of translation-invariant signals. In *IEEE CAMSAP*, 2013.
- [37] R. Tibshirani. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B. Methodological*, 58(1):267–288, 1996.
- [38] M. Unser, J. Fageot, and J. P. Ward. Splines are universal solutions of linear inverse problems with generalized-tv regularization. *arXiv preprint arXiv:1603.01427*, 2016.
- [39] M. Vetterli, P. Marziliano, and T. Blu. Sampling signals with finite rate of innovation. *IEEE transactions on Signal Processing*, 50(6):1417–1428, 2002.
- [40] S. Wenger, S. Darabi, P. Sen, K.-H. Glassmeier, and M. Magnor. Compressed sensing for aperture synthesis imaging. In *2010 IEEE International Conference on Image Processing*, pages 1381–1384. IEEE, 2010.
- [41] Y. Wu, C. D’Agostino, D. J. Holland, and L. F. Gladden. In situ study of reaction kinetics using compressed sensing nmr. *Chemical Communications*, 50(91):14137–14140, 2014.