Exact Support Recovery
for Sparse Spikes Deconvolution

Vincent Duval [*]   Gabriel Peyré [†]

January 8, 2017

Abstract

This paper studies sparse spikes deconvolution over the space of measures. We focus on the recovery properties of the support of the measure (i.e. the location of the Dirac masses) using total variation of measures (TV) regularization. This regularization is the natural extension of the ℓ^1 norm of vectors to the setting of measures. We show that support identification is governed by a specific solution of the dual problem (a so-called dual certificate) having minimum L^2 norm. Our main result shows that if this certificate is non-degenerate (see the definition below), when the signal-to-noise ratio is large enough TV regularization recovers the exact same number of Diracs. We show that both the locations and the amplitudes of these Diracs converge toward those of the input measure when the noise drops to zero. Moreover the non-degeneracy of this certificate can be checked by computing a so-called vanishing derivative pre-certificate. This proxy can be computed in closed form by solving a linear system. Lastly, we draw connections between the support of the recovered measure on a continuous domain and on a discretized grid. We show that when the signal-to-noise level is large enough, and provided the aforementioned dual certificate is non-degenerate, the solution of the discretized problem is supported on pairs of Diracs which are neighbors of the Diracs of the input measure. This gives a precise description of the convergence of the solution of the discretized problem toward the solution of the continuous grid-free problem, as the grid size tends to zero.

Keywords: ℓ^1 minimization, sparsity, stable signal recovery, sparse spike trains, super-resolution, deconvolution, dual certificates, interpolation.

AMS Subjsect Classification (2010): 65J20, 46E27, 49K40.

---
[*]INRIA, Domaine de Voluceau, Rocquencourt (vincent.duval@inria.fr)

[†]CNRS, CEREMADE, Université Paris-Dauphine (gabriel.peyre@ceremade.dauphine.fr)

1 Introduction

1.1 Sparse Spikes Deconvolution

Super-resolution is a central problem in imaging science, and loosely speaking corresponds to recovering fine scale details from a possibly noisy input signal or image. This thus encompasses the problems of data interpolation (recovering missing sampling values on a regular grid) and deconvolution (removing acquisition blur). We refer to the review articles [27, 24] and the references therein for an overview of these problems.

We consider in our article an idealized super-resolution problem, known as sparse spikes deconvolution. It corresponds to recovering 1-D spikes (i.e. both their positions and amplitudes) from blurry and noisy measurements. These measurements are obtained by a convolution of the spikes train against a known kernel. This setup can be seen as an approximation of several imaging devices. A method of choice to perform this recovery is to introduce a sparsity-enforcing prior, among which the most popular is a ℓ^1-type norm, which favors the emergence of spikes in the solution.

1.2 Previous Works

Discrete ℓ^1 regularization. ℓ^1-type techniques were initially proposed in geophysics [10, 28, 23] to recover the location of density changes in the underground for seismic exploration. They were later studied in depth by David Donoho and co-workers, see for instance [14]. Their popularity in signal processing and statistics can be traced back to the development of the basis pursuit method [9] for approximation in redundant dictionaries and the Lasso method [31] for statistical estimation.

The theoretical analysis of the ℓ^1-regularized deconvolution was initiated by Donoho [14]. Assessing the performance of discrete ℓ^1 regularization methods is challenging and requires to take into account both the specific properties of the operator to invert and of the signal that is aimed at being recovered. A popular approach is to assess the recovery of the positions of the non-zero coefficients. This requires to impose a well-conditioning constraint that depends on the signal of interest, as initially introduced by Fuchs [20], and studied in the statistics community under the name of "irrepresentability condition", see [34]. A similar approach is used by Dossal and Mallat in [15] to study the problem of support stability over a discrete grid.

Imposing the exact recovery of the support of the signal to recover might be a too strong assumption. The inverse problem community rather focuses on the L^2 recovery error, which typically leads to a linear convergence rate with respect to the noise amplitude. The seminal paper of Grasmair et al. [21] gives a necessary and sufficient condition for such a convergence, which corresponds to the existence of a non-saturating dual certificate (see Section 2 for a precise definition of certificates). This can be understood as an abstract condition, which is often difficult to check on practical problems such as deconvolution.

Note that the continuous setting adopted in the present paper might be seen as a limit of such discrete problems, and in Section 5, we relate our results to well-known results on discrete grids.

Let us also note that, although we focus here on ℓ^1-based methods, there is a vast literature on various non-linear super-resolution schemes. This includes for instance greedy [26, 25], root finding [3, 11], matrix pencils [13] and compressed sensing [18, 16] approaches.

Inverse problems regularization with measures. Working over a discrete grid makes the mathematical analysis difficult. Following recent proposals [12, 4, 8, 2], we consider here this sparse deconvolution over a continuous domain, i.e. in a grid-free setting. This shift from the traditional discrete domain to a continuous one offers considerable advantages in term of mathematical analysis, allowing for the first time the emergence of almost sharp signal-dependent criteria for stable spikes recovery (see references below). Note that while the corresponding continuous recovery problem is infinite dimensional in nature, it is possible to find its solution using either provably convergent algorithms [4] or root finding methods for ideal low pass filters [8].

Inverse problem regularization over the space of measures is now well understood (see for instance [29, 4]), and requires to perform variational analysis over a non-reflexive Banach space (as in [22]), which leads to some mathematical technicalities. We capitalize on these earlier works to build our analysis of the recovery performance.

Theoretical analysis of deconvolution over the space of measures. For deconvolution from ideal low-pass measurements, the ground-breaking paper [8] shows that it is indeed possible to construct a dual certificate by solving a linear system when the input Diracs are well-separated. This work is further refined in [7] that studies the robustness to noise. In a series of paper [2, 30] the authors study the prediction (i.e. denoising) error using the same dual certificate, but they do not consider the reconstruction error (recovery of the spikes). In our work, we use a different certificate to assess the exact recovery of the spikes when the noise is small enough.

In view of the applications of superresolution, it is crucial to understand the precise location of the recovered Diracs locations when the measurements are noisy. Partial answers to this questions are given in [19] and [1], where it is shown (under different conditions on the signal-to-noise level) that the recovered spikes are clustered tightly around the initial measure's Diracs. In this article, we fully answer the question of the position of the recovered Diracs in the setting where the signal-to-noise ratio is large enough.

1.3 Formulation of the Problem and Contributions.

Let $m_0 = \sum_{i=1}^N a_{0,i} \delta_{x_{0,i}}$ be a discrete measure defined on the torus $\mathbb{T} = \mathbb{R}/\mathbb{Z}$, where $a_0 \in \mathbb{R}^N$ and $x_0 \in \mathbb{T}^N$. We assume we are given some low-pass filtered

3

observation $y_0 = \Phi m_0 \in L^2(\mathbb{T})$. Here Φ denotes a convolution operator with some kernel $\varphi \in C^2(\mathbb{T})$. The observation might be noisy, in which case we are given $y_0 + w = \Phi m_0 + w$, with $w \in L^2(\mathbb{T})$, instead of y_0.

Following [8, 12], we hope to recover m_0 by solving the problem

$$\min_{\Phi m = y_0} \|m\|_{\mathrm{TV}}. \qquad (\mathcal{P}_0(y_0))$$

among all Radon measures, where $\|m\|_{\mathrm{TV}}$ refers to the total variation (defined below) of m. Note that in our setting, the total variation is the natural extension of the ℓ^1 norm of finite dimensional vectors to the setting of Radon measures, and it should not be mistaken for the total variation of functions, which is routinely used to recover signals or images.

We may also consider reconstructing m_0 by solving the following penalized problem for $\lambda > 0$, also known as the Beurling LASSO (see for instance [1]):

$$\min_m \frac{1}{2}\|\Phi m - y_0\|_2^2 + \lambda\|m\|_{\mathrm{TV}}. \qquad (\mathcal{P}_\lambda(y_0))$$

This is especially useful if the observation is noisy, in which case y_0 should be replaced with $y_0 + w$.

Four questions immediately arise:

1. Does the resolution of $(\mathcal{P}_0(y_0))$ for $y_0 = \Phi m_0$ actually recover interesting measures m_0?

2. How close is the solution of $(\mathcal{P}_\lambda(y_0))$ to the solution of $(\mathcal{P}_0(y_0))$ when λ is small enough?

3. How close is the solution of $(\mathcal{P}_\lambda(y_0 + w))$ to the solution of $(\mathcal{P}_\lambda(y_0))$ when both λ and w/λ are small enough?

4. What can be said about the above questions when solving $(\mathcal{P}_\lambda(y_0))$ with measures supported on a fixed finite grid?

The first question is addressed in the landmark paper [8] in the case of ideal low-pass filtering: measures m_0 whose spikes are separated enough are the unique solution of $(\mathcal{P}_0(y_0))$ (for data $y_0 = \Phi m_0$). Several other cases (using observations different from convolutions) are also tackled in [12], particularly in the case of non-negative measures.

The second and third questions receive partial answers in [4, 7, 1, 19]. In [4] it is shown that if the solution of $(\mathcal{P}_0(y_0))$ is unique, the measures recovered by $(\mathcal{P}_\lambda(y_0 + w))$ converge to the solution of $(\mathcal{P}_0(y_0))$ in the sense of the weak-* convergence when $\lambda \to 0$ and $\frac{\|w\|_2^2}{\lambda} \to 0$. In [7], the authors measure the reconstruction error using the L^2 norm of a low-pass filtered version of recovered measures. In [1], error bounds are derived from the amplitudes of the reconstructed measure. In [19], bounds are given in terms of the original measure. However, those works provide little information about the structure of the measures recovered by $(\mathcal{P}_\lambda(y_0 + w))$: *are they made of less spikes than m_0 or, in the*

contrary, do they present lots of parasitic spikes? What happens if one compels the spikes to belong to a finite grid?

The fourth question is of primary importance since most numerical schemes for sparse regularization solve a finite dimensional optimization problem over a fixed discretization grid. Following [8], one can remark that in the noiseless setting, if m_0 is recovered over the continuous domain and if its support is included in the grid, m_0 is also guaranteed to be recovered by the discretized problem. But this is of little interest in practice because the noise is likely to impact in a different manner the discrete problem and the input measure might fall outside the grid locations. Dossal and Mallat in [15] study the stability of the position of the Diracs on the grid, which leads to overly pessimistic conclusions because noise typically forces the spikes to translate over the domain. Studying the convergence of the discretized problem toward the continuous one is thus important to obtain a precise description of the discretized solution. To the best of our knowledge, the work of [2] is the only one to provide some conclusion about this convergence in term of denoising error. No previous work has studied the capability of the discretized problem to estimate in a precise manner the location of the spikes of the input measure.

Contributions. The present paper studies in detail the structure of the recovered measure. For this purpose, we define the minimal L^2-norm certificate. This certificate fully governs the behavior of the regularization when both λ and $\|w\|_2/\lambda$ are small.

Our first contribution is a set of results indicating that the regions of saturation of the certificate (when it reaches $+1$ or -1) are approximately stable when λ and $\|w\|_2/\lambda$ are small enough. This means that the recovered measures are supported closely to the support of the input measure if the latter is identifiable (solution of the noiseless problem $(\mathcal{P}_0(y_0))$).

Our second contribution introduces the *Non Degenerate Source Condition*, which imposes that the second derivative of the minimal-norm certificate does not vanish on the saturation points. Under this condition, we show that for λ and $\|w\|_2/\lambda$ small enough, the reconstructed measure has exactly the same number of spikes as the original measure and that their locations and amplitudes converge to those of the original one.

Our third contribution shows that under the *Non Degenerate Source Condition*, the minimal norm certificate can actually be computed in closed form by simply solving a linear system. This in turn also implies that the errors in the amplitudes and locations decay linearly with respect to the noise level.

Our fourth and last contribution focuses on the regularization over a discrete finite grid, which corresponds to the so-called Lasso or Basis Pursuit Denoising problem. We show that when λ and $\|w\|_2/\lambda$ are small enough, and provided that the Non Degenerate Source Condition holds, the discretized solution is located on pairs of Diracs adjacent to the input Diracs location. This gives a precise description of how the solution to the discretized problem converges to the one of the continuous problem when the stepsize of the grid vanishes.

Throughout the paper, the proposed definitions and results are illustrated in the case of the ideal low-pass filter, showing that the assumptions are actually relevant. Note that the code to reproduce the figures of this article is available online[1].

Outline of the paper. Section 2 defines the framework for the recovery of Radon measures using total variation minimization. We also expose basic results that are used throughout the paper. Section 3 is devoted to the main result of the paper: we define the *Non Degenerate Source Condition* and we show that it implies the robustness of the reconstruction using $(\mathcal{P}_\lambda(y_0 + w))$. In Section 4 we show how the specific dual certificate involved in the *Non Degenerate Source Condition* can be computed numerically by solving a linear system. Lastly, Section 5 focuses on the recovery of measures on a discrete grid.

1.4 Notations

For any Radon measure m defined on $\mathbb{T}$, we denote its support by $\mathrm{Supp}(m)$. If $\mathrm{Supp}(m)$ is a finite set (in which case we say that m is a discrete measure) and $m \neq 0$, then m is of the form $m = \sum_{i=1}^N a_i \delta_{x_i}$, where $N \in \mathbb{N}^*$, $a \in \mathbb{R}^N$, $x \in \mathbb{T}^N$ and $a_i \neq 0$ and $x_i \neq x_j$ for all $1 \leqslant i,j \leqslant N$. In the rest of the paper, we shall write $m = m_{a,x}$ to hint that m has the above decomposition (implying that $a_i \neq 0$ and $x_i \neq x_j$ for all $1 \leqslant i,j \leqslant N$).

We also define the *signed support*:

$$\mathrm{Supp}^\pm m = (\mathrm{Supp}\, m_+) \times \{1\} \cup (\mathrm{Supp}\, m_-) \times \{-1\} \subset \mathbb{T} \times \{+1, -1\}$$

where m_+ (resp. m_-) denotes the positive (resp. negative) part of m. For a discrete measure $m = m_{a,x}$,

$$\begin{aligned}
\mathrm{Supp}^\pm m &= \{(t,v) \in \mathbb{T} \times \{+1, -1\},\ m(\{t\}) \neq 0 \text{ and } \mathrm{sign}\, m(\{t\}) = v\} \\
&= \{(x_i, \mathrm{sign}\, a_i),\ 1 \leqslant i \leqslant N\}.
\end{aligned}$$

We shall consider restrictions of measures and functions to subsets of $\mathbb{T}$. For $m \in \mathcal{M}(\mathbb{T})$ a discrete measure and $J = \{x_1, \ldots, x_k\} \subset \mathbb{T}$ a finite set, we define

$$m|_J = a \in \mathbb{T}^k \quad \text{where} \quad \forall i = 1, \ldots, k, \quad a_i = m(\{x_i\}).$$

For $\eta \in C(\mathbb{T})$ a continuous function defined on $\mathbb{T}$, we define

$$\eta|_J = (\eta(x_j))_{j=1}^k \in \mathbb{T}^k.$$

Given a convolution operator Φ with kernel $t \mapsto \varphi(-t)$, we define $\Phi_x : \mathbb{R}^N \to$

---

[1] https://github.com/gpeyre/2013-FOCM-SparseSpikes/

$L^2(\mathbb{T})$ (resp. Φ_x', Φ_x'') by

$$\forall a \in \mathbb{R}^N, \ \Phi_x(a) = \Phi(m_{a,x}) = \sum_{i=1}^{N} a_i \varphi(x_i - \cdot), \tag{1}$$

$$\Phi_x'(a) = (\Phi_x(a))' = \sum_{i=1}^{N} a_i \varphi'(x_i - \cdot), \tag{2}$$

$$\Phi_x''(a) = (\Phi_x(a))'' = \sum_{i=1}^{N} a_i \varphi''(x_i - \cdot). \tag{3}$$

We define

$$\Gamma_x = (\Phi_x, \Phi_x') : (u,v) \in \mathbb{R}^N \times \mathbb{R}^N \mapsto \Phi_x u + \Phi_x' v \in \mathrm{L}^2(\mathbb{T}), \tag{4}$$
$$\Gamma_x' = (\Phi_x', \Phi_x'') : (u,v) \in \mathbb{R}^N \times \mathbb{R}^N \mapsto \Phi_x' u + \Phi_x'' v \in \mathrm{L}^2(\mathbb{T}). \tag{5}$$

Eventually, in order to study small noise regimes, we shall consider domains D_{α,λ_0}, for $\alpha > 0$, $\lambda_0 > 0$, where:

$$D_{\alpha,\lambda_0} = \left\{ (\lambda, w) \in \mathbb{R}_+ \times L^2(\mathbb{T}) \ ; \ 0 \leqslant \lambda \leqslant \lambda_0 \quad \text{and} \quad \|w\|_2 \leqslant \alpha\lambda \right\}. \tag{6}$$

2 Preliminaries

In this section, we precise the framework and we state the basic results needed in the next sections. We refer to [5] for aspects regarding functional analysis and to [17] as far as duality in optimization is concerned.

2.1 Topology of Radon Measures

Since $\mathbb{T}$ is compact, the space of Radon measures $\mathcal{M}(\mathbb{T})$ can be defined as the dual of the space $C(\mathbb{T})$ of continuous functions on $\mathbb{T}$, endowed with the uniform norm. It is naturally a Banach space when endowed with the dual norm (also known as the total variation), defined as

$$\forall m \in \mathcal{M}(\mathbb{T}), \quad \|m\|_{\mathrm{TV}} = \sup \left\{ \int \psi \mathrm{d}m \ ; \ \psi \in C(\mathbb{T}), \|\psi\|_\infty \leqslant 1 \right\}. \tag{7}$$

In that case, the dual of $\mathcal{M}(\mathbb{T})$ is a complicated space, and it is strictly larger than $C(\mathbb{T})$ as $C(\mathbb{T})$ is not reflexive.

However, if we endow $\mathcal{M}(\mathbb{T})$ with its weak-* topology (i.e. the coarsest topology such that the elements of $C(\mathbb{T})$ define continuous linear forms on $\mathcal{M}(\mathbb{T})$), then $\mathcal{M}(\mathbb{T})$ is a locally convex space whose dual is $C(\mathbb{T})$.

In the following, we endow $C(\mathbb{T})$ (respectively $\mathcal{M}(\mathbb{T})$) with its weak (respectively its weak-*) topology so that both have symmetrical roles: one is the dual of the other, and conversely. Moreover, since $C(\mathbb{T})$ is separable, the set $\{m \in \mathcal{M}(\mathbb{T}) \ ; \ \|m\|_{\mathrm{TV}} \leqslant 1\}$ endowed with the weak-* topology is metrizable.

7

Given a function $\varphi \in C^2(\mathbb{T}, \mathbb{R})$, we define an operator $\Phi : \mathcal{M}(\mathbb{T}) \to L^2(\mathbb{T})$ as

$$\forall\, m \in \mathcal{M}(\mathbb{T}), \quad \Phi(m) : t \mapsto \int_{\mathbb{T}} \varphi(x-t) \mathrm{d}m(x).$$

It can be shown using Fubini's theorem that Φ is weak-* to weak continuous. Moreover, its adjoint operator $\Phi^* : L^2(\mathbb{T}) \to C(\mathbb{T})$ is defined as

$$\forall\, y \in L^2(\mathbb{T}), \quad \Phi^*(y) : t \mapsto \int_{\mathbb{T}} \varphi(t-x) y(x) \mathrm{d}x.$$

2.2 Subdifferential of the Total Variation

It is clear from the definition of the total variation in (7) that it is convex lower semi-continuous with respect to the weak-* topology. Its subdifferential is defined as

$$\partial \|m\|_{\mathrm{TV}} = \left\{ \eta \in C(\mathbb{T}) \;;\; \forall \tilde{m} \in \mathcal{M}(\mathbb{T}), \|\tilde{m}\|_{\mathrm{TV}} \geqslant \|m\|_{\mathrm{TV}} + \int \eta \, \mathrm{d}(\tilde{m} - m) \right\}, \quad (8)$$

for any $m \in \mathcal{M}(\mathbb{T})$ such that $\|m\|_{\mathrm{TV}} < +\infty$.

Since the total variation is a sublinear function, its subgradient has a special structure. One may show (see Proposition 12 in Appendix A) that

$$\partial \|m\|_{\mathrm{TV}} = \left\{ \eta \in C(\mathbb{T}) \;;\; \|\eta\|_\infty \leqslant 1 \quad \text{and} \quad \int \eta \, \mathrm{d}m = \|m\|_{\mathrm{TV}} \right\}. \quad (9)$$

In particular, when m is a measure with finite support, i.e. $m = \sum_{i=1}^N a_i \delta_{x_i}$ for some $N \in \mathbb{N}$, with $(a_i)_{1 \leqslant i \leqslant N} \in (\mathbb{R}^*)^N$ and distinct $(x_i)_{1 \leqslant i \leqslant N} \in \mathbb{T}^N$

$$\partial \|m\|_{\mathrm{TV}} = \{\eta \in C(\mathbb{T}) \;;\; \|\eta\|_\infty \leqslant 1 \text{ and } \forall\, i = 1, \ldots, N, \; \eta(x_i) = \mathrm{sign}(a_i)\}. \quad (10)$$

2.3 Primal and Dual Problems

Given an observation $y_0 = \Phi m_0 \in L^2(\mathbb{T})$ for some $m_0 \in \mathcal{M}(\mathbb{T})$, we consider reconstructing m_0 by solving either the relaxed problem for $\lambda > 0$

$$\min_{m \in \mathcal{M}(\mathbb{T})} \frac{1}{2} \|\Phi(m) - y_0\|_2^2 + \lambda \|m\|_{\mathrm{TV}}, \qquad (\mathcal{P}_\lambda(y_0))$$

or the constrained problem

$$\min_{\Phi(m) = y_0} \|m\|_{\mathrm{TV}}. \qquad (\mathcal{P}_0(y_0))$$

If m_0 is the unique solution of $(\mathcal{P}_0(y_0))$, we say that m_0 is *identifiable*.

In the case where the observation is noisy (i.e. the observation y_0 is replaced with $y_0 + w$ for $w \in L^2(\mathbb{T})$), we attempt to reconstruct m_0 by solving $\mathcal{P}_\lambda(y_0 + w)$ for a well-chosen value of $\lambda > 0$.

Existence of solutions for $(\mathcal{P}_\lambda(y_0))$ is shown in [4], and existence of solutions for $(\mathcal{P}_0(y_0))$ can be checked using the direct method of the calculus of variations (recall that for $(\mathcal{P}_0(y_0))$, we assume that the observation is $y_0 = \Phi m_0$).

A straightforward approach to studying the solutions of Problem $(\mathcal{P}_\lambda(y_0))$ is then to apply Fermat's rule: a discrete measure $m = m_{a,x} = \sum_{i=1}^{N} a_i \delta_{x_i}$ is a solution of $(\mathcal{P}_\lambda(y_0))$ if and only if there exists $\eta \in C(\mathbb{T})$ such that

$$\Phi^*(\Phi m - y_0) + \lambda \eta = 0,$$

with $\|\eta\|_\infty \leqslant 1$ and $\eta(x_i) = \text{sign}(a_i)$ for $1 \leqslant i \leqslant N$.

Another source of information for the study of Problems $(\mathcal{P}_\lambda(y_0))$ and $(\mathcal{P}_0(y_0))$ is given by their associated dual problems. In the case of the ideal low-pass filter, this approach is also the key to the numerical algorithms used in [2, 8, 1]: the dual problem can be recast into a finite-dimensional problem.

The Fenchel dual problem to $(\mathcal{P}_\lambda(y_0))$ is given by

$$\max_{\|\Phi^*p\|_\infty \leqslant 1} \langle y_0, \, p \rangle - \frac{\lambda}{2}\|p\|_2^2, \qquad (\mathcal{D}_\lambda(y_0))$$

which may be reformulated as a projection on a closed convex set (see [4, 1])

$$\min_{\|\Phi^*p\|_\infty \leqslant 1} \left\| \frac{y_0}{\lambda} - p \right\|_2^2. \qquad (\mathcal{D}_\lambda'(y))$$

This formulation immediately yields existence and uniqueness of a solution to $(\mathcal{D}_\lambda(y_0))$.

The dual problem to $(\mathcal{P}_0(y_0))$ is given by

$$\sup_{\|\Phi^*p\|_\infty \leqslant 1} \langle y_0, \, p \rangle. \qquad (\mathcal{D}_0(y_0))$$

Contrary to $(\mathcal{D}_\lambda(y_0))$, the existence of a solution to $(\mathcal{D}_0(y_0))$ is not always guaranteed, so that in the following (see Definition 5) we make this assumption.

Existence is guaranteed when for instance $\text{Im}\,\Phi^*$ is finite-dimensional (as is the case in the framework of [8]). If a solution to $(\mathcal{D}_0(y_0))$ exists, the unique solution of $(\mathcal{D}_\lambda(y_0))$ converges to a certain solution of $(\mathcal{D}_0(y_0))$ for $\lambda \to 0^+$ as shown in Proposition 1 below.

2.4 Dual Certificates

The strong duality between $(\mathcal{P}_\lambda(y_0))$ and $(\mathcal{D}_\lambda(y_0))$ is proved in [4, Prop. 2] by seeing $(\mathcal{D}_\lambda'(y))$ as a predual problem for $(\mathcal{P}_\lambda(y_0))$. As a consequence, both problems have the same value and any solution m_λ of $(\mathcal{P}_\lambda(y_0))$ is linked with the unique solution p_λ of $(\mathcal{D}_\lambda(y_0))$ by the extremality condition

$$\begin{cases} \Phi^* p_\lambda \in \partial \|m_\lambda\|_{\text{TV}}, \\ -p_\lambda = \frac{1}{\lambda}(\Phi m_\lambda - y_0). \end{cases} \qquad (11)$$

Moreover, given a pair $(m_\lambda, p_\lambda) \in \mathcal{M}(\mathbb{T}) \times L^2(\mathbb{T})$, if relations (11) hold, then m_λ is a solution to Problem $(\mathcal{P}_\lambda(y_0))$ and p_λ is the unique solution to Problem $(\mathcal{D}_\lambda(y_0))$.

As for $(\mathcal{P}_0(y_0))$, a proof of strong duality is given in Appendix A (see Proposition 13). If a solution $p^\star$ to $(\mathcal{D}_0(y_0))$ exists, then it is linked to any solution $m^\star$ of $(\mathcal{P}_0(y_0))$ by

$$\Phi^* p^\star \in \partial \|m^\star\|_{\mathrm{TV}}, \tag{12}$$

and similarly, given a pair $(m^\star, p^\star) \in \mathcal{M}(\mathbb{T}) \times L^2(\mathbb{T})$, if relation (12) hold, then $m^\star$ is a solution to Problem $(\mathcal{P}_0(y_0))$ and $p^\star$ is a solution to Problem $(\mathcal{D}_0(y_0)))$.

Since finding $\eta = \Phi^* p^\star$ which satisfies (12) gives a quick proof that $m^\star$ is a solution of $(\mathcal{P}_0(y_0))$, we call η a *dual certificate* for $m^\star$. We may also use a similar terminology for $\eta_\lambda = \Phi^* p_\lambda$ and Problem $(\mathcal{P}_\lambda(y_0))$.

In general, dual certificates for $(\mathcal{P}_0(y_0))$ are not unique, but we consider in the following definition a specific one, which is crucial for our analysis.

Definition 1 (Minimal-norm certificate). *When it exists, the minimal-norm dual certificate associated with $(\mathcal{P}_0(y_0))$ is defined as $\eta_0 = \Phi^* p_0$ where $p_0 \in L^2(\mathbb{T})$ is the solution of $(\mathcal{D}_0(y_0))$ with minimal norm, i.e.*

$$\eta_0 = \Phi^* p_0, \quad where \quad p_0 = \operatorname*{argmin}_p \ \{\|p\|_2 \ ; \ p \ is \ a \ solution \ of \ (\mathcal{D}_0(y_0))\}. \tag{13}$$

Observe that in the above definition, p_0 is well-defined provided there exists a solution to Problem $(\mathcal{D}_0(y_0))$, since p_0 is then the projection of 0 onto the non-empty closed convex set of solutions. Moreover, in view of the extremality conditions (12), given any solution $m^\star$ to $(\mathcal{P}_0(y_0))$, it may be expressed as

$$p_0 = \operatorname*{argmin}_p \ \{\|p\|_2 \ ; \ \Phi^* p \in \partial \|m^\star\|_{\mathrm{TV}}\}. \tag{14}$$

Proposition 1 (Convergence of dual certificates). *Let p_λ be the unique solution of Problem $(\mathcal{D}_\lambda(y_0))$, and p_0 be the solution of Problem $(\mathcal{D}_0(y_0))$ with minimal norm defined in (13). Then*

$$\lim_{\lambda \to 0^+} p_\lambda = p_0 \quad for \ the \ L^2 \ strong \ topology.$$

Moreover the dual certificates $\eta_\lambda = \Phi^* p_\lambda$ for Problem $(\mathcal{P}_\lambda(y_0))$ converge to the minimal norm certificate $\eta_0 = \Phi^* p_0$. More precisely,

$$\forall k \in \{0, 1, 2\}, \quad \lim_{\lambda \to 0^+} \eta_\lambda^{(k)} = \eta_0^{(k)}, \tag{15}$$

in the sense of the uniform convergence.

Proof. Let p_λ be the unique solution of $(\mathcal{D}_\lambda(y_0))$. By optimality of p_λ (resp. p_0) for $(\mathcal{D}_\lambda(y_0))$ (resp. $(\mathcal{D}_0(y_0)))$

$$\langle y_0, p_\lambda \rangle - \lambda \|p_\lambda\|_2^2 \geqslant \langle y_0, p_0 \rangle - \lambda \|p_0\|_2^2, \tag{16}$$

$$\langle y_0, p_0 \rangle \geqslant \langle y_0, p_\lambda \rangle. \tag{17}$$

As a consequence $\|p_0\|_2^2 \geqslant \|p_\lambda\|_2^2$ for all $\lambda > 0$.

Now, let $(\lambda_n)_{n \in \mathbb{N}}$ be any sequence of positive parameters converging to 0. The sequence p_{λ_n} being bounded in $L^2(\mathbb{T})$, we may extract a subsequence (denoted $\lambda_{n'}$) such that $p_{\lambda_{n'}}$ weakly converges to some $p^\star \in L^2(\mathbb{T})$. Passing to the limit in (16), we get $\langle y_0, p^\star \rangle \geqslant \langle y_0, p_0 \rangle$. Moreover, $\Phi^* p_{\lambda_n}$ weakly converges to $\Phi^* p^\star$ in $C(\mathbb{T})$, so that $\|\Phi^* p^\star\|_\infty \leqslant \liminf_{n'} \|\Phi^* p_{\lambda_{n'}}\|_\infty \leqslant 1$, and $p^\star$ is therefore a solution of $(\mathcal{D}_0(y_0))$.

But one has
$$\|p^\star\|_2 \leqslant \liminf_{n'} \|p_{\lambda_{n'}}\|_2 \leqslant \|p_0\|_2,$$

hence $p^\star = p_0$ and in fact $\lim_{n' \to +\infty} \|p_{\lambda_{n'}}\|_2 = \|p_0\|_2$. As a consequence, $p_{\lambda_{n'}}$ converges to p_0 for the $L^2(\mathbb{T})$ strong topology as well. This being true any sequence $\lambda_n \to 0^+$, we get the result claimed for p_λ: assume by contradiction that there exists $\varepsilon_0 > 0$ and a sequence $\lambda_n \searrow 0$ such that $\|p_0 - p_{\lambda_n}\|_2 \geqslant \varepsilon_0$ for all $n \in \mathbb{N}$. By the above argument we may extract a subsequence $\lambda_{n'}$ which converges towards p_0, which contradicts $\|p_0 - p_{\lambda'_n}\|_2 \geqslant \varepsilon_0$. Hence $\lim_{\lambda \to 0} p_\lambda = p_0$ strongly in L^2.

It remains to prove the convergence of the dual certificates. Observing that $\eta_\lambda^{(k)}(t) = \int \varphi^{(k)}(t-x) p_\lambda(x) \mathrm{d}x$, we get

$$|\eta_\lambda^{(k)}(t) - \eta_0^{(k)}(t)| = \left| \int_{\mathbb{T}} \varphi^{(k)}(t-x)(p_\lambda - p_0)(x)\mathrm{d}x \right|$$

$$\leqslant \sqrt{\int_{\mathbb{T}} |\varphi^{(k)}(t-x)|^2 \mathrm{d}x} \sqrt{\int_{\mathbb{T}} |(p_\lambda - p_0)(x)|^2 \mathrm{d}x}$$

$$\leqslant C\|p_\lambda - p_0\|_2,$$

where $C > 0$ does not depend on t nor k, hence the uniform convergence. □

2.5 Application to the ideal Low-pass filter

In this paragraph, we apply the above duality results to the particular case of the Dirichlet kernel, defined as

$$\varphi(t) = \sum_{k=-f_c}^{f_c} e^{2i\pi kt} = \frac{\sin\left((2f_c + 1)\pi t\right)}{\sin(\pi t)}. \tag{18}$$

It is well known that in this case the spaces $\mathrm{Im}\,\Phi$ and $\mathrm{Im}\,\Phi^*$ are finite-dimensional, being the space of real trigonometric polynomials with degree less than or equal to f_c.

We first check that a solution to $(\mathcal{D}_0(y_0))$ always exists. As a consequence, given any measure m_0, the minimal norm certificate is well defined.

Proposition 2 (Existence of p_0). *Let $m_0 \in \mathcal{M}(\mathbb{T})$ and $y_0 = \Phi m_0 \in L^2(\mathbb{T})$. There exists a solution of $(\mathcal{D}_0(y_0))$. As a consequence, $p_0 \in L^2(\mathbb{T})$ is well defined.*

11

Proof. We rewrite $(\mathcal{D}_0(y_0))$ as

$$\sup_{\|\eta\|_\infty \leqslant 1, \eta \in \operatorname{Im} \Phi^*} \langle m_0, \eta \rangle.$$

Let $(\eta_n)_{n \in \mathbb{N}}$ be any maximizing sequence. Then $(\eta_n)_{n \in \mathbb{N}}$ is bounded in the finite-dimensional space of trigonometric polynomials with degree f_c or less. We may extract a subsequence converging to $\eta^\star \in C(\mathbb{T})$. But $\|\eta^\star\|_\infty \leqslant 1$ and $\eta^\star \in \operatorname{Im} \Phi^*$, so that $\eta^\star = \Phi^* p^\star$ for some $p^\star$ solution of $(\mathcal{D}_0(y_0))$. □

A striking result of [8] is that discrete measures are identifiable provided that their support is separated enough, i.e. $\Delta(m_0) \geqslant \frac{C}{f_c}$ for some $C > 0$, where $\Delta(m_0)$ is the so-called minimum separation distance.

Definition 2 (Minimum separation). *The minimum separation of the support of a discrete measure m is defined as*

$$\Delta(m) = \inf_{(t,t') \in \operatorname{Supp}(m)} |t - t'|,$$

where $|t - t'|$ is the distance on the torus between t and $t' \in \mathbb{T}$, and we assume $t \neq t'$.

In [8] it is proved that $C \leqslant 2$ for complex measures (i.e. of the form $m_{a,x}$ for $a \in \mathbb{C}^N$ and $x \in \mathbb{T}^N$) and $C \leqslant 1.87$ for real measures (i.e. of the form $m_{a,x}$ for $a \in \mathbb{R}^N$ and $x \in \mathbb{T}^N$). Extrapolating from numerical simulations on a finite grid, the authors conjecture that for complex measures, one has $C \geqslant 1$. In this section we apply results from Section 2.4 to show that for real measures, necessarily $C \geqslant \frac{1}{2}$.

We rely on the following theorem, proved by P. Turán [32].

Theorem 1 (Turán). *Let $P(z)$ be a non trivial polynomial of degree n such that $|P(1)| = \max_{|z|=1} |P(z)|$. Then for any root z_0 of P on the unit circle, $|\arg(z_0)| \geqslant \frac{\pi}{n}$. Moreover, if $|\arg(z_0)| = \frac{\pi}{n}$, then $P(z) = c(1 + z^n)$ for some $c \in \mathbb{C}^*$.*

From this theorem we derive necessary conditions for measures that can be reconstructed by $(\mathcal{P}_0(y_0))$.

Corollary 1 (Non identifiable measures). *There exists a discrete measure m_0 with $\Delta(m_0) = \frac{1}{2f_c}$ such that m_0 is not a solution of $(\mathcal{P}_0(y_0))$ for $y_0 = \Phi m_0$.*

Proof. Let $m_0 = \delta_{-\frac{1}{2f_c}} + \delta_0 - \delta_{\frac{1}{2f_c}}$, assume by contradiction that m is a solution of $(\mathcal{P}_0(y_0))$, and let $\eta \in C(\mathbb{T})$ be an associated dual certificate (which exists since $\operatorname{Im} \Phi^*$ is finite-dimensional). Then necessarily $\eta(-\frac{1}{2f_c}) = \eta(0) = 1$ and $\eta(\frac{1}{2f_c}) = -1$ and by the intermediate value theorem, there exists $t_0 \in (0, \frac{1}{2f_c})$ such that $\eta(t_0) = 0$.

Writing $\eta(t) = \sum_{k=-f_c}^{f_c} d_k e^{2i\pi kt}$, the polynomial $P(z) = \sum_{k=0}^{2f_c} d_{k-f_c} z^k$ satisfies $P(1) = 1 = \sup_{|z|=1} |P(z)| = |P(e^{\frac{2i\pi}{2f_c}})|$, and $P(e^{2i\pi t_0}) = 0$.

By Theorem 1, we cannot have $|2\pi t_0 - 0| < \frac{\pi}{2f_c}$ nor $|2\pi t_0 - \frac{2\pi}{2f_c}| < \frac{\pi}{2f_c}$, hence $t_0 = \frac{1}{4f_c}$ and $P(z) = c(1 + z^{2f_c})$, so that $\eta(t) = \cos(2\pi f_c t)$. But this implies $\eta(-\frac{1}{2f_c}) = -1$, which contradicts the optimality of η. $\qquad\square$

In a similar way, we may also deduce the following corollary.

Corollary 2 (Opposite spikes separation). *Let $m^\star \in \mathcal{M}(\mathbb{T})$ be any discrete measure solution of Problem $\mathcal{P}_\lambda(y_0 + w)$ or $\mathcal{P}_0(y_0)$ where $y_0 = \Phi m_0$ for any data $m_0 \in \mathcal{M}(\mathbb{T})$ and any noise $w \in L^2(\mathbb{T})$. If there exists $x_0^\star \in \mathbb{T}$ (resp. $x_1^\star \in \mathbb{T}$) such that $m^\star(\{x_0^\star\}) > 0$ (resp. $m^\star(\{x_1^\star\}) < 0$), then $|x_0^\star - x_1^\star| \geqslant \frac{1}{2f_c}$.*

3   Noise Robustness

This section is devoted to the study of the behavior of solutions to $\mathcal{P}_\lambda(y_0 + w)$ for small values of λ and $\|w\|$. In order to study such regimes, as already defined in (6), we consider sets of the form

$$D_{\alpha,\lambda_0} = \left\{ (\lambda, w) \in \mathbb{R}_+ \times L^2(\mathbb{T}) \; ; \; 0 \leqslant \lambda \leqslant \lambda_0 \quad \text{and} \quad \|w\|_2 \leqslant \alpha\lambda \right\},$$

for $\alpha > 0$ and $\lambda_0 > 0$.

First, we introduce the notion of extended support of a measure. Then we show that this concept governs the structure of solutions at small noise regime. After introducing the Non Degenerate Source Condition, we state the main result of the paper, *i.e.* that under this assumption, the solutions of $\mathcal{P}_\lambda(y_0 + w)$ have the same number of spikes as the original measure, and that these spikes converge smoothly to those of the original measure.

3.1   Extended signed support

Our first step in understanding the behavior of solutions to $\mathcal{P}_\lambda(y_0 + w)$ at low noise regime is to introduce the notion of extended signed support.

Definition 3 (Extended signed support). *Let $m_0 \in \mathcal{M}(\mathbb{T})$ such that there exists a solution to $(\mathcal{D}_0(y_0))$ (where as usual $y_0 = \Phi m_0$), and let $\eta_0 \in C(\mathbb{T})$ be the associated minimal norm certificate.*

The extended support of m_0 is defined as:

$$\text{Ext}(m_0) = \{t \in \mathbb{T} \; ; \; \eta_0(t) = \pm 1\}, \tag{19}$$

and the extended signed support of m_0 as:

$$\text{Ext}^\pm(m_0) = \{(t, v) \in \mathbb{T} \times \{+1, -1\} \; ; \; \eta_0(t) = v\}. \tag{20}$$

Notice that $\text{Ext}\, m_0$ and $\text{Ext}^\pm m_0$ actually depend on $y_0 = \Phi m_0$ rather than on m_0 itself. For any measure $m_0 \in \mathcal{M}(\mathbb{T})$, the (signed) support and the extended (signed) support of m_0 are in general not related. Yet, from the optimality conditions (12) we observe:

Proposition 3. *Let $m_0 \in \mathcal{M}(\mathbb{T})$ and $y_0 = \Phi m_0$ such that there exists a solution to $(\mathcal{D}_0(y_0))$. Then:*

- *m_0 is a solution to $(\mathcal{P}_0(y_0))$ if and only if $\mathrm{Supp}^{\pm} m_0 \subset \mathrm{Ext}^{\pm} m_0$.*

- *In any case, if $\Phi_{\mathrm{Ext}\, m_0}$ has full rank, the solution to $(\mathcal{P}_0(y_0))$ is unique.*

Here, following the notation (1), we have denoted by $\Phi_{\mathrm{Ext}\, m_0}$ the restriction of Φ to the space of measures with support in $\mathrm{Ext}\, m_0$. The link between Proposition 3 and the source condition [6] is discussed in Section 3.3

3.2 Local behavior of the support

In this paragraph, we focus on the local properties of the support of solutions to $\mathcal{P}_\lambda(y_0 + w)$ at low noise regime. As usual, we denote $y_0 = \Phi m_0$ for some $m_0 \in \mathcal{M}(\mathbb{T})$. For now, we make as few assumptions as possible on m_0. In particular, we do not assume that $\Phi_{\mathrm{Ext}\, m_0}$ has full rank. Any solution to $\mathcal{P}_\lambda(y_0 + w)$ (which is not necessarily unique) is denoted by $\tilde{m}_\lambda$.

Lemma 1. *Assume that there exists a solution to $(\mathcal{D}_0(y_0))$ and let $\varepsilon > 0$. Then there exists $\alpha > 0$, $\lambda_0 > 0$ such that for all $(\lambda, w) \in D_{\alpha, \lambda_0}$,*

$$\mathrm{Supp}^{\pm} \tilde{m}_\lambda \subset \left(\mathrm{Ext}^{\pm} m_0\right) \oplus \left((-\varepsilon, +\varepsilon) \times \{0\}\right), \tag{21}$$

where given two sets A and B, $A \oplus B = \{a + b\,;\; a \in A, b \in B\}$ denotes their Minkowski sum.

In particular, if $\mathrm{Ext}\, m_0$ consists in isolated points $x_{0,1}, \ldots x_{0,N}$, Lemma 1 states that all the mass of $\tilde{m}_\lambda$ is concentrated in boxes $(x_{i,0} - \varepsilon, x_{i,0} + \varepsilon)$, where $\varepsilon \to 0$ when $\lambda, \|w\| \to 0$. Moreover, in each box, $\tilde{m}_\lambda$ has the sign of $\eta_0(x_{0,i})$.

Also, if $\mathrm{Ext}^{\pm} m_0 = \emptyset$ (i.e. $y = 0$), we see that $\tilde{m}_\lambda = 0$ for λ and $\frac{\|w\|_2}{\lambda}$ small enough (in fact, any $\lambda_0 > 0$ and $\alpha = \frac{1}{\|\Phi^*\|_{2,\infty}}$ suffices, as can be seen from (11)).

Proof. We split the proof in several parts.

Behavior of the minimal norm certificate. Let us consider the sets:

$$\mathrm{Ext}^+ = \{t \in \mathbb{T}\,;\; \eta_0(t) = 1\}, \quad \mathrm{Ext}^- = \{t \in \mathbb{T}\,;\; \eta_0(t) = -1\},$$
$$\mathrm{Ext}^{+,\varepsilon} = \mathrm{Ext}^+ \oplus(-\varepsilon, \varepsilon), \quad \mathrm{Ext}^{-,\varepsilon} = \mathrm{Ext}^- \oplus(-\varepsilon, \varepsilon).$$

From the uniform continuity of η_0, for ε small enough, $\eta_0 > \frac{1}{2}$ in $\mathrm{Ext}^{+,\varepsilon}$ and $\eta_0 < -\frac{1}{2}$ in $\mathrm{Ext}^{-,\varepsilon}$, so that $\mathrm{Ext}^{+,\varepsilon} \cap \mathrm{Ext}^{-,\varepsilon} = \emptyset$.

If $\mathrm{Ext}^{+,\varepsilon} \cup \mathrm{Ext}^{-,\varepsilon} \subsetneq \mathbb{T}$, the set $K_\varepsilon = \mathbb{T} \setminus \left(\mathrm{Ext}^{+,\varepsilon} \cup \mathrm{Ext}^{-,\varepsilon}\right)$ being compact, $\sup_{K_\varepsilon} |\eta_0| < 1$. We define $r = 1 - \sup_{K_\varepsilon} |\eta_0|$.

If $\mathrm{Ext}^{+,\varepsilon} \cup \mathrm{Ext}^{-,\varepsilon} = \mathbb{T}$, the connectedness of $\mathbb{T}$ implies that $\mathrm{Ext}^{+,\varepsilon} = \mathbb{T}$ and $\mathrm{Ext}^{-,\varepsilon} = \emptyset$, or conversely. In that case we define $r = 1$.

In any case, we see that for all $g \in C(\mathbb{T})$, if $\|g - \eta_0\|_\infty < r$, then

$$\{t \in \mathbb{T}\,;\; g(t) = 1\} \subset \mathrm{Ext}^{+,\varepsilon} \quad \text{and} \quad \{t \in \mathbb{T}\,;\; g(t) = -1\} \subset \mathrm{Ext}^{-,\varepsilon}. \tag{22}$$

Variations of dual certificates. Let p_λ be the solution of the noiseless problem $(\mathcal{D}_\lambda(y_0))$ and $\tilde{p}_\lambda$ be the solution of the noisy dual problem $\mathcal{D}_\lambda(y_0 + w)$ for $w \in L^2(\mathbb{T})$. Since the mapping $\frac{y_0}{\lambda} \mapsto p_\lambda$ is a projection onto a convex set (see $(\mathcal{D}'_\lambda(y))$), it is non-expansive, *i.e.*

$$\|p_\lambda - \tilde{p}_\lambda\|_2 \leqslant \frac{\|w\|_2}{\lambda}. \tag{23}$$

As a consequence, if $\eta_\lambda = \Phi^* p_\lambda$ (resp. $\tilde{\eta}_\lambda = \Phi^* \tilde{p}_\lambda$) is the dual certificate of the noiseless (resp. noisy) problem, we have

$$\|\eta_\lambda - \tilde{\eta}_\lambda\|_\infty \leqslant M \frac{\|w\|_2}{\lambda} \tag{24}$$

for some $M > 0$ (in fact $M = \sqrt{\int_{\mathbb{T}} |\varphi(t)|^2 dt} = \|\Phi^*\|_{\infty,2}$).

From now on, we set $\alpha = \frac{r}{2M}$ and we impose $\frac{\|w\|_2}{\lambda} \leqslant \alpha$. Writing

$$\|\eta_0 - \tilde{\eta}_\lambda\|_\infty \leqslant \|\eta_0 - \eta_\lambda\|_\infty + \|\eta_\lambda - \tilde{\eta}_\lambda\|_\infty,$$
$$\leqslant \|\eta_0 - \eta_\lambda\|_\infty + \frac{r}{2},$$

we see using Proposition 1 that for λ small enough $\tilde{\eta}_\lambda$ satisfies (22).

Structure of the reconstructed measure. By (22) for $g = \tilde{\eta}_\lambda$ and using the extremality conditions we obtain that $|\tilde{m}_\lambda|(K_\varepsilon) = 0$ and that $\tilde{m}_\lambda$ (resp. $-\tilde{m}_\lambda$) is non-negative in $\mathrm{Ext}^{+,\varepsilon}$ (resp. $\mathrm{Ext}^{-,\varepsilon}$). Indeed, the extremality conditions impose that $\tilde{\eta}_\lambda = \mathrm{sign} \frac{d\tilde{m}_\lambda}{d|\tilde{m}_\lambda|}$, $\tilde{m}_\lambda$-almost everywhere, hence the claimed result. $\qquad\square$

Lemma 1 does not make any assumption on the local structure of $\mathrm{Ext}^{\pm} m_0$, and does not provide any information on the local structure of $\tilde{m}_\lambda$ either (it might even not be discrete). If we assume that $\eta_0''(x) \neq 0$ for some $x \in \mathrm{Ext}\, m_0$, then the reconstructed measure has at most one spike in the neighborhood of x.

Lemma 2. *Assume that there exists a solution to $(\mathcal{D}_0(y_0))$ and that $\eta_0''(x) \neq 0$ for some $x \in \mathrm{Ext}\, m_0$. Then for $\varepsilon > 0$ small enough, there exists $\alpha > 0$, $\lambda_0 > 0$ such that for all $(\lambda, w) \in D_{\alpha,\lambda_0}$, the restriction of $\tilde{m}_\lambda$ to $(x - \varepsilon, x + \varepsilon)$ is*

- *either the null measure,*

- *or of the form $\tilde{a}_{\lambda,w} \delta_{\tilde{x}_{\lambda,w}}$ where $\mathrm{sign}\, \tilde{a}_{\lambda,w} = \eta_0(x)$ and $\tilde{x}_{\lambda,w} \in (x - \varepsilon, x + \varepsilon)$.*

If, in addition, m_0 is identifiable and $|m_0|((x - \varepsilon, x + \varepsilon)) \neq 0$, only the second case may happen.

Proof. The proof follows the same steps as those of Lemma 1.

Behavior of the minimal norm certificate. First, observe that if $\eta_0''(x) \neq 0$ and $\eta_0(x) = 1$ (resp. -1) for $x \in \mathrm{Ext}\, m_0$, then $\eta_0''(x) < 0$ (resp. > 0). As a consequence, x is an isolated point of $\mathrm{Ext}\, m_0$. For $\varepsilon > 0$ small enough, $\mathrm{Ext}\, m_0 \cap (x - \varepsilon, x + \varepsilon) = \{x\}$ and $|\eta_0''(t)| \geqslant \frac{|\eta_0''(x)|}{2} > 0$ for all $t \in (x - \varepsilon, x + \varepsilon)$.

15

Variations of dual certificates. From (23), we infer that

$$\|\eta_\lambda'' - \tilde{\eta}_\lambda''\|_\infty \leqslant M \frac{\|w\|_2}{\lambda} \tag{25}$$

with $M > 0$ (here $M = \sqrt{\int_\mathbb{T} |\varphi''(t)|^2 \mathrm{d}t} = \|(\Phi'')^*\|_{\infty,2}$).

We set $\alpha = \frac{r}{2M}$ with $r = \frac{|\eta_0''(x)|}{2}$ and we impose $\frac{\|w\|_2}{\lambda} \leqslant \alpha$, so that

$$\|\eta_0'' - \tilde{\eta}_\lambda''\|_\infty \leqslant \|\eta_0'' - \eta_\lambda''\|_\infty + \|\eta_\lambda'' - \tilde{\eta}_\lambda''\|_\infty,$$
$$\leqslant \|\eta_0'' - \eta_\lambda''\|_\infty + \frac{r}{2},$$

thus $\|\eta_0'' - \tilde{\eta}_\lambda''\|_\infty < \frac{|\eta_0''(x)|}{2}$ for λ small enough.

Structure of the reconstructed measure. From the above inequality, we know that $\tilde{\eta}_\lambda$ is strictly concave (resp. strictly convex) in $(x - \varepsilon, x + \varepsilon)$. As a result, there is at most one point $\tilde{x}_{\lambda,w}$ in $(x - \varepsilon, x + \varepsilon)$ such that $\tilde{\eta}_\lambda(\tilde{x}_{\lambda,w}) = 1$ (resp. -1).

If m_0 is identifiable, it remains to prove that there is indeed one spike in $(x - \varepsilon, x + \varepsilon)$. This is obtained by relying on a result by Bredies and Pikkarainen [4] which is an application of [22, Th. 3.5]. It guarantees that $\tilde{m}_\lambda$ converges to m for the weak-* topology when $\lambda, \|w\|_2 \to 0$. We recall the result below (see Proposition 4) for the convenience of the reader.

By weak-* convergence of $\tilde{m}_\lambda$ to m for $\lambda \to 0^+$ and $\|w\|_2 \to 0$, $\tilde{m}_\lambda((x - \varepsilon, x + \varepsilon))$ must converge to $m_0((x - \varepsilon, x + \varepsilon))$. By the optimality conditions, we see that $|m_0((x - \varepsilon, x + \varepsilon))| = |m_0(\{x\})|$, so that $m_0(\{x\}) \neq 0$ and $\operatorname{sign} m_0(\{x\}) = \eta_0(x)$, hence the result. $\qquad\square$

In the proof of Lemma 2 we have relied on the following result.

Proposition 4 ([22, Th. 3.5],[4, Prop. 5]). *Let m_0 be an identifiable measure, if $\lambda \to 0$ and $\|w\| \to 0$ with $\frac{\|w\|_2^2}{\lambda} \to 0$, then $\tilde{m}_\lambda$ converges to m_0 with respect to the weak-* topology.*

3.3 Non Degenerate Source Condition

The notion of extended signed support has strong connections with the source condition introduced in [6] to derive convergence rates for the Bregman distance.

Definition 4 (Source Condition). *A measure m_0 satisfies the source condition if there exists $p \in L^2(\mathbb{T})$ such that*

$$\Phi^* p \in \partial \|m_0\|_{\mathrm{TV}}.$$

In a finite-dimensional framework, the source condition is simply equivalent to the optimality of m_0 for $(\mathcal{P}_0(y_0))$ given $y_0 = \Phi m_0$. In the framework of Radon measures, the source condition amounts to assuming that m_0 is a solution of $(\mathcal{P}_0(y_0))$ *and* that there exists a solution to $(\mathcal{D}_0(y_0))$. In fact, the source condition simply means that the conditions of Proposition 3 hold.

If one is interested in m_0 being the *unique* solution of $(\mathcal{P}_0(y_0))$ for $y_0 = \Phi m_0$ (in which case we say that m_0 is *identifiable*), the source condition may be strengthened to give a sufficient condition.

Proposition 5 ([12, Lemma 1.1]). *Let $m_0 = m_{x_0,a_0}$ be a discrete measure. If Φ_{x_0} has full rank, and if*

- *there exists $\eta \in \operatorname{Im}\Phi^*$ such that $\eta \in \partial\|m_0\|_{TV}$,*

- $\forall s \notin \operatorname{Supp}(m_0), \quad |\eta(s)| < 1,$

then m_0 is the unique solution of $(\mathcal{P}_0(y_0))$.

In this paper, in view of Lemma 2, we strengthen a bit more the Source Condition so as to derive a global stability result concerning the support of the solutions of $\mathcal{P}(y_0 + w)$ (see Theorem 2).

Definition 5 (Non Degenerate Source Condition). *Let $m_0 = m_{x_0,a_0}$ be a discrete measure, and $\{x_{0,1}, \ldots x_{0,N}\} = \operatorname{Supp} m_0$. We say that m_0 satisfies the Non Degenerate Source Condition (NDSC) if*

- *there exists $\eta \in \operatorname{Im}\Phi^*$ such that $\eta \in \partial\|m_0\|_{TV}$.*

- *the minimal norm certificate η_0 satisfies*

$$\forall s \in \mathbb{T} \setminus \{x_{0,1}, \ldots x_{0,N}\}, \qquad |\eta_0(s)| < 1,$$
$$\forall i \in \{1, \ldots N\}, \qquad \eta_0''(x_{0,i}) \neq 0.$$

In that case, we say that η_0 is not degenerate.

The first assumption in the above definition is the standard Source Condition. The last two assumptions impose conditions on the extended signed support, namely that $\operatorname{Supp}^{\pm} m_0 = \operatorname{Ext}^{\pm}(m_0)$ and $\eta_0''(t) \neq 0$ for all $t \in \operatorname{Supp} m_0$.

When Φ is an ideal low-pass filter with cutoff frequency f_c, there are numerical evidences that measures having a large enough separation distance (proportional to f_c) satisfy the non degenerate source condition, see Section 4.

3.4 Main Result

The following theorem, which is the main result of this paper, gives a global result on the precise structure of the solution when the signal-to-noise ratio is large enough and λ is small enough.

Theorem 2 (Noise robustness). *Let $m_0 = m_{a_0,x_0} = \sum_{i=1}^N a_{0,i} \delta_{x_{0,i}}$ be a discrete measure. Assume that Γ_{x_0} (defined in (4)) has full rank and that m_0 satisfies the Non Degenerate Source Condition.*

Then there exists $\alpha > 0, \lambda_0 > 0$, such that for $(\lambda, w) \in D_{\alpha,\lambda_0}$, the solution $\tilde{m}_\lambda$ of $\mathcal{P}_\lambda(y + w)$ is unique and is composed of exactly N spikes.

Moreover, up to a permutation of indices, we may write $\tilde{m}_\lambda = \sum_{i=1}^N \tilde{a}_{\lambda,i} \delta_{\tilde{x}_{\lambda,i}}$ with $\tilde{a}_{\lambda,i} \neq 0$ and $\text{sign}(\tilde{a}_{\lambda,i}) = \text{sign}(a_{0,i})$ (for $1 \leqslant i \leqslant N$), and writing $(\tilde{a}_0, \tilde{x}_0) = (a_0, x_0)$, the mapping

$$(\lambda, w) \in D_{\alpha,\lambda_0} \mapsto (\tilde{a}_\lambda, \tilde{x}_\lambda) \in \mathbb{R}^N \times \mathbb{T}^N,$$

is C^{k-1} whenever $\varphi \in C^k(\mathbb{T})$ ($k \geqslant 2$).

In particular, for $\lambda = \frac{1}{\alpha}\|w\|_2$, we have

$$\forall i \in \{1, \ldots N\}, \quad |\tilde{x}_{\lambda,i} - x_{0,i}| = O(\|w\|_2) \quad and \quad |\tilde{a}_{\lambda,i} - a_{0,i}| = O(\|w\|_2). \quad (26)$$

Proof. Applying Lemma 2 at each point $x_{0,i}$ for $1 \leqslant i \leqslant N$ and Lemma 1, we see that for $\varepsilon > 0$ small enough, there exists $\alpha > 0, \lambda_0 > 0$ such that $\tilde{m}_\lambda$ has at most one spike in each interval $(x_{i,0} - \varepsilon, x_{i,0} + \varepsilon)$, and

$$|\tilde{m}_\lambda| \left(\mathbb{T} \setminus \bigcup_{i=1}^N (x_{i,0} - \varepsilon, x_{i,0} + \varepsilon) \right) = 0.$$

In fact, since Γ_{x_0} has full rank, $\Phi_{\text{Ext }m_0}$ has full rank as well and m_0 is identifiable (by Proposition 3). Therefore, Lemma 2 ensures that there is indeed one spike in each interval, with sign equal to $\eta_0(x_{0,i})$.

It remains to prove the uniqueness of the amplitudes and locations $(\tilde{a}_\lambda, \tilde{x}_\lambda)$ and their smoothness as function of (λ, w). To this end, we observe that they satisfy the following implicit equation

$$E_{s_0}(\tilde{a}_\lambda, \tilde{x}_\lambda, \lambda, w) = 0$$

where $s_0 = \text{sign}(a_0) = (\eta_0(x_{i,0}))_{1 \leqslant i \leqslant N}$, and

$$E_{s_0}(a, x, \lambda, w) = \begin{pmatrix} \Phi_x^*(\Phi_x a - y_0 - w) + \lambda s_0 \\ \Phi_x'^*(\Phi_x a - y_0 - w) \end{pmatrix} = \Gamma_x^*(\Phi_x a - y_0 - w) + \lambda \begin{pmatrix} s_0 \\ 0 \end{pmatrix}.$$

Indeed, this implicit equation simply states that $\tilde{\eta}_\lambda(\tilde{x}_{\lambda,i}) = \text{sign}(a_{0,i}) = \text{sign}(\tilde{a}_{\lambda,i})$, and that $\tilde{\eta}_\lambda'(\tilde{x}_{\lambda,i}) = 0$.

Since $((a, x), (\lambda, w)) \mapsto E_{s_0}(a, x, \lambda, w)$ is a C^1 function defined on $(\mathbb{R}^N \times \mathbb{T}^N) \times (\mathbb{R} \times L^2(\mathbb{T}^N))$, we may apply the implicit functions theorem.

The derivative of E_{s_0} with respect to x and a reads

$$\frac{\partial E}{\partial a}(a, x, \lambda, w) = \Gamma_x^* \Phi_x$$

$$\frac{\partial E}{\partial x}(a, x, \lambda, w) = \begin{pmatrix} \text{diag}(\Phi_x^{*\prime}(\Phi_x a - y_0 - w)) \\ \text{diag}(\Phi_x^{*\prime\prime}(\Phi_x a - y_0 - w)) \end{pmatrix} + \Gamma_x^* \Phi_x' \, \text{diag}(a).$$

so that for $\lambda = 0$, $w = 0$ and using $y_0 = \Phi_{x_0} a_0$, one obtains

$$\frac{\partial E_s}{\partial(a,x)}(a_0, x_0, 0, 0) = \Gamma^*_{x_0}\left(\Phi_{x_0},\ \Phi'_{x_0}\operatorname{diag}(a_0)\right)$$
$$= (\Gamma^*_{x_0}\Gamma_{x_0})\begin{pmatrix}\operatorname{Id} & 0 \\ 0 & \operatorname{diag}(a_0)\end{pmatrix}.$$

Since we assume Γ_{x_0} has full rank, then $\frac{\partial E_{s_0}}{\partial(a,x)}(a_0, x_0, 0, 0)$ is invertible and the implicit functions theorem applies: there is a neighborhood $V \times W$ of $(a_0, x_0) \times \{(0,0)\}$ in $(\mathbb{R}^N \times \mathbb{T}^N) \times (\mathbb{R} \times L^2(\mathbb{T}))$ and a function $f : W \to V$ such that

$$((a,x), \lambda, w) \in V \times W \quad \text{and} \quad E_{s_0}(a, x, \lambda, w) = 0$$
$$\iff \quad (\lambda, w) \in W \quad \text{and} \quad (a,x) = f(\lambda, w).$$

Moreover, writing $(\hat{a}_{\lambda,w}, \hat{x}_{\lambda,w}) = f(\lambda, w) \in \mathbb{R}^N \times \mathbb{T}^N$, we have

- $(\hat{a}_{0,0}, \hat{x}_{0,0}) = (a_0, x_0)$,

- for any $(\lambda, w) \in W$, $\operatorname{sign}(\hat{a}_{\lambda,w}) = s_0$,

- if $\varphi \in C^k(\mathbb{T})$ (for $k \geqslant 2$), then $f \in C^{k-1}(W)$.

The constructed amplitudes and locations $(\hat{a}_{\lambda,w}, \hat{x}_{\lambda,w})$ coincide with those of the solutions of $\mathcal{P}_\lambda(y_0 + w)$ for all $(\lambda, w) \in W$ such that $\|w\|_2 \leqslant \alpha\lambda$. Possibly changing the value of λ_0 so that $D_{\alpha,\lambda_0} \subset W$, we obtain the desired result. $\quad \square$

Remark 1. Although this paper focuses on identifiable measures, Theorem 2 describes the evolution of the solutions of $\mathcal{P}_\lambda(y_0 + w)$ for any input measure m_1 such that there exists m_0 which satisfies the non degenerate source condition and $y_0 = \Phi m_1 = \Phi m_0$. Instead of converging towards m_1, the solutions will converge towards m_0.

3.5   Extensions

Theorem 2 extends in a straightforward manner to higher dimensions, i.e. when replacing $\mathbb{T}$ by $\mathbb{T}^d$ for $d \geqslant 1$. In the NDSC introduced in Definition 5, one should replace, for $i = 1, \ldots, N$, the constraint $\eta''_0(x_{0,i}) \neq 0$ by the constraint that the Hessian $D^2\eta_0(x_{0,i}) \in \mathbb{R}^{d \times d}$ is invertible.

The proof also extends to non-stationary filtering operators, i.e. which can be written as

$$\forall t \in \mathbb{T}^d, \quad \Phi m(t) = \int_{\mathbb{T}^d} \varphi(x, t)\mathrm{d}m(x)$$

where $\varphi \in C^2(\mathbb{T}^d \times \mathbb{T}^d)$.

3.6 Application to the ideal Low-pass filter

We first observe that the injectivity condition on Γ_x assumed in Theorem 2 always holds.

Proposition 6 (Injectivity of Γ_x). *Let $x = (x_1, \ldots x_N) \in \mathbb{T}^N$ with $x_i \neq x_j$ for $i \neq j$ and $N \leqslant f_c$. Then $\Gamma_x = (\Phi_x, \Phi'_x)$ has full rank.*

The proof is given in Appendix B.

(a) m_0 and η_0 (b) $w = 0$

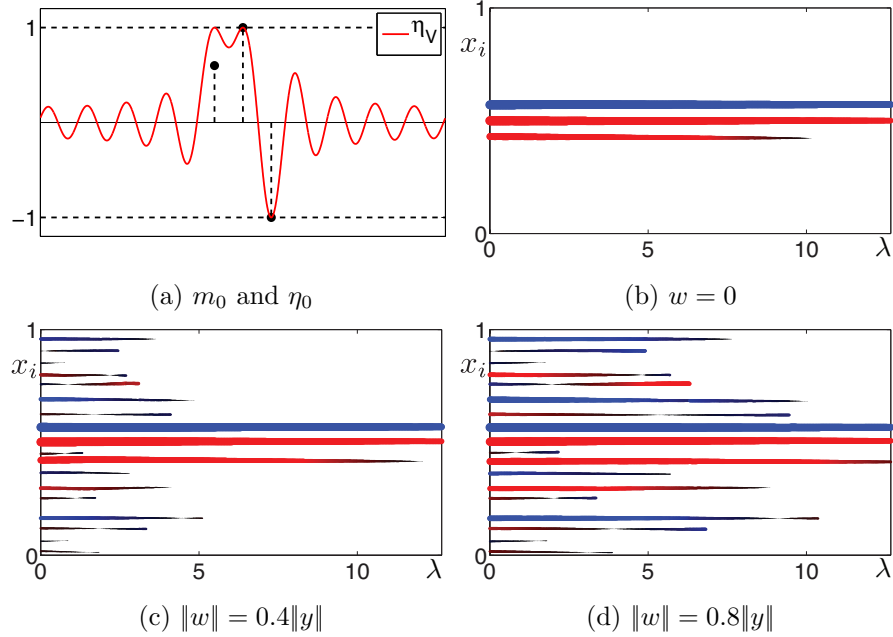(c) $\|w\| = 0.4\|y\|$ (d) $\|w\| = 0.8\|y\|$

Figure 1: (a) Input measure m_0, and corresponding minimal norm certificate. (b,c,d) Regularization paths $\lambda \mapsto \tilde{m}_\lambda$ that are solutions of $\mathcal{P}_\lambda(\Phi m_0 + w)$ for three different noise levels $\|w\|$. Each "strip" represents the evolution of a spike as λ varies. The color refers to the sign of the spike (blue for negative and red for positive) and the (vertical) width is proportional to its amplitude. The exact location is given by the middle of each band.

As to whether or not the Non Degenerate Source Condition holds for discrete measures, we will discuss this matter in Section 4 more in depth. For now, let us mention that we have observed empirically that this condition holds under the hypotheses of Theorem 1.2 in [8], namely that $\Delta(m) > \frac{1.87}{f_c}$, but also with measures with far smaller values of $\Delta(m)$.

Figure 1 shows the whole solution path $\lambda \mapsto \tilde{m}_\lambda$ of the solutions of $\mathcal{P}_\lambda(\Phi m_0 + w)$ when $f_c = 10$ and the input measure is identifiable and has three spikes separated by $\Delta(m_0) = 0.7/f_c$. Such a measure satisfies the Non-degenerate Source Condition as shown in plot (a). The plots (b,c,d) illustrate the conclusion

(a) m_0 and η_0

(b) $\|w_0\| = 0.07\|y\|$

(c) $\|w_0\| = 0.1\|y\|$
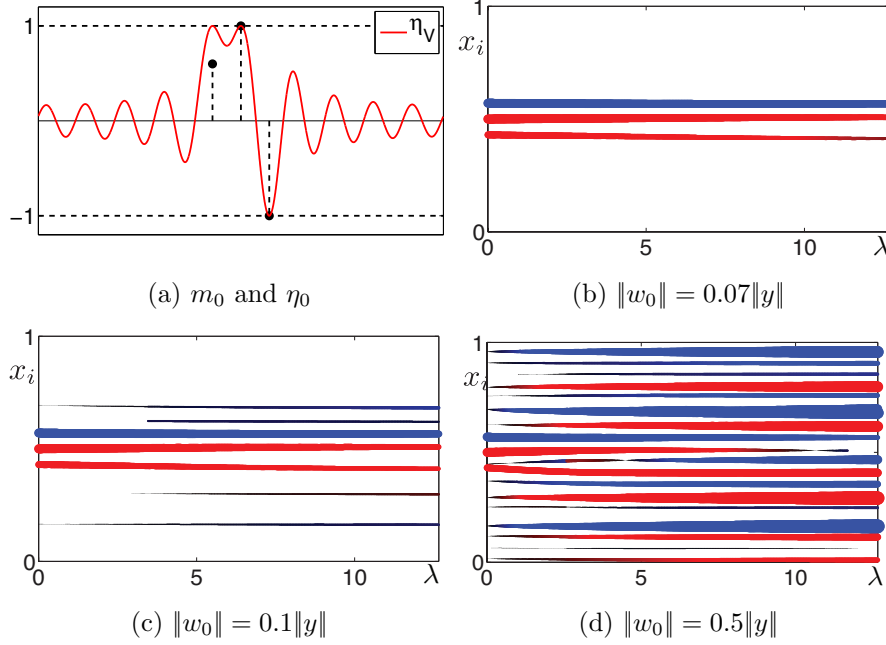
(d) $\|w_0\| = 0.5\|y\|$

Figure 2: Same plots as Figure 1 except that the solutions of $\mathcal{P}(\Phi m_0 + \lambda w_0)$ are displayed instead of those of $\mathcal{P}_\lambda(\Phi m_0 + w)$.

of Theorem 2. For values of λ which are too small with respect to $\|w\|$, the solution $\tilde{m}_\lambda$ is perturbed with spurious spikes, but as soon as λ is large enough, $\tilde{m}_\lambda$ has a support that closely (but not exactly) matches the one of m_0. For large value of λ, spikes starts disappearing, and the support is not correctly estimated. Figure 2 shows the solutions of $\mathcal{P}_\lambda(\Phi m_0 + \lambda w_0)$, i.e. the noise $w = \lambda w_0$ is scaled by the regularization parameter λ. In accordance with Theorem 2, this shows that for $\|w\|_2/\lambda = \|w_0\|_2 \leqslant 0.07$, the support of the spikes is precisely estimated.

4   Vanishing Derivatives Pre-certificate

We show in this section that, if the Non Degenerate Source Condition holds, the minimal norm certificate η_0 is characterized by its values on the support of m_0 and the fact that its derivative must vanish on the support of m_0. Thus, one may compute the minimal norm certificate simply by solving a linear system, without handling the cumbersome constraint $\|\eta_0\|_\infty \leqslant 1$.

4.1   Dual Pre-certificates

Loosely speaking, we call *pre-certificate* any "good candidate" for a solution of (12). Typically, a pre-certificate is built by solving a linear system (with possibly a condition on its norm). The following pre-certificate appears naturally in our analysis.

Definition 6 (Vanishing derivative pre-certificate). *The vanishing derivative pre-certificate associated with a measure $m_0 = m_{a_0,x_0}$ is $\eta_V = \Phi^* p_V$ where*

$$p_V = \underset{p \in L^2(\mathbb{T})}{\operatorname{argmin}} \ \|p\|_2 \quad subj. \ to \quad \forall 1 \leqslant i \leqslant N, \quad \left\{ \begin{array}{l} (\Phi^* p)(x_{0,i}) = \operatorname{sign}(a_{0,i}), \\ (\Phi^* p)'(x_{0,i}) = 0. \end{array} \right. \tag{27}$$

It is clear that if the Source Condition (see Definition 4) holds, then p_V exists (since Problem (27) is feasible). Observe that, in general, η_V is not a certificate for m_0 since it does not satisfy the constraint $\|\eta_V\|_\infty \leqslant 1$. The following proposition gathers several facts about the vanishing derivative pre-certificate which show that it is indeed a good candidate for the minimal norm certificate.

Proposition 7. *Let $m_0 = m_{a_0,x_0} = \sum_{i=1}^N a_{0,i} \delta_{x_{0,i}}$ be a discrete measure. The following assertions hold.*

- *Problem (27) is feasible and $\|\eta_V\|_\infty \leqslant 1$ if and only if the Source Condition holds and $\eta_V = \eta_0$.*

- *If Problem (27) is feasible and Γ_{x_0} has full rank, i.e. $\Gamma_{x_0}^* \Gamma_{x_0} \in \mathbb{R}^{2N \times 2N}$ is invertible, then*

$$\eta_V = \Phi^* \Gamma_{x_0}^{+,*} \begin{pmatrix} \operatorname{sign}(a_0) \\ 0 \end{pmatrix} \quad where \quad \Gamma_{x_0}^{+,*} = \Gamma_{x_0}(\Gamma_{x_0}^* \Gamma_{x_0})^{-1}.$$

- *If Γ_{x_0} has full rank, then m_0 satisfies the Non Degenerate Source Condition if and only if Problem (27) is feasible and*

$$\begin{array}{ll} \forall s \in \mathbb{T} \setminus \{x_{0,1}, \dots x_{0,N}\}, & |\eta_V(s)| < 1, \\ \forall i \in \{1, \dots N\}, & \eta_V''(x_{0,i}) \neq 0. \end{array}$$

The third assertion of Proposition 7 states that it is equivalent to check the Non Degenerate Source Condition on η_0 (Definition 5) or to check the same conditions on η_V. In case those conditions hold, one even has $\eta_V = \eta_0$ (first assertion). The main point of this equivalence is that the second assertion yields a practical expression to compute η_V which may be used in numerical experiments (see Section 4.3).

Proof. For the first assertion, we observe that if Problem (27) is feasible (and thus p_V exists) and $\|\eta_V\|_\infty \leqslant 1$, then $\eta_V \in \partial \|m_0\|_{\mathrm{TV}}$ and the Source Condition holds. Hence, $\|p_V\|_2 \geqslant \|p_0\|_2$. On the other hand the minimal norm certificate η_0 must satisfy all the constraints of (27), thus the minimality of the norms of both η_V and η_0 implies that $\eta_V = \eta_0$. The converse implication is obvious.

For the second assertion, Problem (27) can be written as

$$\eta_V = \underset{\eta = \Phi^* p}{\operatorname{argmin}} \ \|p\|_2. \quad subj. \ to \quad \left\{ \begin{array}{l} \Phi_{x_0}^* p = \operatorname{sign}(a_0), \\ \Phi_{x_0}'^* p = 0, \end{array} \right.$$

which is a quadratic optimization problem in a Hilbert space with a finite number of affine equality constraints. Moreover, the assumption that Γ_{x_0} has full rank implies that the constraints are qualified. Hence it can be solved by introducing Lagrange multipliers u and v for the constraints. One should therefore solve the following linear system to obtain the value of $p = p_{\mathrm{v}}$

$$
\begin{pmatrix} \mathrm{Id} & \Phi_{x_0} & \Phi'_{x_0} \\ \Phi^*_{x_{0*}} & 0 & 0 \\ \Phi'^*_{x_0} & 0 & 0 \end{pmatrix} \begin{pmatrix} p \\ u \\ v \end{pmatrix} = \begin{pmatrix} 0 \\ s \\ 0 \end{pmatrix}.
$$

Solving for (u, v) in these equations gives the result.

For the third assertion, if the Non Degenerate Source condition holds, we apply Theorem 2 which yields a C^1 path $\lambda \mapsto (\tilde{a}_\lambda, \tilde{x}_\lambda)$ of solutions of $\mathcal{P}_\lambda(y_0)$ (we consider the case $w = 0$). Then from Proposition 8 below, we obtain that η_{v} is a valid certificate and $\eta_{\mathrm{v}} = \eta_0$, hence η_{v} is non-degenerate. The converse implication is a straightforward consequence of the first assertion. $\qquad\square$

4.2 Necessary condition for support recovery

There is a priori no reason for the vanishing derivative pre-certificate η_{v} to satisfy $\|\eta_{\mathrm{v}}\|_\infty \leqslant 1$. Here, we prove that that is in fact a necessary condition for (noiseless) exact support recovery to hold on some interval $[0, \lambda_0)$ with $\lambda_0 > 0$, *i.e.* the solutions of $\mathcal{P}_\lambda(y_0)$ having exactly N spikes which converge smoothly towards those of the original measure.

Proposition 8. *Let $m_0 = m_{a_0,x_0} = \sum_{i=1}^N a_{0,i} \delta_{x_{0,i}}$ be a discrete measure such that Γ_{x_0} has full rank. Assume that there exists $\lambda_0 > 0$ and a C^1 path $[0, \lambda_0) \rightarrow \mathbb{R}^N \times \mathbb{T}^N$, $\lambda \mapsto (a_\lambda, x_\lambda)$ such that for all $\lambda \in [0, \lambda_0)$ the measure $m_\lambda = m_{a_\lambda,x_\lambda}$ is a solution to $\mathcal{P}_\lambda(y_0)$ (the noiseless problem).*

Then η_{v} exists, $\|\eta_{\mathrm{v}}\|_\infty \leqslant 1$ and $\eta_{\mathrm{v}} = \eta_0$.

Proof. Let $p_\lambda = \frac{1}{\lambda}(y_0 - \Phi m_\lambda) = \frac{1}{\lambda}(\Phi_{x_0} a_0 - \Phi_{x_\lambda} a_\lambda)$ be the certificate defined by the optimality conditions (11). We show that $\Phi^* p_\lambda$ converges towards $\Phi^* \Gamma_{x_0}^{+,*} \begin{pmatrix} \mathrm{sign}(a_0) \\ 0 \end{pmatrix} = \eta_{\mathrm{v}}$ (and that the latter exists).

Writing

$$
a'_\lambda = \frac{\mathrm{d}a_\lambda}{\mathrm{d}\lambda} \in \mathbb{R}^N \quad \text{and} \quad x'_\lambda = \frac{\mathrm{d}x_\lambda}{\mathrm{d}\lambda} \in \mathbb{R}^N,
$$

we observe that for any $i \in \{1, \dots N\}$ and any $x \in \mathbb{T}$,

$$
\frac{a_{\lambda,i}\varphi(x_{\lambda,i} - x) - a_{0,i}\varphi(x_{0,i} - x)}{\lambda} - \left[a_{0,i}\varphi'(x_{0,i} - x)x'_{0,i} + a'_{0,i}\varphi(x_{0,i} - x) \right]
$$

$$
= \int_0^1 \left[a_{\lambda t,i}\varphi'(x_{\lambda t,i} - x)x'_{\lambda t,i} + a'_{\lambda t,i}\varphi(x_{\lambda t,i} - x) \right]
$$

$$
- \left[a_{0,i}\varphi'(x_{0,i} - x)x'_{0,i} + a'_{0,i}\varphi(x_{0,i} - x) \right] \mathrm{d}t,
$$

23

and the latter integral converges (uniformly in x) to zero when $\lambda \to 0^+$ by uniform continuity of its integrand (since a, x and φ are C^1). As a consequence, we obtain that $\frac{y_0 - \Phi_{x_\lambda} a_\lambda}{\lambda}$ converges uniformly to $-\Gamma_{x_0} \begin{pmatrix} \mathrm{Id} & 0 \\ 0 & \mathrm{diag}(a_0) \end{pmatrix} \begin{pmatrix} a_0' \\ x_0' \end{pmatrix}$.

On the other hand, we observe that for λ small enough, $\mathrm{sign}(a_\lambda) = \mathrm{sign}(a_0)$, and using the notations of the proof of Theorem 2, the implicit equation $E_{s_0}(a_\lambda, x_\lambda, \lambda, 0) = 0$ holds. Differentiating that equation at $\lambda = 0$ we obtain:

$$\left(\frac{\partial E_{s_0}}{\partial(a,x)}(a_0, x_0, 0, 0) \right) \begin{pmatrix} a_0' \\ x_0' \end{pmatrix} + \frac{\partial E_{s_0}}{\partial \lambda}(a_0, x_0, 0, 0) = 0,$$

or equivalently

$$(\Gamma_{x_0}^* \Gamma_{x_0}) \begin{pmatrix} \mathrm{Id} & 0 \\ 0 & \mathrm{diag}(a_0) \end{pmatrix} \begin{pmatrix} a_0' \\ x_0' \end{pmatrix} = - \begin{pmatrix} s_0 \\ 0 \end{pmatrix}.$$

As a consequence, Problem (27) is feasible and we see that $\frac{y_0 - \Phi_{x_\lambda} a_\lambda}{\lambda}$ converges uniformly (and thus in the L^2 strong topology) to $\Gamma_{x_0} (\Gamma_{x_0}^* \Gamma_{x_0})^{-1} \begin{pmatrix} \mathrm{sign}(a_0) \\ 0 \end{pmatrix}$ and $\Phi^* \left(\frac{y_0 - \Phi_{x_\lambda}}{\lambda} \right)$ converges uniformly to $\Phi^* \Gamma_{x_0}^{+,*} \begin{pmatrix} \mathrm{sign}(a_0) \\ 0 \end{pmatrix}$ (which is precisely η_{V} from the second assertion of Proposition 7).

Since $\| \Phi^* \left(\frac{y_0 - \Phi_{x_\lambda} a_\lambda}{\lambda} \right) \|_\infty = \| \Phi^* p_\lambda \|_\infty \leqslant 1$ for all $\lambda > 0$, we obtain that $\| \eta_{\mathrm{V}} \|_\infty \leqslant 1$, hence the claimed result. $\qquad \square$

4.3 Application to the Ideal Low-pass Filter

In order to prove their identifiability result for measures, the authors of [8] also introduce a "good candidate" for a dual certificate associated with $m = m_{a,x}$ for $a \in \mathbb{C}^N$ and $x \in \mathbb{R}^N$. For K being the square of the Fejer kernel, they build a trigonometric polynomial

$$\eta_{\mathrm{CF}}(t) = \sum_{i=1}^N (\alpha_i K(t - x_i) + \beta_i K'(t - x_i)) \ \text{ with } K(t) = \left(\frac{\sin \left(\left(\frac{f_c}{2} + 1 \right) \pi t \right)}{\left(\frac{f_c}{2} + 1 \right) \sin \pi t} \right)^4$$

and compute $(\alpha_i, \beta_i)_{i=1}^N$ by imposing that $\eta_{\mathrm{CF}}(x_i) = \mathrm{sign}(a_i)$ and $(\eta_{\mathrm{CF}})'(x_i) = 0$.

They show that the constructed pre-certificate is indeed a certificate, i.e. that $\| \eta_{\mathrm{CF}} \|_\infty \leqslant 1$, provided that the support is separated enough (i.e. when $\Delta(m) \geqslant C/f_c$). This result is important since it proves that measures that have sufficiently separated spikes are identifiable. Furthermore, using the fact that η_{CF} is not degenerate (i.e. $(\eta_{\mathrm{CF}})''(x_i) \neq 0$ for all $i = 1, \ldots, N$), the same authors derive an L^2 robustness to noise result in [7], and Fernandez-Granda and Azais et al. use the constructed certificate to analyze finely the local averages of the spikes in [19, 1].

From a numerical perspective, we have investigated how this pre-certificate compares with the vanishing derivative pre-certificate that appears naturally

in our analysis, by generating real-valued measures for different separation distances and observing when each pre-certificate η satisfies $\|\eta\|_\infty \leqslant 1$.
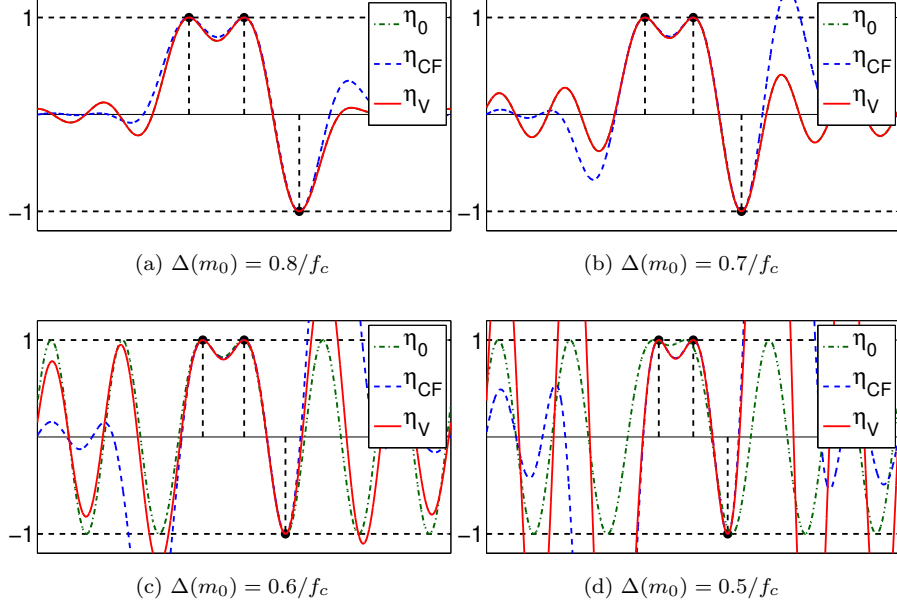

(a) $\Delta(m_0) = 0.8/f_c$ (b) $\Delta(m_0) = 0.7/f_c$

(c) $\Delta(m_0) = 0.6/f_c$ (d) $\Delta(m_0) = 0.5/f_c$

Figure 3: Pre-certificates for three equally spaced masses. The blue curves with dots is the Fejer pre-certificate η_{CF}, while red continuous line is the vanishing derivative η_{V}. The black dashed line is the minimal norm precertificate η_0.

As predicted by the result of [8], we observe numerically that the pre-certificate η_{CF} is a certificate (i.e. $\|\eta_{\mathrm{CF}}\|_\infty \leqslant 1$) for any measure with $\Delta(m_0) \geqslant 1.87/f_c$. We also observe that this continues to hold up to $\Delta(m_0) \geqslant 1/f_c$. Yet, below $1/f_c$, it may happen that some measures are still identifiable (as asserted using the vanishing derivative pre-certificate η_{V}) but η_{CF} stops being a certificate, i.e. $\|\eta_{\mathrm{CF}}\|_\infty > 1$. A typical example is shown in Figure 3, where, for $f_c = 6$ we have used three equally spaced masses as an input, their separation distance being $\Delta(m_0) \in \{\frac{0.8}{f_c}, \frac{0.7}{f_c}, \frac{0.6}{f_c}, \frac{0.5}{f_c}\}$. Here, we have computed an approximation of the minimal norm certificate η_0 by solving $(\mathcal{D}_\lambda(y_0))$ with very small λ.

For $\Delta(m_0) = \frac{0.8}{f_c}$, both η_{V} and η_{CF} are certificates, so that the vanishing derivatives pre-certificate η_{V} is equal to the minimal norm certificate η_0. For $\Delta(m_0) = \frac{0.7}{f_c}$, η_{CF} violates the constraint $\|\eta_{\mathrm{CF}}\|_\infty \leqslant 1$ but the vanishing derivative pre-certificates is still a certificate (even showing that the measure is identifiable). For $\Delta(m_0) = \frac{0.6}{f_c}$ and $\frac{0.5}{f_c}$, neither η_{V} nor η_{CF} satisfy the constraint, hence $\eta_{\mathrm{V}} \neq \eta_0$. Yet, η_0 ensures that m_0 is a solution to $(\mathcal{P}_0(y_0))$.

From the experiments we have carried out, we have observed that the vanishing derivative pre-certificate η_{V} behaves in general at least as well as the square
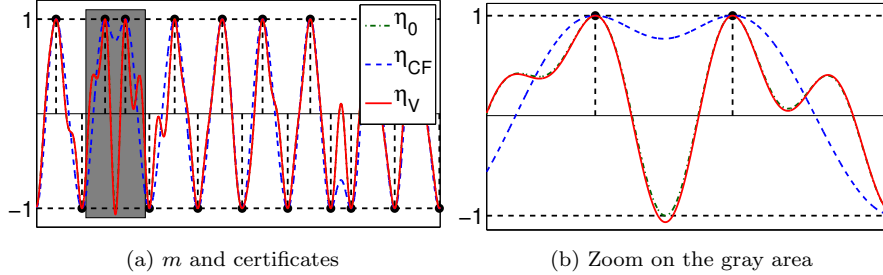
(a) m and certificates   (b) Zoom on the gray area

Figure 4: Example of measure for which $\eta_{\mathrm{V}} \neq \eta_0$.

Fejer η_{CF}. The only exceptions we have noticed is for a large number of peaks (when N is close to f_c), with $\Delta(m_0) \leqslant \frac{1.5}{f_c}$. This is illustrated in Figure 4 which shows a measure m_0 for which η_{CF} is a non-degenerate certificate (which shows that it is identifiable), but for which $\eta_0 \neq \eta_{\mathrm{V}}$ since $\|\eta_{\mathrm{V}}\|_\infty > 1$ (thus η_{V} is not a certificate). Typically, we have in this case $\mathrm{Supp}^\pm m_0 \subsetneq \mathrm{Ext}^\pm(m_0)$. Such a measure is identifiable but there is no support recovery for $\lambda > 0$ (in the sense of Proposition 8), hence its support is not stable.

Such pathological cases are relatively rare. An intuitive explanation for this is the fact that having $\eta_0(x) = \pm 1$ for $x \in \mathbb{T} \setminus \mathrm{Supp}(m_0)$ or $\eta_0''(x) = 0$ for some $x \in \mathrm{Supp}(m_0)$ tend to impose a large L^2 norm, thus contradicting the minimality of p_0 (recall that when φ is an ideal low pass filter $\|\eta\|_2 = \|p\|_2$).

5 Discrete Sparse Spikes Deconvolution

5.1 Finite Dimensional ℓ^1 Regularization

A popular way to compute approximate solutions to $(\mathcal{P}_\lambda(y_0))$ with fast algorithms is to solve this problem on a finite discrete grid $\mathcal{G} \subset \mathbb{T}$. Denoting by P the cardinal of the grid $\mathcal{G}$, and by $g \in \mathbb{T}^P$ the finite sequence of elements of $\mathcal{G}$, the idea is to solve $\mathcal{P}_\lambda(y_0)$ (or $\mathcal{P}_0(y_0)$) with the additional constraint that $m = \sum_{i=1}^P a_i \delta_{g_i}$ for some $a \in \mathbb{R}^P$.

This is nothing but the so-called basis pursuit denoising problem [9], also known as the Lasso [31] in statistics. Indeed, defining the linear operator Ψ through

$$\Psi a = \Phi m = \sum_{i=1}^P (\Phi \delta_{g_i}) a_i,$$

the problem amounts to:

$$\min_{a \in \mathbb{R}^P} \frac{1}{2}\|y_0 - \Psi a\|^2 + \lambda\|a\|_1 \quad \text{where} \quad \|a\|_1 = \sum_{i=1}^P |a_i|, \qquad (\tilde{\mathcal{P}}_\lambda^{\mathcal{G}}(y_0))$$

where $\Psi : \mathbb{R}^P \to L^2(\mathbb{T})$ is a linear operator ($L^2(\mathbb{T})$ may as well be replaced with $\mathbb{R}^Q$ or any Hilbert space), and a_i denotes the mass at each point i of the grid.

In the noiseless case, the exact reconstruction problem reads:

$$\min_{\Psi a = y_0} \|a\|_1. \qquad (\tilde{\mathcal{P}}_0^{\mathcal{G}}(y_0))$$

The aim of the present section is to study the asymptotic of Problems $(\tilde{\mathcal{P}}_\lambda^{\mathcal{G}}(y_0))$ and $(\tilde{\mathcal{P}}_0^{\mathcal{G}}(y_0))$ as the stepsize of the grid $\mathcal{G}$ vanishes. To this end, we keep the framework of measures and we reformulate the constraint that $\mathrm{Supp}(m) \subset \mathcal{G}$, i.e. that m can be written as $m = m_{a,x}$, where $x = (x_1, \ldots, x_N) \in \mathcal{G}^N$. Recall that the notation $m_{a,x}$ hints that $a_i \neq 0$ for all i and that the x_i's are all distinct, so that in general $N \leqslant P$. We thus adopt the following penalization term

$$\|m\|_{\mathrm{TV},\mathcal{G}} = \sup \left\{ \int \psi \mathrm{d}m \; ; \; \psi \in C(\mathbb{T}), \forall t \in \mathcal{G} \; |\psi(t)| \leqslant 1 \right\}, \qquad (28)$$

so that $\|m\|_{\mathrm{TV},\mathcal{G}} = +\infty$ when $\mathrm{Supp}(m) \not\subset \mathcal{G}$, and $\sum_{i=1}^N |a_i|$ otherwise.

Problems $(\tilde{\mathcal{P}}_\lambda^{\mathcal{G}}(y_0))$ and $(\tilde{\mathcal{P}}_0^{\mathcal{G}}(y_0))$ are then respectively equivalent to:

$$\min_{m \in \mathcal{M}(\mathbb{T})} \frac{1}{2}\|\Phi(m) - y_0\|^2 + \lambda\|m\|_{\mathrm{TV},\mathcal{G}} \qquad (\mathcal{P}_\lambda^{\mathcal{G}}(y_0))$$

and

$$\min_{\Phi m = y_0} \|m\|_{\mathrm{TV},\mathcal{G}}, \qquad (\mathcal{P}_0^{\mathcal{G}}(y_0))$$

Let us stress the fact that the results of Sections 5.2 and 5.3 hold for *any finite dimensional matrix* $\Psi \in \mathbb{R}^{P \times Q}$ or linear operator $\Psi : \mathbb{R}^P \to L^2(\mathbb{T})$: the columns of Ψ need not be the samples of a convolution operator.

5.2 Certificates over a Discrete Grid

As in Section 2, we may compute the subdifferential of the ℓ^1 norm. For $m = m_{a,x} = \sum_{i=1}^N a_i \delta_{x_i}$ with support in $\mathcal{G}$:

$$\partial\|m\|_{\mathrm{TV},\mathcal{G}} = \{\eta \in C(\mathbb{T}) \; ; \; \|\eta\|_{\infty,\mathcal{G}} \leqslant 1, \forall i = 1, \ldots, N, \; \eta(x_i) = \mathrm{sign}(a_i)\}. \qquad (29)$$

where

$$\|\eta\|_{\infty,\mathcal{G}} = \max\left\{|\eta(t)| \; ; \; t \in \mathcal{G}\right\}.$$

We also introduce the corresponding dual problems:

$$\min_{\|\Phi^* p\|_{\infty,\mathcal{G}} \leqslant 1} \left\|\frac{y_0}{\lambda} - p\right\|_2^2, \qquad (\mathcal{D}_\lambda^{\mathcal{G}}(y_0))$$

$$\sup_{\|\Phi^* p\|_{\infty,\mathcal{G}} \leqslant 1} \langle y_0, \, p\rangle. \qquad (\mathcal{D}_0^{\mathcal{G}}(y_0))$$

Remark 2. Let us denote by G the image by Φ of all measures with support in $\mathcal{G}$. It may happen (for instance if the grid is too rough) that $y_0 \notin G$, in which case $(\mathcal{P}_0^{\mathcal{G}}(y_0))$ is not feasible and $(\mathcal{D}_0^{\mathcal{G}}(y_0))$ has infinite value. But $(\mathcal{P}_\lambda^{\mathcal{G}}(y_0))$ is

then equivalent to $\mathcal{P}_\lambda^{\mathcal{G}}(y_{0,G})$ where $y_0 = y_{0,G} + y_{0,G^\perp}$ is an orthogonal decomposition. Problem $(\mathcal{P}_\lambda^{\mathcal{G}}(y_0))$ is thus an approximation of $\mathcal{P}_0^{\mathcal{G}}(y_{0,G})$, and the relevant dual problems are $\mathcal{D}_\lambda^{\mathcal{G}}(y_{0,G})$ and $\mathcal{D}_0^{\mathcal{G}}(y_{0,G})$. For the sake of simplicity, we shall assume from now on that $y_0 \in G$, but the reader may keep in mind that this hypothesis can be withdrawn by replacing y with $y_{0,G}$.

In view of Remark 2, we observe that problems $(\mathcal{D}_\lambda^{\mathcal{G}}(y_0))$ and $(\mathcal{D}_0^{\mathcal{G}}(y_0))$ are in fact finite dimensional. Indeed, their constraints being invariant by addition of elements of $G^\perp$, we may consider their quotient with the space $G^\perp$. Therefore the condition $p \in L^2(\mathbb{T})$ may be reduced to $p \in G$ where G is a finite dimensional space.

As a consequence, a solution to $(\mathcal{D}_0^{\mathcal{G}}(y_0))$ always exists, so that we may define the *discrete minimal norm certificate*:

$$\eta_0^{\mathcal{G}} = \Phi^* p_0^{\mathcal{G}}, \quad \text{where} \quad p_0^{\mathcal{G}} = \underset{p}{\operatorname{argmin}} \ \left\{ \|p\|_2 \ ; \ p \text{ is a solution of } (\mathcal{D}_0^{\mathcal{G}}(y_0)) \right\}. \quad (30)$$

The solutions of $(\mathcal{P}_\lambda^{\mathcal{G}}(y_0))$ and $(\mathcal{D}_\lambda^{\mathcal{G}}(y_0))$ (resp. $(\mathcal{P}_0^{\mathcal{G}}(y_0))$ and $(\mathcal{D}_0^{\mathcal{G}}(y_0))$) are related by the extremality conditions (11) (resp. (12)) where the total variation is replaced with its discrete counterpart $\|\cdot\|_{\mathrm{TV},\mathcal{G}}$.

5.3   Noise Robustness

As in the continuous case (cf. Section 3), the support of the solutions of $\mathcal{P}_\lambda^{\mathcal{G}}(y_0 + w)$ for $\lambda \to 0^+$ and $\|w\|_2 = O(\lambda)$ is governed by the minimal norm certificate. We introduce here the discrete counterpart of the extended support of a measure.

Definition 7 (Extended support). *Let $m_0 \in \mathcal{M}(\mathbb{T})$ such that $y_0 = \Phi(m_0) \in G$, and let $\eta_0^{\mathcal{G}}$ be the discrete minimal norm certificate defined in (30). The extended support of m_0 relatively to $\mathcal{G}$ is defined as*

$$\operatorname{Ext}_{\mathcal{G}}(m_0) = \left\{ t \in \mathcal{G} \ ; \ |\eta_0^{\mathcal{G}}(t)| = 1 \right\}, \quad (31)$$

and the extended signed support relatively to $\mathcal{G}$ as

$$\operatorname{Ext}_{\mathcal{G}}^\pm(m_0) = \left\{ (t,v) \in \mathcal{G} \times \{-1,+1\} \ ; \ \eta_0^{\mathcal{G}}(t) = v \right\}. \quad (32)$$

It is important to notice that the assumption $y_0 \in G$ does not mean that the support of m_0 is included in $\mathcal{G}$, but that there exists a measure with support included in $\mathcal{G}$ which produces the same observation y_0. Therefore the support of m_0 and its extended support may even be disjoint.

As in the continuous case, notice that m_0 is a solution of $(\mathcal{P}_0^{\mathcal{G}}(y_0))$ if and only if $\operatorname{Supp}^\pm(m_0) \subset \operatorname{Ext}_{\mathcal{G}}^\pm(m_0)$.

Theorem 3 (Noise robustness, discrete case). *Let $m_0 \in \mathcal{M}(\mathbb{T})$ such that $y_0 = \Phi(m_0) \in G$. Then, there exists $\alpha > 0, \lambda_0 > 0$, such that for $(\lambda, w) \in D_{\alpha,\lambda_0}$ (defined in (6)) any solution $\tilde{m}_{\lambda,w}$ of $\mathcal{P}_\lambda^{\mathcal{G}}(y_0 + w)$ satisfies:*

$$\operatorname{Supp}^\pm(\tilde{m}_{\lambda,w}) \subset \operatorname{Ext}_{\mathcal{G}}^\pm(m_0). \quad (33)$$

If, in addition, $\Phi_{\mathrm{Ext}_{\mathcal{G}}(m_0)}$ has full rank and m_0 is a solution of $(\mathcal{P}_0^{\mathcal{G}}(y_0))$, then the solution $\tilde{m}_{\lambda,w}$ is unique, m_0 is identifiable and choosing $\lambda = \|w\|_2/\alpha$ ensures $\|\tilde{m}_{\lambda,w} - m\|_{2,\mathcal{G}} = O(\|w\|)$, where

$$\|\tilde{m}_{\lambda,w} - m\|_{2,\mathcal{G}}^2 = \sum_{x \in \mathcal{G}} |m(\{x\}) - \tilde{m}_{\lambda,w}(\{x\})|^2.$$

Proof. The proof is essentially the same as in the continuous case, therefore we only sketch it. To simplify the notation, we write $J = \mathrm{Ext}_{\mathcal{G}}(m_0)$. The solutions of $(\mathcal{D}_\lambda^{\mathcal{G}}(y_0))$ converge to $p_0^{\mathcal{G}} \in L^2(\mathbb{T})$ for $\lambda \to 0^+$, where $\Phi^* p_0^{\mathcal{G}} = \eta_0^{\mathcal{G}}$ is the discrete minimal norm certificate.

By the triangle inequality

$$\|\tilde{\eta}_\lambda - \eta_0^{\mathcal{G}}\|_{\infty,\mathcal{G}} \leqslant \underbrace{\|\tilde{\eta}_\lambda - \eta_\lambda^{\mathcal{G}}\|_{\infty,\mathcal{G}}}_{\leqslant C \frac{\|w\|_2}{\lambda}} + \|\eta_\lambda^{\mathcal{G}} - \eta_0^{\mathcal{G}}\|_{\infty,\mathcal{G}}$$

Thus, there exist two constants $\alpha > 0$ and $\lambda_0 > 0$, such that for $\frac{\|w\|_2}{\lambda} \leqslant \alpha$ and $0 < \lambda < \lambda_0$, $|\tilde{\eta}_\lambda(x)| < 1$ for any $x \in \mathcal{G} \setminus J$. Then, the primal-dual extremality conditions imply that for any solution $\tilde{m}_{\lambda,w}$ of $\mathcal{P}_\lambda^{\mathcal{G}}(y_0 + w)$, one has $\mathrm{Supp}(\tilde{m}_{\lambda,w}) \subset J$ and equality of the signs.

Now, if Φ_J has full rank, we can invert the extremality condition:

$$\frac{1}{\lambda} \Phi_J^* \left(y_0 + w - \Phi_J(\tilde{m}_{\lambda,w}|_J) \right) = \tilde{\eta}_\lambda|_J,$$

$$\text{so that} \quad \tilde{m}_{\lambda,w}|_J = m|_J + \Phi_J^+ w - \lambda (\Phi_J \Phi_J^*)^{-1} \tilde{\eta}_\lambda|_J.$$

Observing that $\|\tilde{\eta}_\lambda|_J\|_{\infty,\mathcal{G}} \leqslant 1$, we obtain the ℓ_2-robustness result. $\qquad\square$

Theorem 3 is analogous to Lemma 1 for the continuous problem. The discrete nature of the problem makes its conclusions more precise. Although the ℓ^2-robustness results are similar to those of Theorem 2, the focus here is a bit more general, in the sense that this theorem does not assert that the support of the recovered measures matches the support of the input measure m_0. In fact, if m_0 is a solution to $(\tilde{\mathcal{P}}_0^{\mathcal{G}}(y_0))$, $\mathrm{Supp}^{\pm}(m_0) \subset \mathrm{Ext}_{\mathcal{G}}^{\pm}(m_0)$, so that the recovered solutions to $\mathcal{P}_\lambda^{\mathcal{G}}(y_0 + w)$ have in general more spikes than m_0, and the spikes in $\mathrm{Ext}_{\mathcal{G}}(m_0) \setminus \mathrm{Supp}(m_0)$ must vanish as $\lambda \to 0, \|w\|_2 \to 0$.

In order to get the exact recovery of the signed support for small noise, we may assume in addition that $\mathrm{Supp}^{\pm}(m_0) = \mathrm{Ext}_{\mathcal{G}}^{\pm}(m_0)$ so as to obtain a result analogous to Theorem 2. Precisely, we obtain the following theorem which was initially proved by Fuchs [20]. First, we introduce a pre-certificate.

Definition 8 (Fuchs pre-certificate)**.** *Let $m_0 \in \mathcal{M}(\mathbb{T})$ such that $\mathrm{Supp}(m_0) \subset \mathcal{G}$. We define the Fuchs pre-certificate as*

$$\eta_{\mathrm{F}} = \operatorname*{argmin}_{\eta = \Phi^* p, p \in L^2} \|p\| \quad \text{subject to} \quad \eta|_{\mathrm{Supp}\, m_0} = \mathrm{sign}(m_0|_{\mathrm{Supp}\, m_0}). \tag{34}$$

This pre-certificate, introduced in [20], is a certificate for m_0 if and only if $\|\eta_{\mathrm{F}}\|_{\infty,\mathcal{G}} \leqslant 1$, in which case it is equal to the discrete minimal norm pre-certificate $\eta_0^{\mathcal{G}}$.

If $\Phi_{\mathrm{Supp}\, m_0}$ has full rank, then η_{F} can be computed by solving a linear system:

$$\eta_{\mathrm{F}} = \Phi^* \Phi_I^{+,*} \, \mathrm{sign}(m|_I) \quad \text{where} \quad I = \mathrm{Supp}\, m_0 \quad \text{and} \quad \Phi_I^{+,*} = \Phi_I (\Phi_I^* \Phi_I)^{-1}.$$

Corollary 3 (Exact support recovery, discrete case,[20])**.** *Let $m_0 \in \mathcal{M}(\mathbb{T})$ such that $\mathrm{Supp}(m_0) \subset \mathcal{G}$, and that $\Phi_{\mathrm{Supp}\, m_0}$ has full rank. If $|\eta_{\mathrm{F}}(t)| < 1$ for all $t \in \mathcal{G} \setminus \mathrm{Supp}\, m_0$, then m_0 is identifiable for $\mathcal{G}$ and there exists $\alpha > 0, \lambda_0 > 0$, such that for $(\lambda, w) \in D_{\alpha,\lambda_0}$ the solution $\tilde{m}_{\lambda,w}$ of $\mathcal{P}_\lambda^{\mathcal{G}}(y + w)$ is unique and satisfies $\mathrm{Supp}^{\pm}(\tilde{m}) = \mathrm{Supp}^{\pm}(m_0)$. Moreover*

$$\tilde{m}_{\lambda,w}|_I = m_0|_I + \Phi_I^+ w - \lambda(\Phi_I \Phi_I^*)^{-1} \mathrm{sign}(m_0|_I), \tag{35}$$

where $I = \mathrm{Supp}\, m_0$.

The condition $|\eta_{\mathrm{F}}(t)| < 1$ for all $t \in \mathcal{G} \setminus \mathrm{Supp}\, m_0$ is often called the irrepresentability condition in the statistics literature, see [34]. This condition can be shown to be almost a necessary and sufficient condition to ensure exact recovery of the support of m_0. For instance, if $|\eta_{\mathrm{F}}(t)| > 1$ for some $t \in \mathcal{G} \setminus \mathrm{Supp}\, m_0$, one can show that for all $\lambda > 0$ $\mathrm{Supp}(\tilde{m}_\lambda) \neq \mathrm{Supp}\, m_0$ where $\tilde{m}_\lambda$ is any solution of $\mathcal{P}_\lambda^{\mathcal{G}}(y_0)$, see [33]. In our framework, we see that this irrepresentability condition means that the precertificate η_{F} is indeed a certificate (so that it is equal to the minimal norm certificate), and that its saturation set is equal to the support of m_0.

For deconvolution problems, an important issue is that Corollary 3 is useless when studying the stability of the original infinite dimensional problem $(\mathcal{P}_\lambda(y_0))$. Indeed, the pre-certificate (34) is not constrained to have vanishing derivatives, so that it generally takes some values strictly greater than 1 for a generic discrete input measure m_0. When the stepsize of the grid is small enough, such values are sampled and $\|\eta_{\mathrm{F}}\|_{\infty,\mathcal{G}}$ necessarily becomes strictly larger than one. As detailed in Section 4, when shifting from the discrete grid setting to the continuous setting, the natural pre-certificate to consider is the vanishing derivative pre-certificate η_{V} defined in (27), and not the pre-certificate η_{F}.

5.4 Structure of the Extended Support for Thin Grids

In the previous section, we have introduced the notion of extended signed support of a measure m_0 relatively to a grid $\mathcal{G}$, and we have proved that this set, $\mathrm{Ext}_{\mathcal{G}}^{\pm}\, m_0$, contains the signed supports of all the reconstructed measures for small noise. In this section, we focus on the structure of the extended support. We show that, if the support of m_0 belongs to the grid for a sufficiently small stepsize and if the Non Degenerate Source Condition holds, the extended signed support consists in the signed support of m_0 and possibly one immediate neighbor with the same sign for each spike. Therefore, when the grid stepsize is small enough, the support of the measure is generally not stable for the discrete

problem, but the support of the reconstructed measure is a close approximation of the original one.

From now on, for the sake of simplicity, we consider dyadic grids $\mathcal{G}_n = \left\{ \frac{j}{2^n} \; ; \; 0 \leqslant j \leqslant 2^n - 1 \right\}$. The constraint sets in $\mathcal{D}_\lambda^{\mathcal{G}_n}(y_0)$ and $(\mathcal{D}_\lambda(y_0))$ are denoted respectively by

$$C_n = \left\{ p \in L^2(\mathbb{T}) \; ; \; \left| (\Phi^* p) \left(\frac{j}{2^n} \right) \right| \leqslant 1, \; 0 \leqslant j \leqslant 2^n \right\}, \tag{36}$$

$$\text{and } C = \left\{ p \in L^2(\mathbb{T}) \; ; \; \| \Phi^* p \|_\infty \leqslant 1 \right\} = \bigcap_{n \in \mathbb{N}} C_n. \tag{37}$$

The structure of $\mathrm{Ext}_{\mathcal{G}}^{\pm}(m_0)$ for large n is intimately related to the convergence of $p_0^{\mathcal{G}_n}$ to p_0. First, let us notice the following result, whose proof is given in Appendix C.

Proposition 9 (Convergence for fixed λ). *Let $m_0 \in \mathcal{M}(\mathbb{T})$. Then, for any $\lambda > 0$,*

$$\lim_{n \to +\infty} p_\lambda^{\mathcal{G}_n} = p_\lambda \text{ for the } L^2(\mathbb{T}) \text{(strong) topology,} \tag{38}$$

$$\text{and } \lim_{n \to +\infty} \eta_\lambda^{\mathcal{G}_n} = \eta_\lambda \text{ for the topology of the uniform convergence.} \tag{39}$$

Moreover, if there exists a solution to the continuous dual problem $(\mathcal{D}_0(y_0))$,

$$\lim_{\lambda \to 0^+} \lim_{n \to +\infty} p_\lambda^{\mathcal{G}_n} = p_0, \text{ and } \lim_{\lambda \to 0^+} \lim_{n \to +\infty} \eta_\lambda^{\mathcal{G}_n} = \eta_0. \tag{40}$$

Proposition 9 simply states that the projection onto convex sets C_n which converge (in the sense of set convergence) to C converges to the projection onto C. However, the case $\lambda = 0$ is not as straightforward, and for instance one cannot easily swap the limits in (40). In fact, given any decreasing sequence of polyhedra C_n, it is not true in general that the minimal norm solution of $\sup_{p \in C_n} \langle y_0, p \rangle$ should converge to the minimal norm solution of $\sup_{p \in C} \langle y_0, p \rangle$ where $C = \bigcap_{n \in \mathbb{N}} C_n$. As a consequence it is not clear to us whether this convergence always holds for polyhedra of the form

$$C_n = \left\{ p \in L^2 \; ; \; \| \Phi^* p \|_{\infty, \mathcal{G}_n} \leqslant 1 \right\}.$$

However, when the spikes locations belong to the grid for n large enough, the convergence of the minimal norm certificates holds. In the case of dyadic grids, this is equivalent to $m_0 \in \mathcal{M}(\mathbb{T})$ being a *discrete dyadic measure*, i.e. such that for some $n_0 \in \mathbb{N}$:

$$m = \sum_{i=1}^{N} a_i \delta_{x_i}, \quad \text{with} \quad x_i = \frac{j_i}{2^{n_0}} \text{ and } 0 \leqslant j_i \leqslant 2^{n_0} - 1. \tag{41}$$

The proofs given below make use of a remark given in [8]: if a solution of the continuous problem $(\mathcal{P}_0(y_0))$ has support in the grid $\mathcal{G}$, then it is also a solution of the discrete problem $(\mathcal{P}_0^{\mathcal{G}}(y_0))$.

Proposition 10 (Convergence for dyadic measures). *Let $m_0 \in \mathcal{M}(\mathbb{T})$ be a discrete dyadic measure (see (41)), and assume that the (possibly degenerate) source condition holds. Then*

$$\lim_{n \to +\infty} p_0^{\mathcal{G}_n} = p_0 \text{ for the } L^2 \text{ (strong) topology,} \tag{42}$$

and $\displaystyle \lim_{n \to +\infty} \eta_0^{\mathcal{G}_n(i)} = \eta_0^{(i)} \text{ for } 0 \leqslant i \leqslant 2, \text{ in the sense of the uniform convergence,}$

$$\tag{43}$$

where $\eta_0 = \Phi^* p_0$ (resp. $\eta_0^{\mathcal{G}_n} = \Phi^* p_0^{\mathcal{G}_n}$) denotes the corresponding minimal norm certificate.

Proof. First, following [8], we observe that, since $(\Phi^* p_0)(x_i) = \text{sign}(a_i)$ and $\|\Phi^* p_0\|_\infty \leqslant 1$ (a fortiori $|\Phi^* p_0\left(\frac{j}{2^n}\right)| \leqslant 1$ for $1 \leqslant j \leqslant 2^n - 1$), $\Phi^* p_0$ is also a dual certificate for $(\mathcal{P}_0^{\mathcal{G}_n})$ provided $n \geqslant n_0$. As a consequence $\|p_0^{\mathcal{G}_n}\|_2 \leqslant \|p_0\|_2$.

The sequence $(p_0^{\mathcal{G}_n})_{n \in \mathbb{N}}$ being bounded in $L^2(\mathbb{T})$, we may extract a subsequence (still denoted by $p_0^{\mathcal{G}_n}$) which weakly converges to some $\tilde{p} \in L^2(\mathbb{T})$, and

$$\|\tilde{p}\|_2 \leqslant \liminf_{n \to +\infty} \|p_0^{\mathcal{G}_n}\|_2 \leqslant \limsup_{n \to +\infty} \|p_0^{\mathcal{G}_n}\|_2 \leqslant \|p_0\|_2. \tag{44}$$

Moreover, by optimality of $p_0^{\mathcal{G}_n}$ for the discrete problem, for each $p \in C \subset C_n$, $\langle y_0, p_0^{\mathcal{G}_n} \rangle \geqslant \langle y_0, p \rangle$ so that in the limit $\langle y_0, \tilde{p} \rangle \geqslant \langle y_0, p \rangle$. Observing that $\tilde{p} \in C = \bigcap_{n \in \mathbb{N}} C_n$ (since each C_n is weakly closed) we conclude that $\tilde{p} = p_0$. Since the limit does not depend on the extracted subsequence, we conclude that the whole sequence $(p_0^{\mathcal{G}_n})_{n \in \mathbb{N}}$ converges to p_0, and equality in (44) implies that the convergence is strong.

The consequence regarding $\eta_0^{\mathcal{G}_n}$ is straightforward. $\square$

We may now describe the structure of the extended support for dyadic measures which satisfy the Non Degenerate Source Condition.

Proposition 11 (Extended support). *Let $m_0 = \sum_{i=1}^{N} a_i \delta_{x_i}$ be a discrete dyadic measure which satisfies the Non Degenerate Source Condition. Then, for n large enough, there exists $\varepsilon^n \in \{+1, -1\}^N$ such that:*

$$\text{Supp}^\pm(m_0) \subset \text{Ext}^\pm{}_{\mathcal{G}_n}(m_0) \subset \text{Supp}^\pm(m_0) \cup \left(\text{Supp}^\pm(m_0) + \frac{\varepsilon^n}{2^n} \right), \tag{45}$$

where $\text{Supp}^\pm(m_0) + \frac{\varepsilon^n}{2^n} = \left\{ (x_i + \frac{\varepsilon_i^n}{2^n}, \eta_0^{\mathcal{G}_n}(x_i)) \; ; \; 1 \leqslant i \leqslant N \right\}$.

Corollary 4. *Under the hypotheses of Proposition 11, for n large enough, there exist two constants $\alpha(n) > 0$ and $\lambda_0(n) > 0$ such that, for $\frac{\|w\|_2}{\lambda} < \alpha(n)$ and $0 < \lambda < \lambda_0(n)$, any solution $\tilde{m}_\lambda^{\mathcal{G}_n}$ of $(\mathcal{P}_\lambda^{\mathcal{G}_n})$ has support in $\{x_i, \; 1 \leqslant i \leqslant N\} \cup \{x_i + \frac{\varepsilon_i^n}{2^n}, \; 1 \leqslant i \leqslant N\}$, with signs $\eta_0^{\mathcal{G}_n}(x_i), 1 \leqslant i \leqslant N$.*

Proof of Proposition 11. We describe the points where the value of $\eta_0^{\mathcal{G}_n}$ may be ± 1. By the Non-Degenerate Source Condition, there exists $\varepsilon > 0$ small enough

such that the intervals $(x_{0,i} - \varepsilon, x_{0,i} + \varepsilon)$, $1 \leqslant i \leqslant N$, do not intersect, and that for all $t \in \bigcup_{i=1}^{N}(x_i - \varepsilon, x_i + \varepsilon)$, $|\eta_0(t)| \geqslant C > 0$ and $|\eta_0''(t)| \geqslant C > 0$. Moreover, $\sup_{K_\varepsilon} |\eta_0| < 1$ with $K_\varepsilon = \mathbb{T} \setminus \bigcup_{i=1}^{N}(x_i - \varepsilon, x_i + \varepsilon)$.

Therefore, by Proposition 10, for n large enough:

- $|\eta_0^{\mathcal{G}_n}(t)| \geqslant \frac{C}{2} > 0$ for $t \in (x_{i,0} - \varepsilon, x_{i,0} + \varepsilon)$,

- $|(\eta_0^{\mathcal{G}_n})''(t)| \geqslant \frac{C}{2} > 0$ for $t \in (x_i - \varepsilon, x_i + \varepsilon)$,

- $\sup_{K_\varepsilon} |\eta_0^{\mathcal{G}_n}| < 1$,

and in each interval $(x_{i,0} - \varepsilon, x_i + \varepsilon)$, $\eta_0^{\mathcal{G}_n}$ has the same sign as η_0 and it is strictly concave (resp. strictly convex) if $\eta_0(x_i) = 1$ (resp. -1).

Assume for instance that $\eta_0(x_i) = 1$. The extremality conditions between p_0 and m_0 for $(\mathcal{P}_0(y))$ also imply that m_0 is a solution of $(\mathcal{P}_0^{\mathcal{G}_n}(y_0))$. Then, the extremality conditions between $p_0^{\mathcal{G}_n}$ and m_0 imply that $\eta_0^{\mathcal{G}_n}(x_i) = 1$ as well. By the strict concavity of $\eta_0^{\mathcal{G}_n}$ there is at most one other point $t^\star \in (x_i - \varepsilon, x_i + \varepsilon)$ such that $\eta_0^{\mathcal{G}_n}(t^\star) = 1$, and since $\eta_0^{\mathcal{G}_n}(x_i \pm \frac{1}{2^n}) \leqslant 1$, $|t^\star - x_i| \leqslant \frac{1}{2^n}$. Such a point $t^\star$ contributes to the extended support of m if and only if it belongs to the grid (i.e. $t^\star = x_i \pm \frac{1}{2^n}$).

The argument for $\eta_0(x_i) = -1$ is similar. This concludes the proof. $\qquad \square$

Corollary 4 highlights the difference between the continuous and the discretized problems. In the first case, any small noise would induce a slight perturbation of the spikes locations and amplitudes, but their number would stay the same. In the second case, the spikes cannot "move", so that new spikes may appear, but only at one of the immediate neighbors of the original ones.

For non-dyadic measures, we may show using Proposition 9 that for small, fixed $\lambda > 0$, and n large enough, there is at most one pair of spikes (located at consecutive points of the grid) in the neighborhood of each original spike. From our numerical experiments described below (in the case of the ideal low-pass filter), we conjecture that, in the case where there are indeed two spikes, they surround the location of the original spike.

5.5 Application to the Ideal Low-pass Filter

To conclude this section, we compare the different (pre-)certificates involved in the above discussion, whether on the discrete grid or in the continuous domain. Then we illustrate the convergence of the sets $(C_n)_{n \in \mathbb{N}}$ towards C.

Certificates. Figure 5 illustrates the results of Section 5.4. The numerical values are $f_c = 6$, $n = 7$, and the distance between the two opposite spikes is $\frac{0.6}{f_c}$. The continuous minimal norm certificate η_0 is shown: it satisfies $|\eta_0(t)| \leqslant 1$ for all $t \in \mathbb{T}$ and $\eta_0(x_i) = \operatorname{sign} m_0(\{x_i\})$ for $1 \leqslant i \leqslant N$. The discrete minimal norm certificate $\eta_0^{\mathcal{G}_n}$ satisfies $|\eta_0^{\mathcal{G}_n}(t)| \leqslant 1$ for all t in the grid, and $\eta(u) = \operatorname{sign} m_0(\{x_i\})$ for all $u \in \operatorname{Ext}_{\mathcal{G}_n}$ in the neighborhood of x_i. For a dyadic measure, such points

are x_i and possibly one of its immediate neighbors. For non dyadic measures, we conjecture that such points are the two immediate neighbors of x_i.

The Fuchs precertificate η_{F} is also shown. Some points t of the grid do not satisfy $|\eta_{\mathrm{F}}(t)| \leqslant 1$, hence the Fuchs pre-certificate is not a certificate and the support is not stable. This was already clear from the fact that $\mathrm{Supp}(m_0) \subsetneq \mathrm{Ext}_{\mathcal{G}_n}(m_0)$.
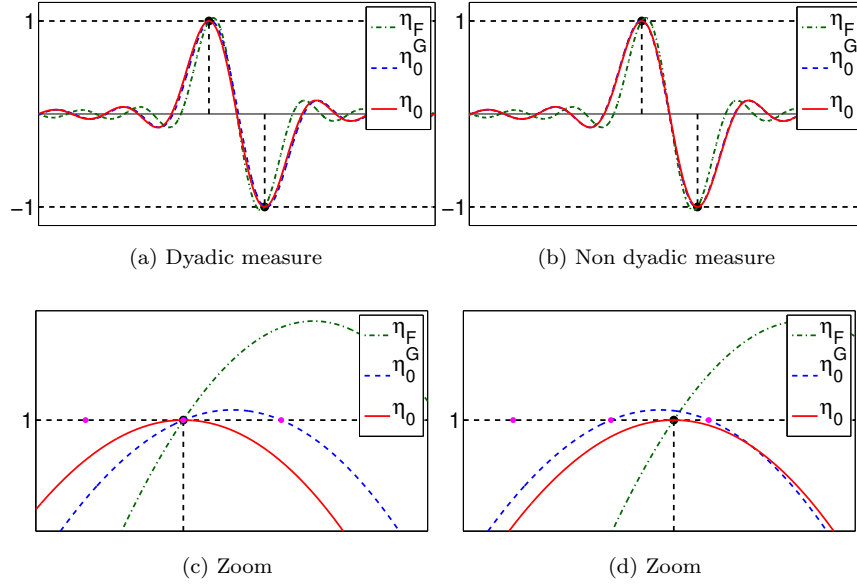

(a) Dyadic measure   (b) Non dyadic measure

(c) Zoom   (d) Zoom

Figure 5: Comparison of certificates for a dyadic (left) and a non-dyadic measure (right). The second row is a zoom of the first one near the left spike. The (continuous) minimum norm certificate η_0 (in continuous red line) is everywhere bounded by 1. The (discrete) minimum norm certificate $\eta_0^{\mathcal{G}_n}$ (in dashed blue line) is bounded by 1 at the grid points. The Fuchs pre-certificate η_{F} (dash-dot green line) is above 1 at some points of the grid: the Fuchs criterion is not satisfied.

Figure 6 focuses on the reconstructed amplitudes $\tilde{a}_i$ using $\mathcal{P}_\lambda(y_0)$ as $\lambda \to 0$. Each curve represents a path $\lambda \mapsto \tilde{a}_i$. Note that for the problem on a finite grid, such paths are piecewise affine. In the dyadic case (left part of the figure), the amplitude at x_i (continuous line) and at the next point of the grid (dashed line) are shown. As $\lambda \to 0$, the spike at the neighbor vanishes and the result tends to the original identifiable measure. In the non dyadic case (right part of the figure), the amplitude at the two immediate neighbors of x_i are shown (continuous and dashed lines). Here $\mathrm{Supp}\, m_0 \not\subset \mathcal{G}$ so that m_0 is not identifiable for the discrete problem. For each spike, the amplitudes of the two neighbors converge to some non zero value. The limit measure as $\lambda \to 0$ is the solution of $\mathcal{P}_0(y_{0G})$.

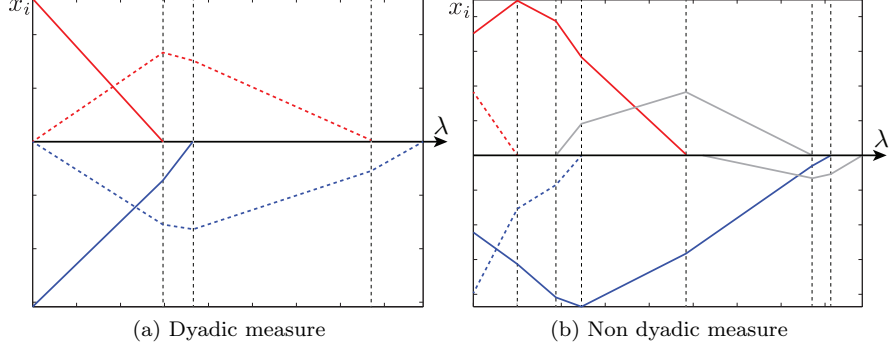(a) Dyadic measure   (b) Non dyadic measure

Figure 6:   Display of the solution path (as a function of λ) for the measure displayed on Figure 5. Left: Amplitudes of the coefficients at x_i (continuous line) and at the next point of the grid (dashed line) as λ varies. Right: idem for the two immediate neighbors of x_i. Some other spikes (grey continuous line) appear and vanish before the last segments, as $\lambda \to 0$.

Set convergence.   Now, we interpret the convergence of the discrete problems through the convergence of the corresponding constraint set for the dual problem. Writing $\Phi^* p(x) = \int p(t)\varphi(x-t)\mathrm{d}t = \langle p, \varphi_x \rangle_{L^2}$ with $\varphi_x : t \mapsto \varphi(x-t)$, we observe that:

$$C_n = \left\{ p \in \operatorname{Im} \Phi \; ; \; \left| \Phi^* p\left(\frac{j}{2^n} \right) \right| \leqslant 1, \; 0 \leqslant j \leqslant 2^n - 1 \right\} \tag{46}$$

$$= \left\{ p \in \operatorname{Im} \Phi \; ; \; |\langle p, \varphi_{\frac{j}{2^n}} \rangle_{L^2}| \leqslant 1, \; 0 \leqslant j \leqslant 2^n - 1 \right\}. \tag{47}$$

As a consequence C_n is the polar set of the convex hull of $\left\{ \pm \varphi_{\frac{j}{2^n}} \; ; \; 0 \leqslant j \leqslant 2^n - 1 \right\}$.

In the case of the Dirichlet kernel, the vector space $\operatorname{Im} \Phi$ is the space of trigonometric polynomials with degree less than or equal to f_c. An orthonormal basis of $\operatorname{Im} \Phi$ is given by: $(c_0, c_1, \ldots c_{f_c}, s_1, \ldots s_{f_c})$ where $c_0 \equiv 1$, $c_k : t \mapsto \sqrt{2}\cos(2\pi kt)$ and $s_k : t \mapsto \sqrt{2}\sin(2\pi kt)$ for $1 \leqslant k \leqslant f_c$.

Moreover,

$$\varphi(x-t) = \frac{1}{2f_c + 1} \left(1 + \sum_{k=1}^{f_c} 2\cos(2\pi k(x-t)) \right)$$

$$= \frac{1}{2f_c + 1} \left(1 + 2\sum_{k=1}^{f_c} \left(\cos(2\pi kx)\cos(2\pi kt) + \sin(2\pi kx)\sin(2\pi kt) \right) \right)$$

so that we may write:

$$\varphi_x = \frac{1}{2f_c + 1} \left(c_0 + \sqrt{2}\sum_{k=1}^{f_c} \left(\cos(2\pi kx)c_k + \sin(2\pi kx)s_k \right) \right).$$
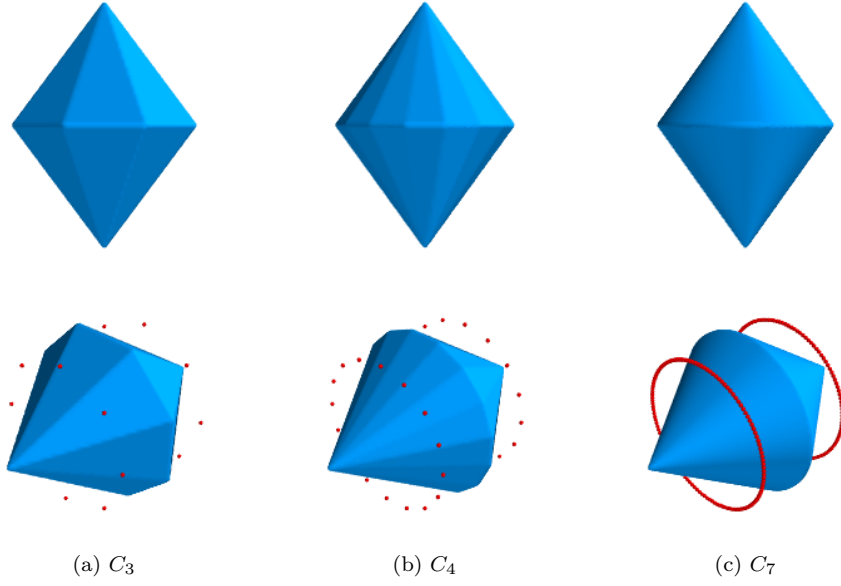
35

(a) C_3 (b) C_4 (c) C_7

Figure 7: Top: The convex set C_n for $f_c = 1$, and $n = 3$, 4 or 7 (from left to right). Bottom: same convex sets, the red spheres indicate the (rescaled) vectors $\varphi_{\frac{j}{2^n}}$.

For $f_c = 1$, we obtain $\varphi_x = \frac{1}{3} \left(c_0 + \sqrt{2} \left(\cos(2\pi x) c_1 + \sin(2\pi x) s_1 \right) \right)$, and the vectors φ_x lie on a circle. The convex hull of $\left\{ \pm \varphi_{\frac{j}{2^n}} \; ; \; 0 \leqslant j \leqslant 2^n - 1 \right\}$ is thus a cylinder, and its polar set C_n is displayed in Figure 7 for $n = 3$, 4, and 7.

Problem $(\mathcal{D}_\lambda^{\mathcal{G}_n}(y_0 + w))$ corresponds to the projection of $\frac{y_0 + w}{\lambda}$ onto the polytope C_n. Each face of C_n corresponds to a possible signed support of the solutions $\tilde{m}_{\lambda,w}$. The large, flat faces of C_n yield stability to the support of $\tilde{m}_{\lambda,w}$ for small noise w, as described by Theorem 3. As $n \to +\infty$ these faces converge into a piecewise smooth manifold and the support of $\tilde{m}_{\lambda,w}$ is allowed to vary smoothly in $\mathbb{T}$, according to Theorem 2.

Conclusion

In this paper, we have given a precise statement about the support recovery property of sparse spikes deconvolution with total variation regularization. This support recovery is governed by the non-degeneracy of a minimal norm certificate. This hypothesis can be checked by computing a vanishing derivative pre-certificate, which can be computed in closed form. We have shown that under this non-degeneracy hypothesis, one recovers the same number of spikes and that these spikes converge to the original ones when λ and $\|w\|/\lambda$ are small enough. While previous stability results [7, 19, 1] hold for an arbitrary noise level and make use of any non-degenerate certificate, they are formulated in terms of local averages of the recovered measure and do not describe precisely the support. In contrast, our result which requires a specific certificate to be non-degenerate and a regime where λ and $\|w\|/\lambda$ are small enough provides exact support stability. These settings and results are thus not comparable, and provide complementary informations about the performance of total variation regularization.

Developing a similar framework for the discrete ℓ^1 setting, we have also improved upon existing results about stability of the support by introducing the notion of extended support of a measure. Our study highlights the difference between the continuous and the discrete case: when the size of the grid is small enough, the stable recovery of the support is generally not possible in the discrete framework. Yet, in the non degenerate case, the reconstructed support at small noise is a slight modification of the original one: each original spike yields at most one pair of consecutive spikes which surround it.

Finally, let us note that the proposed method extends to non-stationary filtering operators and to arbitrary dimensions.

Acknowledgements

A   Auxiliary results

For the convenience of the reader, we give here the proofs of several auxiliary results which are needed in the discussion.

Proposition 12 (Subdifferential of the total variation). *Let us endow $\mathcal{M}(\mathbb{T})$ with the weak-* topology and $C(\mathbb{T})$ with the weak topology. Then, for any $m \in \mathcal{M}(\mathbb{T})$, we have:*

$$\partial \|m\|_{\mathrm{TV}} = \left\{ \eta \in C(\mathbb{T}) \; ; \; \|\eta\|_\infty \leqslant 1 \quad and \quad \int \eta \, \mathrm{d}m = \|m\|_{\mathrm{TV}} \right\}.$$

Proof. Let $A = \{\eta \in C(\mathbb{T}) \; ; \; \forall m \in \mathcal{M}(\mathbb{T}), \; \langle \eta, \, m \rangle \leqslant \|m\|_{\mathrm{TV}}\}$. It is clear that $B_\infty(0,1) \subset A$, where $B_\infty(0,1)$ is the $L^\infty(\mathbb{T})$ closed unit ball. Conversely, we observe that $A \subset B_\infty(0,1)$ by considering the Dirac masses $(\pm \delta_t)_{t \in \mathbb{T}}$.

Let us write $J(m) := \|m\|_{\mathrm{TV}}$. The function $J : \mathcal{M}(\mathbb{T}) \to \mathbb{R} \cup \{+\infty\}$ is convex, proper, lower semi-continuous (for the weak-* topology), positively homogeneous and:

$$J^*(\eta) = \sup_{m \in \mathcal{M}(\mathbb{T})} \sup_{t > 0} \left(\langle \eta, \, tm \rangle - J(tm) \right)$$

$$= \sup_{t > 0} t \left(\sup_{m \in \mathcal{M}(\mathbb{T})} \langle \eta, \, m \rangle - J(m) \right)$$

$$= \begin{cases} 0 & \text{if } \eta \in A, \\ +\infty & \text{otherwise.} \end{cases}$$

By Proposition I.5.1 in [17], for any $\eta \in C(\mathbb{T})$:

$$\eta \in \partial J(m) \iff \langle \eta, \, m \rangle = J(m) + J^*(\eta),$$

which is equivalent to $\|\eta\|_\infty \leqslant 1$ and $\int \eta \mathrm{d}m = \|m\|_{\mathrm{TV}}$. $\qquad\square$

Proposition 13. *There exists a solution to $(\mathcal{P}_0(y_0))$ and the strong duality holds between $(\mathcal{P}_0(y_0))$ and $(\mathcal{D}_0(y_0))$, i.e.*

$$\min_{\Phi(m) = y_0} \|m\|_{\mathrm{TV}} = \sup_{\|\Phi^* p\|_\infty \leqslant 1} \langle y_0, \, p \rangle. \tag{48}$$

Moreover, if a solution $p^\star$ to $(\mathcal{D}_0(y_0))$ exists,

$$\Phi^* p^\star \in \partial \|m^\star\|_{\mathrm{TV}} \tag{49}$$

where $m^\star$ is any solution to $(\mathcal{P}_0(y_0))$. Conversely, if (49) holds, then $m^\star$ and $p^\star$ are solutions of respectively $(\mathcal{P}_0(y_0))$ and $(\mathcal{D}_0(y_0))$.

Proof. We apply [17, Theorem II.4.1] to $(\mathcal{D}_0(y_0))$ (and not to $(\mathcal{P}_0(y_0))$ as would be natural) rewritten as

$$\inf_{\|\Phi^* p\|_\infty \leqslant 1} \langle -y_0, \, p \rangle,$$

The infimum is finite since for any admissible p, $\langle -y_0, p \rangle = \langle m_0, \Phi p \rangle \geqslant -\|m_0\|_{\mathrm{TV}}$. Let $V = L^2(\mathbb{T})$, $Y = C(\mathbb{T})$ (endowed with the strong topology), $Y^* = \mathcal{M}(\mathbb{T})$, $F(u) = \langle -y_0, u \rangle$ for $u \in V$, $G(\psi) = \iota_{\|\cdot\|_\infty \leqslant 1}(\psi)$ for $\psi \in Y$ and $\Lambda = \Phi^*$. It is clear that F and G are proper convex lower semi-continuous functions. Eventually, F is finite at 0, G is finite and continuous at $0 = \Lambda 0$. Hence the result. $\qquad\square$

B   Proof of Proposition 6

Assume that for some $(u, v) \in \mathbb{R}^N \times \mathbb{R}^N$, $\Gamma_x(u, v) = 0$. Then

$$\forall\, t \in \mathbb{T}, \quad 0 = \sum_{j=1}^{N} \left(u_j \varphi(t - x_j) + v_j \varphi'(t - x_j) \right)$$

$$= \sum_{k=-f_c}^{f_c} \left(\sum_{j=1}^{N} (u_j + 2ik\pi v_j) e^{-2ik\pi x_j} \right) e^{2ik\pi t}$$

We deduce that

$$\forall\, k \in \{-f_c, \ldots f_c\}, \quad \sum_{j=1}^{N} (u_j + k\tilde{v}_j) r_j^k = 0 \quad \text{where} \quad \begin{cases} r_j = e^{-2i\pi x_j}, \\ \tilde{v}_j = 2i\pi v_j. \end{cases}$$

It is therefore sufficient to prove that the columns of the following matrix are linearly independent

$$\begin{pmatrix} r_1^{-f_c} & \cdots & r_N^{-f_c} & (-f_c) r_1^{-f_c} & \cdots & (-f_c) r_N^{-f_c} \\ \vdots & & \vdots & \vdots & & \vdots \\ r_1^k & \cdots & r_N^k & k r_1^k & \cdots & k r_N^k \\ \vdots & & \vdots & \vdots & & \vdots \\ r_1^{f_c} & \cdots & r_N^{f_c} & (f_c) r_1^{f_c} & \cdots & (f_c) r_N^{f_c} \end{pmatrix}.$$

If $N < f_c$, we complete the family $\{r_1, \ldots r_N\}$ in a family $\{r_0, r_1, \ldots r_{f_c}\} \subset \mathbb{S}^1$ such that the r_i's are pairwise distinct. We obtain a square matrix M by inserting the corresponding columns

$$M = \begin{pmatrix} r_1^{-f_c} & \cdots & r_{f_c}^{-f_c} & r_0^{-f_c} & (-f_c) r_1^{-f_c} & \cdots & (-f_c) r_{f_c}^{-f_c} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ r_1^k & \cdots & r_{f_c}^k & r_0^k & k r_1^k & \cdots & k r_{f_c}^k \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ r_1^{f_c} & \cdots & r_{f_c}^{f_c} & r_0^{f_c} & (f_c) r_1^{f_c} & \cdots & (f_c) r_{f_c}^{f_c} \end{pmatrix}.$$

We claim that M is invertible. Indeed, if there exists $\alpha \in \mathbb{C}^{(2f_c+1)}$ such that $M^T \alpha = 0$, then the rational function $F(z) = \sum_{k=-f_c}^{f_c} \alpha_k z^k$ satisfies:

$$F(r_j) = 0 \quad \text{and} \quad F'(r_j) = 0 \ \text{ for } 1 \leqslant j \leqslant f_c,$$
$$F(r_0) = 0.$$

Hence, F has at least $2f_c + 1$ roots in $\mathbb{S}^1$, counting the multiplicities. This imposes that $F = 0$, thus $\alpha = 0$, and M is invertible. The result is proved.

C Proof of Proposition 9

Let us denote by $P_{C_n}(x)$ the projection of $x \in L^2(\mathbb{T})$ onto C_n. We have:

$$\left\| P_{C_n}(\frac{y_0}{\lambda}) - P_{C_n}(0) \right\|_2 \leqslant \left\| \frac{y_0}{\lambda} - 0 \right\|_2,$$

so that the sequence $p_\lambda^{\mathcal{G}_n} = P_{C_n}(\frac{y_0}{\lambda})$ is bounded in $L^2(\mathbb{T})$, and we may extract a subsequence $p_\lambda^{\mathcal{G}'_n}$ which weakly converges to some $p_\lambda^\star \in L^2(\mathbb{T})$. Since $C_{n'}$ is (weakly) closed for all n', $p_\lambda^\star \in \bigcap_{n'} C_{n'} = C$.

Moreover, by the characterization of the projection onto convex sets:

$$\forall z \in C \subset C'_n, \; \left\langle \frac{y_0}{\lambda} - p_\lambda^{\mathcal{G}'_n}, z \right\rangle - \left\langle \frac{y_0}{\lambda}, p_\lambda^{\mathcal{G}'_n} \right\rangle + \|p_\lambda^{\mathcal{G}'_n}\|_2^2 \leqslant 0.$$

Passing to the limit $n' \to +\infty$, $\left\langle \frac{y_0}{\lambda} - p_\lambda^\star, z \right\rangle - \left\langle \frac{y_0}{\lambda}, p_\lambda^\star \right\rangle + \liminf_{n'} \|p_\lambda^{\mathcal{G}'_n}\|_2^2 \leqslant 0,$

$$\left\langle \frac{y_0}{\lambda} - p_\lambda^\star, z \right\rangle - \left\langle \frac{y_0}{\lambda}, p_\lambda^\star \right\rangle + \|p_\lambda^\star\|_2^2 \leqslant 0,$$

$$\left\langle \frac{y_0}{\lambda} - p_\lambda^\star, z - p_\lambda^\star \right\rangle \leqslant 0.$$

Thus $p_\lambda^\star$ is the orthogonal projection of $\frac{y_0}{\lambda}$ on C: $p_\lambda^\star = P_C\left(\frac{y_0}{\lambda}\right) = p_\lambda$. Since this is true for any subsequence, the whole sequence $p_\lambda^{\mathcal{G}_n}$ weakly converges to p_λ.

Moreover, by lower semicontinuity and the inclusion $C \subset C_n$ we have:

$$\left\| \frac{y_0}{\lambda} - p_\lambda \right\|_2 \leqslant \liminf_{n \to +\infty} \left\| \frac{y_0}{\lambda} - p_\lambda^{\mathcal{G}_n} \right\|_2 \leqslant \limsup_{n \to +\infty} \left\| \frac{y_0}{\lambda} - p_\lambda^{\mathcal{G}_n} \right\|_2 \leqslant \left\| \frac{y_0}{\lambda} - p_\lambda \right\|_2,$$

so that $\frac{y_0}{\lambda} - p_\lambda^{\mathcal{G}_n}$ converges strongly to $\frac{y_0}{\lambda} - p_\lambda$, hence the strong convergence of $p_\lambda^{\mathcal{G}_n}$ to p_λ.

The rest of the statement follows from Proposition 1.

References

[1] J-M. Azais, Y. De Castro, and F. Gamboa. Spike detection from inaccurate samplings. *Applied and Computational Harmonic Analysis, to appear*, 2014.

[2] B.N. Bhaskar and B. Recht. Atomic norm denoising with applications to line spectral estimation. In *2011 49th Annual Allerton Conference on Communication, Control, and Computing*, pages 261–268, 2011.

[3] T. Blu, P.-L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot. Sparse sampling of signal innovations. *IEEE Signal Processing Magazine*, 25(2):31–40, 2008.

[4] K. Bredies and H.K. Pikkarainen. Inverse problems in spaces of measures. *ESAIM: Control, Optimisation and Calculus of Variations*, 19(1):190–218, 2013.

[5] H. Brézis, P.G. Ciarlet, and J.L. Lions. *Analyse fonctionnelle: Théorie et applications*. Collection Mathématiques appliquées pour la maîtrise. Dunod, 1999.

[6] M. Burger and S. Osher. Convergence rates of convex variational regularization. *Inverse Problems*, 20(5):1411–1421, 2004.

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

[8] 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.

[9] S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM journal on scientific computing*, 20(1):33–61, 1999.

[10] J. F. Claerbout and F. Muir. Robust modeling with erratic data. *Geophysics*, 38(5):826–844, 1973.

[11] L. Condat and A. Hirabayashi. Cadzow denoising upgraded: A new projection method for the recovery of dirac pulses from noisy linear measurements. *Sampling Theory in Signal and Image Processing, to appear*, 2014.

[12] Y. de Castro and F. Gamboa. Exact reconstruction using beurling minimal extrapolation. *Journal of Mathematical Analysis and Applications*, 395(1):336–354, 2012.

[13] L. Demanet, D. Needell, and N. Nguyen. Super-resolution via superset selection and pruning. *CoRR*, abs/1302.6288, 2013.

[14] D. L. Donoho. Superresolution via sparsity constraints. *SIAM J. Math. Anal.*, 23(5):1309–1331, September 1992.

[15] C. Dossal and S. Mallat. Sparse spike deconvolution with minimum scale. In *Proceedings of SPARS*, pages 123–126, November 2005.

[16] M. F. Duarte and R. G. Baraniuk. Spectral compressive sensing. *Applied and Computational Harmonic Analysis*, 35(1):111–129, 2013.

[17] I. Ekeland and R. Témam. *Convex Analysis and Variational Problems*. Number vol. 1 in Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 1976.

[18] A. Fannjiang and W. Liao. Coherence pattern-guided compressive sensing with unresolved grids. *SIAM J. Img. Sci.*, 5(1):179–202, February 2012.

[19] C. Fernandez-Granda. Support detection in super-resolution. *Proc. Proceedings of the 10th International Conference on Sampling Theory and Applications*, pages 145–148, 2013.

[20] J.J. Fuchs. On sparse representations in arbitrary redundant bases. *IEEE Transactions on Information Theory*, 50(6):1341–1344, 2004.

[21] M. Grasmair, O. Scherzer, and M. Haltmeier. Necessary and sufficient conditions for linear convergence of ℓ_1-regularization. *Communications on Pure and Applied Mathematics*, 64(2):161–182, 2011.

[22] B. Hofmann, B. Kaltenbacher, C. Poschl, and O. Scherzer. A convergence rates result for tikhonov regularization in Banach spaces with non-smooth operators. *Inverse Problems*, 23(3):987, 2007.

[23] S. Levy and P. Fullagar. Reconstruction of a sparse spike train from a portion of its spectrum and application to high-resolution deconvolution. *Geophysics*, 46(9):1235–1243, 1981.

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

[25] D. A. Lorenz and D. Trede. Greedy Deconvolution of Point-like Objects. In Rémi Gribonval, editor, *SPARS'09*, Saint Malo, France, 2009.

[26] J. W. Odendaal, E. Barnard, and C. W. I. Pistorius. Two-dimensional superresolution radar imaging using the MUSIC algorithm. *IEEE Transactions on Antennas and Propagation*, 42(10):1386–1391, October 1994.

[27] S.C. Park, M.K. Park, and M.G. Kang. Super-resolution image reconstruction: a technical overview. *IEEE Signal Processing Magazine*, 20(3):21–36, 2003.

[28] 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.

[29] O. Scherzer and B. Walch. Sparsity regularization for radon measures. In *Scale Space and Variational Methods in Computer Vision*, volume 5567 of *Lecture Notes in Computer Science*, pages 452–463. Springer Berlin Heidelberg, 2009.

[30] G. Tang, B. Narayan Bhaskar, and B. Recht. Near minimax line spectral estimation. *CoRR*, abs/1303.4348, 2013.

[31] R. Tibshirani. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B. Methodological*, 58(1):267–288, 1996.

[32] P. Turán. On rational polynomials. *Acta Univ. Szeged, Sect. Sci. Math.*, pages 106–113, 1946.

[33] S. Vaiter, G. Peyré, C. Dossal, and J. Fadili. Robust sparse analysis regularization. *IEEE Transactions on Information Theory*, 59(4):2001–2016, 2013.

[34] P. Zhao and B. Yu. On model selection consistency of Lasso. *J. Mach. Learn. Res.*, 7:2541–2563, December 2006.