Exact Support Recovery for Sparse Spikes Deconvolution

Vincent Duval, Gabriel Peyré

To cite this version:

Vincent Duval, Gabriel Peyré. Exact Support Recovery for Sparse Spikes Deconvolution. 2013. hal-00839635v1

HAL Id: hal-00839635
https://hal.science/hal-00839635v1

Preprint submitted on 28 Jun 2013 (v1), last revised 13 Sep 2014 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Exact Support Recovery for Sparse Spikes Deconvolution

Vincent Duval and Gabriel Peyré
CNRS and Université Paris-Dauphine
{vincent.duval,gabriel.peyre}@ceremade.dauphine.fr

June 28, 2013

Abstract

This paper studies sparse spikes deconvolution over the space of measures. For non-degenerate sums of Diracs, we show that, when the signal-to-noise ratio is large enough, total variation regularization (which the natural extension of ℓ^1 norm of vector to the setting of measures) recovers the exact same number of Diracs. We also show that both the locations and the heights of these Diracs converge toward those of the input measure when the noise drops to zero. The exact speed of convergence is governed by a specific dual certificate, which can be computed by solving a linear system. Finally we draw connections between the performances of sparse recovery on a continuous domain and on a discretized grid.

1 Introduction

1.1 Sparse Spikes Deconvolution

Super-resolution is a central problem in imaging science, and loosely speaking corresponds to recover 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 [25, 22] 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 spike 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 technics were initially proposed in geophysics [9, 26, 21] 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 [12]. Their popularity in signal processing and statistics can be traced back to the development of the basis pursuit method [8] for approximation in redundant dictionaries and the Lasso method [29] for statistical estimation.

The theoretical analysis of the ℓ^1-regularized deconvolution was initiated by Donoho [12], see also [28]. 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 [18], and studied in the statistics community under the name of "irrepresentability condition", see [32]. A similar approach is used by Dossal and Mallat in [13] 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. [19] 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. We draw connections between our work and result on deconvolution on discrete grids in Section 6.

Let us 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 [24, 23], root finding [2], matrix pencils [11] and compressed sensing [16, 14] approaches.

Inverse problems regularization with measures. Working over a discrete grid makes the mathematical analysis difficult and leads to overly pessimistic performance guarantees. Following recent proposals [10, 3, 7], 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 [3] or root finding methods for ideal low pass filters [7].

Inverse problem regularization over the space of measures is now well understood (see for instance [27, 3]), and requires to perform variational analysis over a non-reflexive Banach space (as in [20]), 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 [7] 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 [6] that studies the robustness to noise. 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. Partial answers to this questions are given in [17] 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 to the question of the position of the recovered Diracs in the setting where the signal-to-noise ratio it 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 observation $y = \Phi m_0 \in L^2(\mathbb{T})$. 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 $y + w = \Phi m_0 + w$, with $w \in L^2(\mathbb{T})$, instead of y.

Following [7, 10], we hope to recover m_0 by solving the problem

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

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 use to recover signals or images.

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

$$\min_{m} \frac{1}{2}\|\Phi m - y\|^2 + \lambda \|m\|_{\mathrm{TV}}. \tag{$\mathcal{P}_\lambda(y)$}$$

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

Three questions immediately arise:

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

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

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

The first question is addressed in the landmark paper [7] in the case of ideal low-pass filtering: measures m_0 whose spikes are separated enough are the unique solution of $(\mathcal{P}_0(y))$ (for data $y = \Phi m_0$). Several other cases (using observations different from convolutions) are also tackled in [10], particularly in the case of non-negative measures.

The second and third questions receive partial answers in [3, 1, 17]. In [3] it is shown that if the solution of $(\mathcal{P}_0(y))$ is unique, the measures recovered by $(\mathcal{P}_\lambda(y+w))$ converge to the solution of $(\mathcal{P}_0(y))$ in the sense of the weak-* convergence when $\lambda \to 0$ and $\frac{\|w\|_2^2}{\lambda} \to 0$. In [1], error bounds are derived from the amplitudes of the reconstructed measure. In [17], bounds are given in terms of the original measure.

However, those works provide little information about the structure of the measures recovered by $(\mathcal{P}_\lambda(y+w))$: *are they made of less spikes than m_0 or, in the contrary, do they present lots of parasitic spikes?*

Contributions The present paper focuses on the problem of studying the structure of the recovered measure. We show that under mild assumptions on m_0, 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 smoothly to those of the original one. Moreover, we show that the errors in the amplitudes and locations decay linearly with respect to the noise level. To this end, we introduce the *Non Degenerate Source Condition*, a variant of the so-called source condition which involves the second derivatives of a specific dual certificate. Moreover, we show that if the original measure is known, the corresponding dual certificate can be computed numerically by solving a simple linear system, without handling the difficult constraint that it should belong to the L^∞ unit ball. We apply these results to the case of the ideal low-pass filter, and show how our results can also be used to study the recovery on a discrete grid.

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 $(\mathcal{P}_\lambda(y+w))$. 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. The particular example of the ideal low-pass filter is treated in Section 5. Lastly, Section 6 shows that this framework can also be used to study the recovery on a discrete grid.

Notations For $a \in \mathbb{R}^N$, $x \in \mathbb{T}^N$ we write $m_{a,x} = \sum_{i=1}^N a_i \delta_{x_i}$, with the implicit assumption that $a_i \neq 0$ and $x_i \neq x_j$ for all $1 \leqslant i, j \leqslant N$. Given a convolution operator Φ with kernel $t \mapsto \varphi(-t)$, we define $\Phi_x : \mathbb{R}^N \to L^2(\mathbb{T})$ (resp. Φ'_x, Φ''_x) by

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

We define

$$\Gamma_x = (\Phi_x, \Phi'_x) : (u, v) \in \mathbb{R}^N \times \mathbb{R}^N \mapsto \Phi_x u + \Phi'_x v \in \mathrm{L}^2(\mathbb{T}), \tag{1}$$
$$\Gamma'_x = (\Phi'_x, \Phi''_x) : (u, v) \in \mathbb{R}^N \times \mathbb{R}^N \mapsto \Phi'_x u + \Phi''_x v \in \mathrm{L}^2(\mathbb{T}). \tag{2}$$

We denote by $\mathrm{supp}(m)$ the support of a measure.

2 Preliminaries

In this section, we precise the framework and we state the basic results needed in the next sections. We refer to [4] for aspects regarding functional analysis and to [15] 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

$$\forall m \in \mathcal{M}(\mathbb{T}), \quad \|m\|_{\mathrm{TV}} = \sup \left\{ \int \psi \mathrm{d}m \ ; \ \psi \in C(\mathbb{T}), \|\psi\|_\infty \leqslant 1 \right\}. \tag{3}$$

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.

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

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

2.2 Subdifferential of the Total Variation

It is clear from the definition of the total variation in (3) that it is convex lower semi-continuous with respect to the weak-* topology. Its subdifferential is therefore nonempty and defined as

$$\partial\|m\|_{\mathrm{TV}} = \left\{\eta \in C(\mathbb{T})\ ;\ \forall \tilde{m} \in \mathcal{M}(\mathbb{T}), \|\tilde{m}\|_{\mathrm{TV}} \geqslant \|m\|_{\mathrm{TV}} + \int \eta\, \mathrm{d}(\tilde{m} - m)\right\}, \quad (4)$$

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 7 in Appendix) that

$$\partial\|m\|_{\mathrm{TV}} = \left\{\eta \in C(\mathbb{T})\ ;\ \|\eta\|_\infty \leqslant 1 \quad \text{and} \quad \int \eta\, \mathrm{d}m = \|m\|_{\mathrm{TV}}\right\}. \quad (5)$$

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

$$\partial\|m\|_{\mathrm{TV}} = \{\eta \in C(\mathbb{T})\ ;\ \|\eta\|_\infty \leqslant 1 \quad \text{and} \quad \forall\, i = 1, \ldots, N,\ \eta(x_i) = \mathrm{sign}(a_i)\}. \quad (6)$$

2.3 Primal and Dual problems

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

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

or the constrained problem

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

In the case where the observation is noisy (i.e. the observation y is replaced with $y + w$ for $w \in L^2(\mathbb{T})$), we attempt to reconstruct m_0 by solving $\mathcal{P}_\lambda(y+w)$.

Existence of solutions for $(\mathcal{P}_\lambda(y))$ is shown in [3], and existence of solutions for $(\mathcal{P}_0(y))$ can be checked using the direct method of the calculus of variations (recall that for $(\mathcal{P}_0(y))$, we assume that the observation is $y = \Phi m_0$).

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

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

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

Another source of information for the study of Problems ($\mathcal{P}_\lambda(y)$) and ($\mathcal{P}_0(y)$) 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 [7, 1]: the dual problem can be recast into a finite-dimensional problem.

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

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

which may be reformulated as a projection on a closed convex set (see [3, 1])

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

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

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

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

Contrary to ($\mathcal{D}_\lambda(y)$), the existence of a solution to ($\mathcal{D}_0(y)$) is not always guaranteed, so that in the following (see Definition 3) we make this assumption.

Existence is guaranteed when for instance $\operatorname{Im}\Phi^*$ is finite-dimensional (as is the case in the framework of [7]). If a solution to ($\mathcal{D}_0(y)$) exists, the unique solution of ($\mathcal{D}_\lambda(y)$) converges to a certain solution of ($\mathcal{D}_0(y)$) for $\lambda \to 0^+$ as shown in Proposition 1 below.

2.4 Dual Certificates

The strong duality between ($\mathcal{P}_\lambda(y)$) and ($\mathcal{D}_\lambda(y)$) is proved in [3, Prop. 2] by seeing ($\mathcal{D}'_\lambda(y)$) as a predual problem for ($\mathcal{P}_\lambda(y)$). As a consequence, both problems have the same value and any solution m_λ of ($\mathcal{P}_\lambda(y)$) is linked with the unique solution p_λ of ($\mathcal{D}_\lambda(y)$) by the extremality condition

$$\begin{cases} \Phi^* p_\lambda \in \partial \|m_\lambda\|_{\mathrm{TV}}, \\ -p_\lambda = \frac{1}{\lambda}(\Phi m_\lambda - y). \end{cases} \tag{7}$$

Moreover, given a pair $(m_\lambda, p_\lambda) \in \mathcal{M}(\mathbb{T}) \times L^2(\mathbb{T})$, if relations (7) hold, then m_λ is a solution to Problem ($\mathcal{P}_\lambda(y)$) and p_λ is the unique solution to Problem ($\mathcal{D}_\lambda(y)$).

As for $(\mathcal{P}_0(y))$, a proof of strong duality is given in Appendix (see Proposition 8). If a solution $p^\star$ to $(\mathcal{D}_0(y))$ exists, then it is linked to any solution $m^\star$ of $(\mathcal{P}_0(y))$ by

$$\Phi^* p^\star \in \partial \|m^\star\|_{\mathrm{TV}}, \tag{8}$$

and similarly, given a pair $(m^\star, p^\star) \in \mathcal{M}(\mathbb{T}) \times L^2(\mathbb{T})$, if relations (8) hold, then $m^\star$ is a solution to Problem $(\mathcal{P}_0(y))$ and $p^\star$ is a solution to Problem $(\mathcal{D}_0(y)))$.

Since finding $\eta = \Phi^* p^\star$ which satisfies (8) gives a quick proof that $m^\star$ is a solution of $(\mathcal{P}_0(y))$, we call η a *dual certificate* for $m^\star$. We may also use a similar terminology for $\eta_\lambda = \Phi^* p_\lambda$ and Problem $(\mathcal{P}_\lambda(y))$.

In general, dual certificates for $(\mathcal{P}_0(y))$ 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 to $(\mathcal{P}_0(y))$ is defined as $\eta_0 = \Phi^* p_0$ where $p_0 \in L^2(\mathbb{T})$ is the solution of $(\mathcal{D}_0(y))$ with minimal norm, i.e.*

$$\eta_0 = \Phi^* p_0, \quad \textit{where} \quad p_0 = \underset{p}{\operatorname{argmin}} \left\{ \|p\|_2 \; ; \; p \textit{ is a solution of } (\mathcal{D}_0(y)) \right\}. \tag{9}$$

Observe that in the above definition, p_0 is well-defined provided there exists a solution to Problem $(\mathcal{D}_0(y))$, since p_0 is then the projection of 0 on the non-empty closed convex set of solutions. Moreover, in view of the extremality conditions (8), given any solution $m^\star$ to $(\mathcal{P}_0(y))$, it may be expressed as

$$p_0 = \underset{p}{\operatorname{argmin}} \left\{ \|p\|_2 \; ; \; \Phi^* p \in \partial \|m^\star\|_{\mathrm{TV}} \right\}. \tag{10}$$

Proposition 1 (Convergence of dual certificates). *Let p_λ be the unique solution of Problem $(\mathcal{D}_\lambda(y))$, and p_0 be the solution of Problem $(\mathcal{P}_0(y))$ with minimal norm defined in (9). Then*

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

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

$$\forall k \in \{0, 1, 2\}, \quad \lim_{\lambda \to 0^+} \eta_\lambda^{(k)} = \eta_0^{(k)}, \tag{11}$$

in the sense of the uniform convergence.

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

$$\langle y, p_\lambda \rangle - \lambda \|p_\lambda\|_2^2 \geqslant \langle y, p_0 \rangle - \lambda \|p_0\|_2^2, \tag{12}$$

$$\langle y, p_0 \rangle \geqslant \langle y, p_\lambda \rangle. \tag{13}$$

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 (still denoted λ_n) such that p_{λ_n} weakly converges to some $p^\star \in L^2(\mathbb{T})$. Passing to the limit in (12), we get $\langle y, p^\star\rangle \geqslant \langle y, p_0\rangle$. 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 $(\mathcal{D}_0(y))$.

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 $\lim_{n\to+\infty} \|p_{\lambda_n}\| = \|p_0\|$. As a consequence, p_{λ_n} converges to p_0 for the $L^2(\mathbb{T})$ strong topology as well. This being true for any subsequence of any sequence $\lambda_n \to 0^+$, we get the result claimed for p_λ.

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

$$\begin{aligned} |\eta_\lambda^{(k)}(t) - \eta_0^{(k)}(t)| &= \Big| \int \varphi^{(k)}(t-x)(p_\lambda - p_0)(x)\mathrm{d}x \Big| \\ &\leqslant \sqrt{\int |\varphi^{(k)}(t-x)|^2 \mathrm{d}t} \sqrt{\int |(p_\lambda - p_0)(x)|^2 \mathrm{d}x} \\ &\leqslant C\|p_\lambda - p_0\|_2, \end{aligned}$$

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

3 Noise Robustness

3.1 Non degenerate source condition

Let us first recall the source condition introduced in [5] to derive convergence rates for the Bregman distance.

Definition 2 (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 would simply be equivalent to the optimality of m_0 for $(\mathcal{P}_0(y))$ given $y = \Phi m_0$. In the framework of Radon measures, the source condition amounts to assuming that m_0 is a solution of $(\mathcal{P}_0(y))$ *and* that there exists a solution to $(\mathcal{D}_0(y))$.

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

Proposition 2 ([10])**.** *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 $\eta \in \partial\|m_0\|_{\mathrm{TV}}$,*

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

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

In this paper, we strengthen a bit more the Source Condition so as to derive stability results with respect to the noise and the regularization parameter (see Theorem 1).

Definition 3 (Non Degenerate Source Condition)**.** *Let $m_0 = m_{x_0,a_0}$ be a discrete measure, and $\{x_{0,1}, \dots x_{0,N}\} = \operatorname{supp} m_0$. We say that m_0 satisfies the Non Degenerate Source Condition (NDSC) if*

- *there exists $\eta \in \operatorname{Im}\Phi^*$ such that $\eta \in \partial\|m_0\|_{\mathrm{TV}}$.*
- *the minimal norm certificate η_0 satisfies*

$$\begin{aligned} &\forall\, s \in \mathbb{T} \setminus \{x_{0,1}, \dots x_{0,N}\}, && |\eta_0(s)| < 1, \\ &\forall\, i \in \{1, \dots N\}, && \eta_0''(x_{0,i}) \neq 0. \end{aligned}$$

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

The first assumption in Definition 3 is the standard Source Condition. As explained above, it implies the existence of a solution to $(\mathcal{D}_0(y))$, so that it makes sense to consider the minimal norm certificate in the second assumption.

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 5.

3.2 Main Result

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

Theorem 1 (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 (1)) has full rank and that m_0 satisfies the Non Degenerate Source Condition (see Definition 3). Then there exists $\alpha > 0, \lambda_0 > 0$, such that for $(\lambda, w) \in D_{\alpha,\lambda_0}$ where*

$$D_{\alpha,\lambda_0} = \left\{(\lambda, w) \in \mathbb{R}_+ \times L^2(\mathbb{T})\ ;\ 0 \leqslant \lambda \leqslant \lambda_0 \quad \text{and} \quad \|w\|_2 \leqslant \alpha\lambda \right\}, \tag{14}$$

the solution $\tilde{m}$ of $\mathcal{P}_\lambda(y + w)$ is unique and is composed of exactly N spikes, $\tilde{m} = \sum_{i=1}^N \tilde{a}_{\lambda,i}\delta_{\tilde{x}_{\lambda,i}}$ with $\tilde{a}_{\lambda,i} \neq 0$ and $\operatorname{sign}(\tilde{a}_{\lambda,i}) = \operatorname{sign}(a_{0,i})$ (for $1 \leqslant i \leqslant N$).

Moreover, 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

$$\forall i \in \{1, \dots N\}, \quad |\tilde{x}_{\lambda,i} - x_{0,i}| = O(\|w\|_2) \quad \text{and} \quad |\tilde{a}_{\lambda,i} - a_{0,i}| = O(\|w\|_2). \tag{15}$$

Proof. We split the proof in several steps.

Behavior of the minimal norm certificate. Let $i \in \{1, \dots N\}$. Since m satisfies the non degenerate source condition, $\eta_0''(x_i) \neq 0$. In fact, $\eta_0''(x_{0,i}) > 0$ (resp. < 0) whenever $\eta_0(x_{0,i}) = -1$ (resp. $+1$). Moreover, by continuity, there exists $\varepsilon_i > 0$, $C_i > 0$ such that for all $t \in (x_{0,i} - \varepsilon_i, x_{0,i} + \varepsilon_i)$, one has

- $|\eta_0(t)| \geqslant C_i > 0$,
- $|\eta_0''(t)| \geqslant C_i > 0$.

The number of points $x_{0,1}, \dots x_{0,N}$ being finite, we may choose $\varepsilon_i = \varepsilon$ and $C_i = C$ independent of i and such that the sets $(x_{0,i} - \varepsilon, x_{0,i} + \varepsilon)$ for $1 \leqslant i \leqslant N$ do not intersect. The set $K_\varepsilon = \mathbb{T} \setminus \bigcup_{i=1}^N (x_{0,i} - \varepsilon, x_{0,i} + \varepsilon)$ being compact, we observe that $\sup_{K_\varepsilon} |\eta_0| < 1$.

As a consequence, there exists $r > 0$ such that for any function $g \in C^2(\mathbb{T})$ satisfying $\|\eta_0^{(j)} - g^{(j)}\|_\infty \leqslant r$ for all $j \in \{0, 1, 2\}$

$$\begin{cases} |g(t)| & \geqslant \frac{C}{2} > 0 \text{ for } t \in (x_{0,i} - \varepsilon, x_{0,i} + \varepsilon), \\ |g''(t)| & \geqslant \frac{C}{2} > 0 \text{ for } t \in (x_{0,i} - \varepsilon, x_{0,i} + \varepsilon), \\ \sup_{K_\varepsilon} |g| & < 1. \end{cases} \tag{16}$$

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

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

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 for $j \in \{0, 1, 2\}$

$$\|\eta_\lambda^{(j)} - \tilde{\eta}_\lambda^{(j)}\|_2 \leqslant M \frac{\|w\|_2}{\lambda} \tag{17}$$

for some constant $M > 0$ which depends only on φ and its derivatives.

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

$$\begin{aligned} \|\eta_0^{(j)} - \tilde{\eta}_\lambda^{(j)}\|_\infty &\leqslant \|\eta_0^{(j)} - \eta_\lambda^{(j)}\|_\infty + \|\eta_\lambda^{(j)} - \tilde{\eta}_\lambda^{(j)}\|_\infty, \\ &\leqslant \|\eta_0^{(j)} - \eta_\lambda^{(j)}\|_\infty + \frac{r}{2}, \end{aligned}$$

we see that for λ small enough $\tilde{\eta}_\lambda$ satisfies (16).

Structure of the reconstructed measure. By (16) for $g = \tilde{\eta}_\lambda$ and using the extremality conditions we obtain that $|\tilde{m}_\lambda|(K_\varepsilon) = 0$ and that $\tilde{m}_\lambda$ has at most one spike in each interval $(x_{0,i} - \varepsilon, x_{0,i} + \varepsilon)$. Indeed, the extremality conditions impose that $\operatorname{supp}(\tilde{m}_\lambda)$ is included in the set of points t such that $\tilde{\eta}_\lambda(t) = \pm 1$. But since η_λ is strictly convex in $(x_{0,i} - \varepsilon, x_{0,i} + \varepsilon)$ (or strictly concave, depending on the sign of $\eta_0(x_{0,i})$), there is at most one point $\tilde{x}_{\lambda,i} \in (x_{0,i} - \varepsilon, x_{0,i} + \varepsilon)$ such that $\{\tilde{x}_{\lambda,i}\} \subset \operatorname{supp}(\tilde{m}_\lambda) \cap (x_{0,i} - \varepsilon, x_{0,i} + \varepsilon)$.

It remains to prove that there is indeed one spike in each interval $(x_{0,i} - \varepsilon, x_{0,i}+\varepsilon)$. This is obtained by relying on a result by Bredies and Pikkarainen [3] which is an application of [20, Th. 3.5] and guarantees that $\tilde{m}_\lambda$ converges to m for the weak-* topology when $\lambda, \|w\|_2 \to 0$. We sketch the proof in Appendix (see Lemma 1) for the convenience of the reader.

By weak-* convergence of $\tilde{m}_\lambda$ to m for $\lambda \to 0^+$ and $\|w\|_2 \to 0$, there are necessarily N spikes located at some $\tilde{x}_{\lambda,i} \in (x_{0,i} - \varepsilon, x_{0,i} + \varepsilon)$ which converges to $x_{0,i}$, and their associated mass $\tilde{a}_{\lambda,i}$ converges to a_i.

Smoothness of the locations and amplitudes. To derive smoothness results for $(\tilde{a}_\lambda, \tilde{x}_\lambda)$ as a function of (λ, w), we observe that for λ small enough and $\|w\| \leqslant \alpha\lambda$, it satisfies the following implicit equation

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

where $s_0 = \mathrm{sign}(a_0)$, and

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

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

$$\begin{aligned} \frac{\partial E}{\partial a}(a,x,\lambda,w) &= \Gamma_x \Phi_x \\ \frac{\partial E}{\partial x}(a,x,\lambda,w)[\delta] &= \mathrm{diag}(\delta)\Gamma'_x(\Phi_x a - y) + \Gamma_x {\Phi'_x}^* \mathrm{diag}(a)\delta. \end{aligned}$$

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

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

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

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

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 $(\lambda, w) \in W$, $\text{sign}(\hat{a}_{\lambda,w}) = s_0$,
- 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 $(\mathcal{P}_\lambda(y))$ for all $(\lambda, w) \in W$ such that $\|w\|_2 \leqslant \alpha\lambda$, hence the result. □

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

3.3 Extensions

Theorem 1 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 3, 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 $\varphi \in C^2(\mathbb{T}^d \times \mathbb{T}^d)$.

4 Vanishing Derivatives Pre-certificate

As a by-product of the proof of Theorem 1, we show in this section that 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. As a consequence, one may compute he minimal norm certificate simply by solving a linear system, without handling the cumbersome constraint $\|\eta_0\|_\infty \leqslant 1$.

4.1 Dual Pre-certificates

We begin by introducing a "good candidate" for a dual certificate.

Definition 4 (Vanishing derivative pre-certificate)**.** *The vanishing derivative pre-certificate associated to a measure $m_0 = m_{x_0,a_0}$ is $\bar{\eta}_0 = \Phi^* q_0$ where*

$$q_0 = \underset{q\in L^2(\mathbb{T})}{\text{argmin}} \ \|q\| \quad \textit{subj. to} \quad \forall\, 1 \leqslant i \leqslant N, \quad \begin{cases} (\Phi^* q)(x_{0,i}) = \text{sign}(a_{0,i}), \\ (\Phi^* q)'(x_{0,i}) = 0. \end{cases} \tag{18}$$

It is clear that if m_0 is a solution to $(\mathcal{P}_0(y))$ (for $y = \Phi m_0$), and if there exists a solution to $(\mathcal{D}_0(y))$, then the minimal norm certificate η_0 must satisfy all the constraints defining $\bar{\eta}_0$. Conversely, if $\|\bar{\eta}_0\|_\infty \leqslant 1$, the conditions imposed on $\bar{\eta}_0$ imply that m_0 is a solution to $(\mathcal{P}_0(y))$ and that $\bar{\eta}_0$ is a certificate for m_0. As a consequence, for an identifiable measure m_0, $\eta_0 = \bar{\eta}_0$ if and only if $\|\bar{\eta}_0\|_\infty \leqslant 1$.

The following proposition details the computation of $\bar{\eta}_0$.

Proposition 3. *We assume Γ_{x_0} has full rank, i.e. $\Gamma_{x_0}^*\Gamma_{x_0} \in \mathbb{R}^{2N\times 2N}$ is invertible, and that Problem (18) is feasible. Then $\bar{\eta}_0$ is uniquely defined and*

$$\bar{\eta}_0 = \Phi^*\Gamma_{x_0}^{+,*}\begin{pmatrix}\operatorname{sign}(a_0)\\ 0\end{pmatrix} \quad \textit{where} \quad \Gamma_{x_0}^{+,*} = \Gamma_{x_0}(\Gamma_{x_0}^*\Gamma_{x_0})^{-1}.$$

Proof. The problem (18) can be written as

$$\bar{\eta}_0 = \underset{\eta=\Phi^* q}{\operatorname{argmin}}\ \|q\|. \quad \text{subj. to} \quad \left\{\begin{array}{l}\Phi_{x_0}^* q = \operatorname{sign}(a_0),\\ \Phi_{x_0}'^* q = 0,\end{array}\right.$$

which is a finite-dimensional quadratic optimization problem with 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 $q = \bar{q}_0$

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

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

4.2 The vanishing Derivative Pre-certificate is a Certificate

A priori, the vanishing derivative pre-certificate $\bar{\eta}_0$ introduced above is not a certificate for m_0 since we do not impose $\|\bar{\eta}_0\|_\infty \leqslant 1$. Yet, the fact that the derivative vanishes on the support of m_0 and the non degenerate source condition imply that $\bar{\eta}_0$ is indeed a certificate.

Proposition 4. *Under the hypothesis of Theorem 1, the vanishing derivative pre-certificate $\bar{\eta}_0$ is equal to the minimal norm certificate η_0, and one has*

$$\eta_0 = \Phi^*\Gamma_{x_0}^{+,*}\begin{pmatrix}\operatorname{sign}(a_0)\\ 0\end{pmatrix} \quad \textit{where} \quad \Gamma_{x_0}^{+,*} = \Gamma_{x_0}(\Gamma_{x_0}^*\Gamma_{x_0})^{-1}. \tag{19}$$

Proof. We consider the case where $w = 0$, we introduce the C^1 path $\lambda \mapsto (\hat{a}_\lambda, \hat{x}_\lambda)$ constructed in the proof of Theorem 1 (in the noiseless case) which coincides for λ small enough with the amplitudes and locations of the solution $m_\lambda = m_{a_\lambda, x_\lambda}$ of $(\mathcal{P}_\lambda(y))$.

Writing
$$\hat{a}_0' = \frac{\partial \hat{a}_\lambda}{\partial \lambda}(0) \in \mathbb{R}^N \quad \text{and} \quad \hat{x}_0' = \frac{\partial \hat{x}_\lambda}{\partial \lambda}(0) \in \mathbb{R}^N,$$
we observe that for any $i \in \{1, \dots N\}$ and any $x \in \Omega$,
$$\begin{aligned} &\frac{\hat{a}_{\lambda,i}\varphi(\hat{x}_{\lambda,i} - x) - \hat{a}_{0,i}\varphi(\hat{x}_{0,i} - x)}{\lambda} - \left[\hat{a}_{0,i}\varphi'(\hat{x}_{0,i} - x)\hat{x}_{0,i}' + \hat{a}_{0,i}'\varphi(\hat{x}_{0,i} - x)\right] \\ &\quad = \int_0^1 \left[\hat{a}_{\lambda t,i}\varphi'(\hat{x}_{\lambda t,i} - x)\hat{x}_{\lambda t,i}' + \hat{a}_{\lambda t,i}'\varphi(\hat{x}_{\lambda t,i} - x)\right] \\ &\qquad\quad - \left[\hat{a}_{0,i}\varphi'(\hat{x}_{0,i} - x)\hat{x}_{0,i}' + \hat{a}_{0,i}'\varphi(\hat{x}_{0,i} - x)\right] \mathrm{d}t, \end{aligned}$$
and the latter integral converges (uniformly in x) to zero when $\lambda \to 0^+$ by uniform continuity of its integrand (since $\hat{a}$, $\hat{x}$ and φ are C^1). As a consequence, we obtain that $\frac{y - \Phi_{\hat{x}_\lambda}}{\lambda}$ converges uniformly to $-\Gamma_{x_0}\begin{pmatrix}\mathrm{Id} & 0 \\ 0 & \mathrm{diag}(a_0)\end{pmatrix}\begin{pmatrix}\hat{a}_0' \\ \hat{x}_0'\end{pmatrix}$.

On the other hand, the implicit functions theorem yields
$$\begin{aligned} \begin{pmatrix}\hat{a}_0' \\ \hat{x}_0'\end{pmatrix} &= \frac{\partial(\hat{a}_\lambda, \hat{x}_\lambda)^T}{\partial \lambda}(0) \\ &= -\left(\frac{\partial E_{s_0}}{\partial(a,x)}(a_0, x_0, 0)\right)^{-1} \frac{\partial E_{s_0}}{\partial \lambda}(a_0, x_0, 0) \\ &= -\begin{pmatrix}\mathrm{Id} & 0 \\ 0 & \mathrm{diag}(a_0)^{-1}\end{pmatrix} (\Gamma_{x_0}^* \Gamma_{x_0})^{-1} \begin{pmatrix} s_0 \\ 0 \end{pmatrix}. \end{aligned}$$
As a consequence, $\frac{y - \Phi_{\hat{x}_\lambda}}{\lambda}$ converges uniformly to $\Gamma_{x_0}(\Gamma_{x_0}^*\Gamma_{x_0})^{-1}\begin{pmatrix}\mathrm{sign}(a_0) \\ 0\end{pmatrix}$ and $\Phi^*\left(\frac{y - \Phi_{\hat{x}_\lambda}}{\lambda}\right)$ converges uniformly to $\bar{\eta}_0$.

On the other hand, by Proposition 1, we know that $\eta_\lambda = \Phi^*\left(\frac{y - \Phi_{x_\lambda}}{\lambda}\right)$ converges uniformly to the minimal norm certificate η_0. We conclude that $\bar{\eta}_0 = \eta_0$. $\square$

5 Application to the Ideal Low-pass Filter

In this section, we apply the results of the previous sections to the particular case of the Dirichlet kernel, defined as
$$\varphi(t) = \sum_{k=-f_c}^{f_c} e^{2i\pi kt} = \frac{\sin\left((2f_c + 1)\pi t\right)}{\sin(\pi t)}. \tag{20}$$

5.1 Elementary Results

We first check that the assumptions made in Section 3 hold in the case of the ideal low-pass filter.

Proposition 5 (Existence of p_0)**.** *Let $m \in \mathcal{M}(\mathbb{T})$ and $y = \Phi m \in L^2(\mathbb{T})$. There exists a solution of $(\mathcal{D}_0(y))$. As a consequence, $p_0 \in L^2(\mathbb{T})$ is well defined.*

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

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

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 $\eta^* \in C(\mathbb{T})$. But $\|\eta^*\|_\infty \leqslant 1$ and $\eta^* \in \operatorname{Im} \Phi^*$, so that $\eta^* = \Phi^* p^*$ for some p^* solution of $(\mathcal{D}_0(y))$. □

Proposition 6 (Injectivity of Γ_x)**.** *Let $x = (x_1, \dots x_N) \in \mathbb{T}^N$ 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.*

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

$$\begin{aligned}\forall t \in \mathbb{T}, \quad 0 &= \sum_{j=1}^{N} (u_j \varphi(t - x_j) + v_j \varphi'(t - x_j)) \\ &= \sum_{k=-f_c}^{f_c} \left(\sum_{j=1}^{N} (u_j + 2ik\pi v_j) e^{-2ik\pi x_j} \right) e^{2ik\pi t}\end{aligned}$$

We deduce that

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

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

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

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

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

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:

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

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. □

5.2 Identifiable Measures

In [7], Candès and Fernandez-Granda have proved that discrete measures are identifiable provided that their support is separated enough, i.e. $\Delta(m) \geqslant \frac{C}{f_c}$ for some $C > 0$, where $\Delta(m)$ is the so-called minimum separation distance.

Definition 5 (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 [7] 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 show that for real measures, necessarily $C \geqslant \frac{1}{2}$.

We rely on the following theorem, proved by P. Turán [30].

Theorem 2 (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, $|\arg(z_0)| \geqslant \frac{2\pi}{n}$. Moreover, if $|\arg(z_0)| = \frac{2\pi}{n}$, then $P(z) = c(1+z)^n$ for some $c \in \mathbb{C}^*$.*

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

Corollary 1 (Non identifiable measures). *Let $\theta \in (0, \frac{1}{2}]$. Then there exists a discrete measure m with $\Delta(m) = \frac{\theta}{f_c}$ such that m is not a solution of $(\mathcal{P}_0(y))$ for $y = \Phi m$.*

Proof. Let $m = \delta_{-\theta/f_c} + \delta_0 - \delta_{\theta/f_c}$. By contradiction, assume that m is a solution to $(\mathcal{P}_0(y))$, and let $\eta \in C(\mathbb{T})$ be an associated solution of $(\mathcal{D}_0(y))$ (which exists since $\operatorname{Im}\Phi^*$ is finite-dimensional). Then necessarily $\eta(0) = 1$ and $\eta(\theta/f_c) = -1$. The real trigonometric polynomial $D(t) = \frac{1+\eta(t)}{2}$ is non-negative, $D(0) = 1 = \|D\|_\infty$ and $D(\theta/f_c) = 0$. If $D(t) = \sum_{k=-f_c}^{f_c} d_k e^{2i\pi kt}$, the polynomial $P(z) = \sum_{k=0}^{2f_c} d_{k-f_c} z^k$ satisfies $P(1) = 1 = \sup_{|z|=1} |P(z)|$, and $P(e^{2i\pi\frac{\theta}{f_c}}) = 0$.

If $\theta < \frac{1}{2}$ this contradicts Theorem 2. If $\theta = \frac{1}{2}$, then Theorem 2 asserts that $P(z) = c(1+z)^{2f_c}$, and from $P(1) = 1$ we get $c = \frac{1}{2^{2f_c}}$, so that

$$D(t) = e^{-2i\pi f_c} P(e^{2i\pi t}) = (\cos(\pi f_c t))^{2f_c}.$$

Then $D(-\frac{\theta}{f_c}) = 0$, which contradicts the optimality of η. As a conclusion, m is not a solution of $(\mathcal{P}_0(y))$. □

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

Corollary 2. *Let $\tilde{m}_\lambda = m_{\tilde{a}_\lambda, \tilde{x}_\lambda}$ be a discrete solution of Problem $\mathcal{P}_\lambda(y+w)$ where $y = \Phi m$ for any data $m \in \mathcal{M}(\mathbb{T})$ and any noise $w \in L^2(\mathbb{T})$. Let $(\tilde{x}_{\lambda,i}, \tilde{x}_{\lambda,j}) \in \operatorname{supp}(m_\lambda)^2$. If $|\tilde{x}_{\lambda,i} - \tilde{x}_{\lambda,j}| \leqslant \frac{0.5}{f_c}$ then $\operatorname{sign}(\tilde{a}_{\lambda,j}) = \operatorname{sign}(\tilde{a}_{\lambda,i})$.*

5.3 Performance of Pre-certificates

In order to prove their identifiability result for measures, the authors of [7] introduce a "good candidate" for a dual certificate associate to $m = m_{a,x}$ 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

$$\hat{\eta}_0(t) = \sum_{i=1}^{N} \left(\alpha_i K(t - x_i) + \beta_i K'(t - x_i)\right) \text{ with } K(t) = \left(\frac{\sin\left(\left(\frac{f_c}{2} + 1\right)\pi t\right)}{\left(\frac{f_c}{2} + 1\right)\sin \pi t} \right)^4$$

and compute $(\alpha_i, \beta_i)_{i=1}^N$ by imposing that $\hat{\eta}_0(x_i) = \operatorname{sign}(a_i)$ and $\hat{\eta}_0'(x_i) = 0$.

They show that the constructed "pre-certificate" is indeed a certificate, i.e. that $\|\hat{\eta}\|_\infty \leqslant 1$, provided that the support is separated enough (i.e. when $\Delta(m) \geqslant C/f_c$). This results is important since it proves that measures that have sufficiently separated spikes are identifiable. Furthermore, using the fact that $\hat{\eta}_0$ is not degenerate (i.e. $\hat{\eta}_0(x_i)'' \neq 0$ for all $i = 1, \ldots, N$), the same authors derive an L^2 robustness to noise result in [6], and Fernandez-Granda and Azais et al. use the constructed certificate to analyze finely the positions of the spikes in [17, 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 random real-valued measures for different separation distances and observing when each pre-certificate satisfies $\|\eta\|_\infty \leqslant 1$.

As predicted by the result of [7], we observe numerically that the pre-certificate $\hat{\eta}_0$ is a a certificate (i.e. $\|\hat{\eta}_0\|_\infty \leqslant 1$) for any measure with $\Delta(m) \geqslant 1.87/f_c$. We also observe that this continues to hold up to $\Delta(m) \geqslant 1/f_c$. Yet, below $1/f_c$, we observe numerically that some measures are still identifiable (as asserted using the vanishing derivative pre-certificate $\bar{\eta}_0$) but $\hat{\eta}_0$ stops being a certificate, i.e. $\|\hat{\eta}_0\|_\infty > 1$. An illustration is given in Figure 1, where the chosen parameters are $f_c = 26$ and $N = 7$. For the cases $\Delta(m) = 2.50/f_c$ and $\Delta(m) = 1.26/f_c$, both pre-certificates $\bar{\eta}_0$ and $\hat{\eta}_0$ are certificates, showing that the generated measure is identifiable. Notice how the

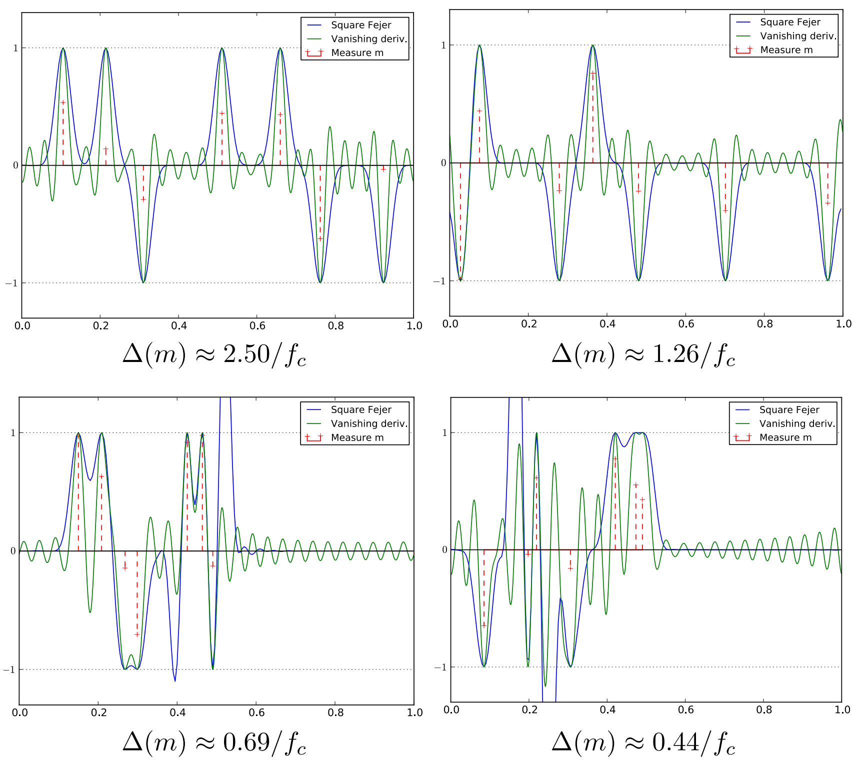

Figure 1: The blue curve is the square Fejer kernel pre-certificate $\hat{\eta}_0$ (introduced in [7]) and the green curve shows the vanishing derivative pre-certificate $\bar{\eta}_0$, defined in (18), for several measures m (whose Diracs' locations and elevation are display in dashed red) with different separation distances $\Delta(m)$. Eventually when $\Delta(m)$ is small enough, both pre-certificates break (i.e. are not anymore certificates), but the square Fejer always breaks before the vanishing derivative pre-certificate.

vanishing derivative certificate $\bar{\eta}_0$ oscillates much more than the square Fejer certificate $\hat{\eta}_0$. For $\Delta(m) = 0.69/f_c$, the square Fejer pre-certificate breaks the constraint ($\|\eta\|_\infty \approx 2.39$) whereas the vanishing derivative certificate still satisfies $\|\eta\|_\infty \leqslant 1$. Eventually, for $\Delta(m) = 0.44/f_c$, both pre-certificates violate the constraint, with $\|\hat{\eta}_0\|_\infty \approx 3.39$ and $\|\bar{\eta}_0\|_\infty = 1.17$ respectively.

In all the experiments that we have led, the vanishing derivative pre-certificate behaved at least as well as the square Fejer. We are not able to prove rigorously this observation, but several facts advocate for this:

- Whenever the square Fejer pre-certificate works, m is a solution of $(\mathcal{P}_0(y))$, and the minimal norm certificate η_0 is then associated to m.
- Provided that $|\eta_0| < 1$ on $\mathbb{T} \setminus \mathrm{supp}(m)$ and that $\eta_0''(x_i) \neq 0$ for all $1 \leqslant i \leqslant N$, Proposition 4 implies that the vanishing derivative pre-certificate $\bar{\eta}_0$ is equal to η_0.

- It the minimum separation condition holds, the authors of [7] have proved that $|\hat{\eta}_0| < 1$ on $\mathbb{T} \setminus \text{supp}(m)$ and $\hat{\eta}_0''(x_i) \neq 0$.

Intuitively it seems very unlikely that the minimal norm certificate should fail to satisfy $|\eta_0| < 1$ on $\mathbb{T} \setminus \text{supp}(m)$ and $\eta_0''(x_i) \neq 0$ whereas the square Fejer certificate does not. Indeed the failure of any of those conditions tend to impose a large L^2 norm on the considered pre-certificate η (recall that when φ is an ideal low pass filter $\|\eta\|_2 = \|p\|_2$).

6 Deconvolution Over a Discrete Grid

6.1 Finite Dimensional ℓ^1 Regularization

A popular way to compute approximate solutions to $(\mathcal{P}_\lambda(y))$ with fast algorithms is to solve this problem on a finite discrete grid. This corresponds to imposing that the optimized measures m are of the form $m = m_{a,\sigma}$, where $\sigma = (\sigma_1, \ldots, \sigma_P) \in \mathbb{T}^P$ is a fixed computational grid. The optimal measures $m_{a,\sigma}$ solving the corresponding problem are obtained by solving the following finite dimensional convex program on a

$$\min_{a \in \mathbb{R}^P} \frac{1}{2}\|y - \Psi a\|^2 + \lambda \|a\|_1 \quad \text{where} \quad \|a\|_1 = \sum_{i=1}^{P} |a_i|. \tag{$\bar{\mathcal{P}}_\lambda(y)$}$$

where $\Psi = \Phi_\sigma : \mathbb{R}^P \to L^2(\mathbb{T})$. The problem $(\bar{\mathcal{P}}_\lambda(y))$ is the so-called basis pursuit denoising problem [8], also known as the Lasso [29] in statistics. In the noiseless setting, one considers

$$\min_{\Psi a = y} \|a\|_1. \tag{$\bar{\mathcal{P}}_0(y)$}$$

We have already remarked in Section 3.3 that Theorem 1 extends to non-stationary smooth operators (i.e. not necessarily convolutions). Similarly, the results exposed in this section hold for any operator $\Psi : \mathbb{R}^P \to \mathcal{H}$ where $\mathcal{H} \subset \mathrm{L}^2(\mathbb{T})$ (i.e. not necessarily of the form $\Psi = \Phi_\sigma$). For instance, if $\mathcal{H} = \mathbb{R}^Q$, then Ψ is a finite dimensional matrix, i.e. $\Psi \in \mathbb{R}^{P \times Q}$. Without loss of generality, we assume that $\mathcal{H} = \mathbb{R}^Q$ with $Q \geqslant P$.

6.2 Certificates over a Discrete Grid

Similarly to its infinite dimensional counterpart, for any observation $y \in \mathcal{H}$, we can associate a minimal norm certificate $\eta_0 \in \mathbb{R}^P$, which reads

$$\eta_0 = \underset{\eta = \Psi^* p, p \in \mathcal{H}}{\text{argmin}} \|p\| \quad \text{subject to} \quad \eta \in \partial \|a^\star\|_1 \tag{21}$$

where $a^\star$ is any solution of the constrained problem $(\bar{\mathcal{P}}_0(y))$. Note that in this setting, the source condition is always satisfied, so that this η_0 is always well defined.

The sub-differential of the ℓ^1 norm reads

$$\partial\|a\|_1 = \left\{\eta \in \mathbb{R}^P \ ; \ \|\eta\|_\infty \leqslant 1 \quad \text{and} \quad \eta_I = \operatorname{sign}(a_I)\right\}$$

where $I = \operatorname{supp}(a) = \{i \in \{1, \ldots, N\} \ ; \ a_i \neq 0\}$ and $a_I = (a_i)_{i \in I}$.

Following [18], we introduce the following pre-certificate by dropping the constraint $\|\eta\|_\infty \leqslant 1$ in (21)

$$\hat{\eta}_0 = \underset{\eta = \Psi^* p, p \in \mathcal{H}}{\operatorname{argmin}} \ \|p\| \quad \text{subject to} \quad \eta_I = \operatorname{sign}(a_I^\star), \tag{22}$$

where $I = \operatorname{supp}(a^\star)$ for any solution $a^\star$ of the constrained problem $(\bar{\mathcal{P}}_0(y))$. We denote by Ψ_I the operator obtained by considering only the columns of Ψ indexed by I. If Ψ_I has full rank, $\hat{\eta}_0$ is uniquely defined and can be computed by solving a linear system

$$\hat{\eta}_0 = \Psi^* \Psi_I^{+,*} \operatorname{sign}(a_I^\star) \quad \text{where} \quad \Psi_I^{+,*} = \Psi_I (\Psi_I^* \Psi_I)^{-1}.$$

Like the vanishing derivative precertificate for the continuous framework, if $\hat{\eta}_0$ satisfies $\|\hat{\eta}_0\|_\infty \leqslant 1$ then it is equal to η_0.

6.3 Noise Robustness

The following theorem is a finite dimensional counter-part to Theorem 1.

Theorem 3 (Noise robustness, discrete case)**.** *Let $a_0 \in \mathbb{R}^N$ be a solution of $\bar{\mathcal{P}}_0(y)$ for $y = \Psi a_0$. We assume that Ψ_J has full rank, where*

$$J = \{i \in \{1, \ldots, N\} \ ; \ |\eta_{0,i}| = 1\}$$

and η_0 is the minimal norm certificate defined in (21)*.*

Then, there exists $\alpha > 0, \lambda_0 > 0$, such that for $(\lambda, w) \in D_{\alpha,\lambda_0}$ (defined in (14)*) the solution $\tilde{a}$ of $\bar{\mathcal{P}}_\lambda(y + w)$ is unique and satisfies* $\operatorname{supp}(\tilde{a}) \subset J$*.*

Moreover, choosing $\lambda = \|w\|_2/\alpha$ ensures $\|\tilde{a} - a_0\| = O(\|w\|)$.

Proof. The dual problem to $(\bar{\mathcal{P}}_\lambda(y))$ is :

$$\min_{\|\Psi^* p\|_\infty \leqslant 1} \left\|\frac{y}{\lambda} - p\right\|_2^2, \tag{$\bar{\mathcal{D}}_\lambda'(y)$}$$

and its solutions converge to $p_0 \in \mathcal{H}$ for $\lambda \to 0^+$, where $\Psi^* p_0 = \eta_0$ is the minimal norm certificate.

By the triangle inequality:

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

Thus, there exist two constants $\alpha > 0$ and $\lambda_0 > 0$, such that for $\frac{\|w\|_2}{\lambda} \leqslant \alpha$ and $0 < \lambda < \lambda_0$, $|\tilde{\eta}_{\lambda,i}| < 1$ for any $i \notin J$. Then, the primal-dual extremality conditions imply that for any solution $\tilde{a}_\lambda$ of $\bar{\mathcal{P}}_\lambda(y + w)$, one has $\operatorname{supp}(\tilde{a}_\lambda) \subset J$.

Such a solution satisfies

$$\frac{1}{\lambda}\Psi_J^* (y + w - \Psi_J \tilde{a}_{\lambda,J}) = \tilde{\eta}_{\lambda,J},$$

and since Ψ_J has full rank and $y = \Psi_J a_{0,J}$, one obtains

$$\tilde{a}_{\lambda,J} = a_{0,J} + \Psi_J^+ w - \lambda(\Psi_J \Psi_J^*)^{-1} \tilde{\eta}_{\lambda,J}.$$

But the certificate satisfies $\|\tilde{\eta}_{\lambda,J}\|_\infty \leqslant 1$, hence, choosing $\lambda = \frac{\|w\|_2}{\alpha}$ we get $\|\tilde{a} - a_0\| = O(\|w\|)$.

□

Note that although the hypotheses are similar to the one of Theorem 1, the conclusion is different, in the sense that this theorem does not assert that the support J of the recovered vector matches the support I of the input vector a_0. In fact, we have $I \subset J$, so that the recovered solutions of $\bar{\mathcal{P}}_\lambda(y + w)$ have in general more spikes than a_0, and the spikes in $J \setminus I$ vanish as $\lambda \to 0, \|w\|_2 \to 0$.

By assuming that $J = I$, we get the exact recovery of the support for small noise. In particular, we obtain the following theorem, which was initially proved by Fuchs [18], and which gives a more precise statement at the price of a more stringent condition on the input signal a_0.

Corollary 3 (Exact support recovery, discrete case,[18])**.** *Let $a_0 \in \mathbb{R}^N$ and $I = \text{supp}(a_0)$, assuming Ψ_I has full rank. If $\|\hat{\eta}_{0,I^c}\|_\infty < 1$ where $\hat{\eta}_0$ is defined in* (22)*, then there exists $\alpha > 0, \lambda_0 > 0$, such that for $(\lambda, w) \in D_{\alpha,\lambda_0}$ (defined in* (14)*) the solution $\tilde{a}$ of $\bar{\mathcal{P}}_\lambda(y + w)$ is unique and satisfies* $\text{supp}(\tilde{a}) = I$ *and reads*

$$\tilde{a}_I = a_{0,I} + \Psi_I^+ w - \lambda(\Psi_I \Psi_I^*)^{-1} \operatorname{sign}(a_{0,I}). \tag{23}$$

The condition $\|\hat{\eta}_{0,I^c}\|_\infty < 1$ is often called the irrepresentability condition in the statistics literature, see [32]. This condition can be shown to be almost a necessary and sufficient condition to ensure exact recovery of the support I. For instance, if $\|\hat{\eta}_{0,I^c}\|_\infty > 1$, one can show that $\text{supp}(\tilde{a}) \neq I$ where $\tilde{a}$ is any solution of $\bar{\mathcal{P}}_\lambda(y)$ for all $\lambda > 0$, see [31]. In our framework, we see that the assumption $\|\hat{\eta}_{0,I^c}\|_\infty < 1$ means that the precertificate $\hat{\eta}_0$ 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 a_0. The result [20, Th. 3.5] ensures that $\text{sign}(\tilde{a}_{\lambda,I}) = \text{sign}(a_{0,I})$ for λ small enough, hence Equation (23).

For deconvolution problems, an important issue is that Corollary 3 becomes useless to study the stability of the original infinite dimensional problem $(\mathcal{P}_\lambda(y))$. Even if computed on a continuous grid, the pre-certificate (22) is not constrained to have vanishing derivatives. So for a generic input measure $m_0 = m_{a_0,\sigma}$, $\|\hat{\eta}_0\|_\infty$ necessarily becomes larger than 1 for N large enough. 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 $\bar{\eta}_0$ defined in (18), and not the pre-certificate $\hat{\eta}_0$.

Conclusion

In this paper, we have given for the first time a precise statement about the structure of the measures recovered using sparse deconvolution. We have shown that for non degenerate measures, one recovers the same number of spikes and that these spikes converge to the original ones when λ and $\|w\|/\lambda$ are small enough. Moreover, we have pointed the importance of a specific "minimal norm" certificate in the asymptotic behavior of the sparse recovery. We have provided a closed form solution that enables its computation by solving a linear system. We have shown how a similar analysis applies to ℓ^1 minimization over a finite grid, which shed some lights on the connexions between the finite and the infinite dimensional recovery problems. Finally, let us note that the proposed method extends to non-stationary filtering operators and to arbitrary dimensions.

Acknowledgements

The authors would like to thank Jalal Fadili, Charles Dossal and Samuel Vaiter for fruitful discussions. This work has been supported by the European Research Council (ERC project SIGMA-Vision).

Appendix

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

Proposition 7 (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:*

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

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 $A \subset B_\infty(0,1)$, where $B_\infty(0,1)$ is the $L^\infty(\mathbb{T})$ closed unit ball. Conversely, we observe that $B_\infty(0,1) \subset A$ 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:

$$\begin{aligned} J^*(\eta) &= \sup_{m \in \mathcal{M}(\mathbb{T})} \sup_{t>0} \left(\langle \eta, \, tm\rangle - J(tm)\right) \\ &= \sup_{t>0} t \left(\sup_{m \in \mathcal{M}(\mathbb{T})} \langle \eta, \, m\rangle - J(m) \right) \\ &= \begin{cases} 0 & \text{if } \eta \in A, \\ +\infty & \text{otherwise.} \end{cases} \end{aligned}$$

By Proposition I.5.1 in [15], 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 $\int \eta \mathrm{d}m = \|m\|_{\mathrm{TV}}$. □

Proposition 8. *There exists a solution to $(\mathcal{P}_0(y))$ and the strong duality holds between $(\mathcal{P}_0(y))$ and $(\mathcal{D}_0(y))$, i.e.*

$$\min_{\Phi(m)=y} \|m\|_{\mathrm{TV}} = \sup_{\|\Phi^* p\|_\infty \leqslant 1} \langle y,\, p\rangle. \tag{24}$$

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

$$\Phi^* p^\star \in \partial\|m^\star\|_{\mathrm{TV}} \tag{25}$$

where $m^\star$ is any solution to $(\mathcal{P}_0(y))$. Conversely, if (25) holds, then $m^\star$ and $p^\star$ are solutions of respectively $(\mathcal{P}_0(y))$ and $(\mathcal{D}_0(y))$.

Proof. We apply [15, Theorem II.4.1] to $(\mathcal{D}_0(y))$ (and not to $(\mathcal{P}_0(y))$ as would be natural) rewritten as

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

The infimum is finite since for any admissible p, $\langle -y,\, p\rangle = \langle m_0,\, \Phi p\rangle \geqslant -\|m_0\|_{\mathrm{TV}}$. Let $V = L^2(\mathbb{T})$, $Y = C(\mathbb{T})$ (endowed with the strong topology), $Y^* = \mathcal{M}(\mathbb{T})$, $F(u) = \langle -y,\, u\rangle$ 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. □

Lemma 1 ([20, Th. 3.5],[3, Prop. 5])**.** *Let m 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 with respect to the weak-* topology.*

Proof. We follow the proof given in [20], simply adapting it to our framework. By optimality of $\tilde{m}_\lambda$, one has

$$\frac{1}{2}\|\Phi\tilde{m}_\lambda - y - w\|_2^2 + \lambda\|\tilde{m}_\lambda\|_{\mathrm{TV}} \leqslant \frac{1}{2}\|\underbrace{\Phi m - y}_{=0} - w\|_2^2 + \lambda\|m\|_{\mathrm{TV}}, \tag{26}$$

so that

$$\lim_{\lambda\to 0, \|w\|_2 \to 0} \|\Phi\tilde{m}_\lambda - y\|_2 = 0 \quad \text{and} \quad \limsup_{\lambda\to 0, \|w\|_2 \to 0} \|\tilde{m}_\lambda\|_{\mathrm{TV}} \leqslant \|m\|_{\mathrm{TV}}.$$

Thus,

$$\begin{aligned}
&\limsup_{\lambda\to 0, \|w\|_2 \to 0} \left[\frac{1}{2}\|\Phi\tilde{m}_\lambda - y - w\|_2^2 + \lambda_{\max}\|\tilde{m}_\lambda\|_{\mathrm{TV}}\right] \\
&\leqslant \limsup_{\lambda\to 0, \|w\|_2 \to 0} \left[\frac{1}{2}\|\Phi\tilde{m}_\lambda - y - w\|_2^2 + \lambda\|\tilde{m}_\lambda\|_{\mathrm{TV}}\right] + \limsup_{\lambda\to 0, \|w\|_2 \to 0} (\lambda_{\max} - \lambda)\|\tilde{m}_\lambda\|_{\mathrm{TV}} \\
&\leqslant \lambda_{\max}\|m\|_{\mathrm{TV}} < +\infty
\end{aligned}$$

Since closed balls for the total variation norm are sequentially compact for the weak-* topology, from any sequence we may extract a subsequence (λ_k, w_k) such that $\tilde{m}_{\lambda_k}$ converges to some $\tilde{m}$ with respect to the weak-* topology. From the weak-* to weak continuity of Φ we get $\Phi\tilde{m} = y$, and from the lower semi-continuity of the total variation, one has

$$\|\tilde{m}\|_{\mathrm{TV}} \leqslant \liminf_{k\to+\infty} \|\tilde{m}_{\lambda_k}\|_{\mathrm{TV}} \leqslant \limsup_{k\to+\infty} \|\tilde{m}_{\lambda_k}\|_{\mathrm{TV}} \leqslant \|m\|_{\mathrm{TV}}.$$

The original measure m being identifiable, we see that $\tilde{m} = m$. Since this is true for any subsequence, the claimed result is proved. □

References

[1] J-M. Azais, Y. De Castro, and F. Gamboa. Spike detection from inaccurate samplings. Technical report, January 2013.

[2] 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.

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

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

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

[6] E. J. Candès and C. Fernandez-Granda. Super-resolution from noisy data. *CoRR*, abs/1211.0290, 2012.

[7] E. J. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics. To appear.*, 2013.

[8] 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.

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

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

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

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

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

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

[15] 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.

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

[17] C. Fernandez-Granda. Support detection in super-resolution. *CoRR*, abs/1302.3921, 2013.

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

[19] 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.

[20] 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.

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

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

[23] D. A. Lorenz and D. Trede. Greedy Deconvolution of Point-like Objects. In Rémi Gribonval, editor, *SPARS'09 - Signal Processing with Adaptive Sparse Structured Representations*, Saint Malo, France, 2009.

[24] 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*, 42:1386–1391, October 1994.

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

[26] 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.

[27] 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.

[28] D. Slepian. Prolate spheroidal wave functions, Fourier analysis, and uncertainty. V - The discrete case. *AT&T Technical Journal*, 57:1371–1430, June 1978.

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

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

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

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