Sparse Spikes Super-resolution on Thin Grids I: the LASSO

Vincent Duval[1,2], **Gabriel Peyré**[2,1]
[1] INRIA, MOKAPLAN and CEREMADE, Université Paris-Dauphine

E-mail: vincent.duval@inria.fr
[2] CNRS and DMA, École Normale Supérieure

E-mail: gabriel.peyre@ens.fr

Abstract. This article analyzes the recovery performance in the presence of noise of sparse ℓ^1 regularization, which is often referred to as the Lasso or Basis-Pursuit. We study the behavior of the method for inverse problems regularization when the discretization step size tends to zero. We assume that the sought after sparse sum of Diracs is recovered when there is no noise (a condition which has been thoroughly studied in the literature) and we study what is the support (in particular the number of Dirac masses) estimated by the Lasso when noise is added to the observation. We identify a precise non-degeneracy condition that guarantees that the recovered support is close to the initial one. More precisely, we show that, in the small noise regime, when the non-degeneracy condition holds, this method estimates twice the number of spikes as the number of original spikes. Indeed, we prove that the Lasso detects two neighboring spikes around each location of an original spike. While this paper is focussed on cases where the observations vary smoothly with the spikes locations (e.g. the deconvolution problem with a smooth kernel), an interesting by-product is an abstract analysis of the support stability of discrete ℓ^1 regularization, which is of an independent interest. We illustrate the usefulness of this abstract analysis to analyze for the first time the support instability of compressed sensing recovery.

1. Introduction

We consider the problem of estimating an unknown Radon measure on the torus $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ (i.e. an interval with periodic boundary conditions), $m_0 \in \mathcal{M}(\mathbb{T})$, from low-resolution noisy observations in a separable Hilbert space $\mathcal{H}$,

$$y = \Phi(m_0) + w \in \mathcal{H} \tag{1}$$

where $w \in \mathcal{H}$ is some measurement noise, and $\Phi : \mathcal{M}(\mathbb{T}) \to \mathcal{H}$ is a bounded linear map such that

$$\forall\, m \in \mathcal{M}(\mathbb{T}), \quad \Phi(m) = \int_{\mathbb{T}} \varphi(x)\mathrm{d}m(x), \tag{2}$$

where $\varphi \in \mathscr{C}^2(\mathbb{T}, \mathcal{H})$.

A typical example of such an operation is a convolution, where $\mathcal{H} = \mathrm{L}^2(\mathbb{T})$ and $\varphi(x) : x' \mapsto \tilde{\varphi}(x' - x)$ for some smooth function $\tilde{\varphi}$ defined on $\mathbb{T}$. Another example is a partial Fourier transform, where $\mathcal{H} = \mathbb{C}^P$, and $\varphi(x) = (e^{2\mathrm{i}\pi\omega_k x})_{k=1}^P \in \mathcal{H}$ where $\omega_k \in \mathbb{Z}$ are the measured frequencies. For instance, using low frequency $-f_c \leqslant \omega_k = k - f_c - 1 \leqslant f_c$ with $P = 2f_c + 1$ is equivalent to using a convolution with the ideal low-pass filter

$$\forall\, x \in \mathbb{T}, \quad \tilde{\varphi}(x) = \sum_{k=-f_c}^{f_c} e^{2\mathrm{i}\pi k x}, \tag{3}$$

with cutoff frequency f_c. To simplify the notation, we shall assume that $\mathcal{H}$ is a real Hilbert space, and we leave to the reader the straightforward adaptations to the complex case.

1.1. Sparse Regularization

The problem of inverting (1) is severely ill-posed. A particular example is when Φ is a low pass filter, which is a typical setting for many problems in imaging. In several applications, it makes sense to impose some sparsity assumption on the data to recover. This idea has been introduced first in the geoseismic literature, to model the layered structure of the underground using sparse sums of Dirac masses [12]. Sparse regularization has later been studied by David Donoho and co-workers, see for instance [16].

In order to recover sparse measures (i.e. sums of Diracs), it makes sense to consider the following regularization

$$\min_{m\in\mathcal{M}(\mathbb{T})} \frac{1}{2}\|y - \Phi(m)\|^2 + \lambda|m|(\mathbb{T}) \tag{4}$$

where $|m|(\mathbb{T})$ is the total variation of the measure m, defined as

$$|m|(\mathbb{T}) \overset{\text{def.}}{=} \sup\left\{ \int_{\mathbb{T}} \psi(x)\mathrm{d}m(x) \;;\; \psi \in \mathscr{C}(\mathbb{T}),\ \|\psi\|_\infty \leqslant 1 \right\}. \tag{5}$$

This formulation of the recovery of sparse Radon measures has recently received lots of attention in the literature, see for instance the works of [5, 14, 9]. In the case where there is no noise, $w = 0$, it makes sense to consider $\lambda \to 0$ and to solve the following limit problem

$$\min_{m\in\mathcal{M}(\mathbb{T})} \left\{ |m|(\mathbb{T}) \;;\; \Phi(m) = \Phi(m_0) \right\}. \tag{6}$$

1.2. Lasso

The optimization problem (4) is convex but infinite dimensional, and while there exists solvers when Φ is measuring a finite number of Fourier frequency (see [9]), they do not scale well with the number of frequencies. Furthermore, the case of an arbitrary linear operator Φ is still difficult to handle, see [5] for an iterative scheme. The vast majority of practitioners thus approximate (4) by a finite dimensional problem computed over a finite grid $\mathcal{G} \overset{\text{def.}}{=} \{z_i \,;\, i \in [\![0, G-1]\!]\} \subset \mathbb{T}$, by restricting their attention to measures of the form

$$m_{a,\mathcal{G}} \overset{\text{def.}}{=} \sum_{i=0}^{G-1} a_i \delta_{z_i} \in \mathcal{M}(\mathbb{T}).$$

For such a discrete measure, one has $|m|(\mathbb{T}) = \sum_{i=0}^{G-1} |a_i| = \|a\|_1$, which can be interpreted as the fact that $|\cdot|(\mathbb{T})$ is the natural extension of the ℓ^1 norm from finite dimensional vectors to the infinite dimensional space of measures. Inserting this parametrization in (4) leads to the celebrated Basis-Pursuit problem [11], which is also known as the LASSO method in statistics [32],

$$\min_{a \in \mathbb{R}^N} \frac{1}{2} \|y - \Phi_{\mathcal{G}} a\|^2 + \lambda \|a\|_1 \tag{7}$$

where in the following we make use of the notations

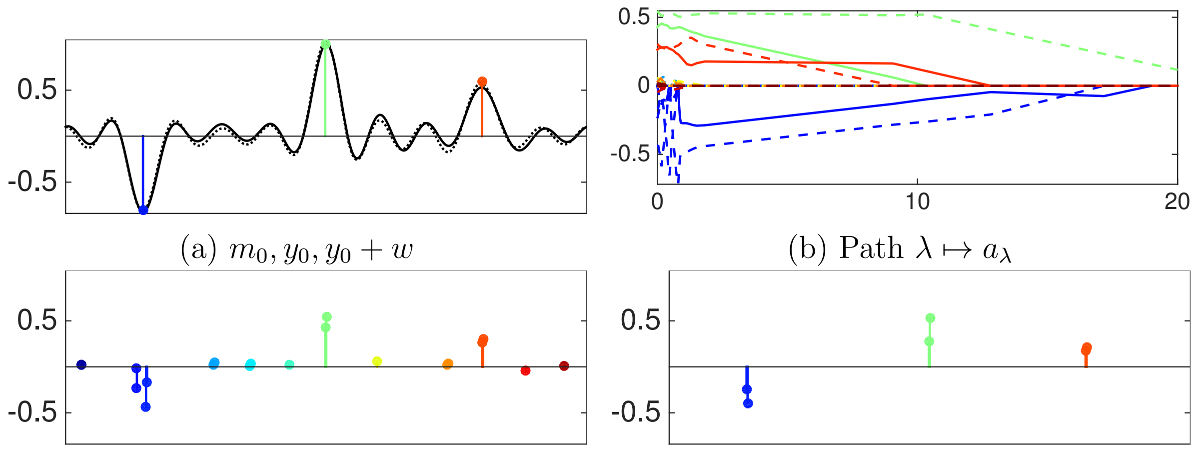
$$\Phi_{\mathcal{G}} a \overset{\text{def.}}{=} \Phi(m_{a,\mathcal{G}}) = \sum_{i=0}^{G-1} a_i \varphi(z_i), \tag{8}$$

One can understand (7) as performing a nearest neighbor interpolation of the Dirac's locations.

Note that while we focus in this paper on convex recovery method, and in particular ℓ^1-type regularization, there is a vast literature on the subject, which makes use of alternative algorithms, see for instance [27, 4] and the references therein.

1.3. Motivating Example

Figure (1) illustrates the typical behavior of the Lasso method (7) to estimate a sparse input measure m_0 (shown in (a)) from observations $y = \Phi m_0 + w$, where Φ is the ideal low-pass filter with cutoff frequency f_c, i.e. $\varphi(x) = \tilde{\varphi}(x - \cdot)$ where $\tilde{\varphi}$ is defined in (3). In the numerical simulation, we used $f_c = 12$ and an uniform grid of $G = 512$ points. Here w is a small input noise, and its impact can be visualized in (a) where both $y_0 = \Phi m_0$ (plain black curve) and $y = y_0 + w$ (dashed black curve) are displayed. As can be expected, the recovered a_λ (solution of (7)) with a small value of λ (here $\lambda = 0.05$ is displayed in (c)) is bad because too much noise contaminates the result. A well chosen value of λ (here $\lambda = 4$ is displayed in (d)) is able to remove the noise, and to detect spikes located near the input spikes composing m_0. However, as showed in [20], in this small noise setting, one can recover up to twice as many spikes as the input measures, because the spikes of m_0 can get duplicated on immediate nearest neighbors on the grid $\mathcal{G}$. Figure 1, (b), further refines this analysis by displaying the whole path $\lambda \mapsto a_\lambda$ (dashed curves indicate spurious spikes whose locations do not match those of the input measure m_0). It is the goal of this paper to precisely analyze and quantify this behavior. In particular, we precisely characterize the "extended support" (those grid locations that are selected when the noise is small and λ well chosen) and show that for deconvolution, it is exactly composed of pairs of nearest neighbors.



(a) $m_0, y_0, y_0 + w$ (b) Path $\lambda \mapsto a_\lambda$

(c) $a_\lambda, \lambda = 0.05$ (d) $a_\lambda, \lambda = 4$

Figure 1: Sparse spikes deconvolution results obtained by computing the solution a_λ of (7). The color reflects the positions of the spikes on the 1-D grid. (a) shows the input measure m_0 and the observation $y = y_0 + w$. (b) shows how the solution a_λ (vertical axis) evolves with λ (horizontal axis). Each curve shows the evolution of $\lambda \mapsto (a_\lambda)_i$ for indexes $i \in \{1, \ldots, G-1\}$. The color encodes the value of i. Plain curves correspond to correct spikes locations i associated to the input measure m_0. Dashed curves correspond to incorrect spikes (not present in the input measure m_0). (c,d) show the results a_λ obtained for two different values of λ.

1.4. Previous Works

Most of the early work to assess the performance of convex sparse regularization has focussed its attention on the finite dimensional case, thus considering only the Lasso problem (7). While the literature on this subject is enormous, only very few works actually deal with deterministic and highly correlated linear operators such as low-pass convolution kernels. The initial works of Donoho [16] study the Lipschitz behavior of the inverse map $y \mapsto a^\star$, where $a^\star$ is a solution of (7), as a function of the bandwidth of the bandpass filter. The first work to address the question of spikes identification (i.e. recovery of the exact location of the spikes over a discrete grid) is [18]. This work uses the analysis of ℓ^1 regularization introduced by Fuchs in [23]. This type of analysis ensures that the support of the input measure is stable under small noise perturbation of the measurements. Our finding is that this is however never the case (the support is always unstable) when the grid is thin enough, and we thus introduce the notion of "extended support", which is in some sense the smallest extension of the support which is stable. The idea of extending the support to study the recovery performance of ℓ^1 methods can be found in the work of Dossal [17] who focusses on noiseless recovery and stability in term of ℓ^2 error.

Recently, a few works have studied the theoretical properties of the recovery over measures (4). Candès and Fernandez-Granda show in [9] that this convex program does recover exactly the initial sparse measure when $w = 0$ and $\lambda \to 0$ (i.e. program (6)) under a minimum-separation condition, i.e. if the spikes are well-separated. The robustness to noisy measurements is analyzed by the same authors in [8] using an Hilbertian norm, and in [22, 2] in terms of spikes localization. The work of [30] analyzes the reconstruction error. Lastly, [20] provides a condition ensuring that (4) recovers the same number of spikes as the input measure and that the error in terms of spikes localization and elevation has the same order as the noise level. It is important to note that in the special case where m_0 is a positive measure, then m_0 is always a solution to (6), as shown in [14] (see also [15] for a refined analysis of the stability to noise in this special case).

Very few works have tried to bridge the gap between these grid-free methods over the space of measures, and finite dimensional discrete approximations that are used by practitioners. These theoretical questions are however relevant from a practitioner's point of view, and we refer [26] for experimental observations of the impact of discretization and the corresponding recovery bias. The convergence (in the sense of measures) of the solutions of the discrete problem toward to ones of the grid-free problem is shown in [31], where a speed of convergence is shown using tools from semi-infinite programming [29]. The same authors show in [3] that the discretized problem achieves a similar prediction L^2 error as the grid-free method. Γ-convergence results on ℓ^1 but also ℓ^0 regularization are provided in the PhD work of [25]. In [20], we have shown that solutions of the discrete Lasso problem estimate in general as much as twice the number of spikes as the input measure. We detail in the following section how the present work gives a much more precise and general analysis of this phenomenon.

1.5. Contributions

Our paper is composed of two contributions (Theorems 1 and 2) that study the robustness to noise of the support of the solution of Lasso finite dimensional recovery problems. We stress the fact that we always suppose that the sought after sparse measure m_0 is identifiable, i.e. is the solution of the BLASSO program (6) (i.e. in the noiseless case $w = 0, \lambda = 0$). This mandatory hypothesis is now well understood, as detailed in Section 1.4, and is always true if the measure m_0 is positive, or under a minimum separation distance between the spikes. Our main contributions study whether the support of the recovered solution is close from the one of m_0 in the presence of a small noise. Such a stability cannot hold in full generality, and requires a strengthening of the optimality condition for m_0 being identifiable, which we refers in the following as a "non-degeneracy" condition.

Section 2 presents our first contribution. This is an improvement over the known analysis of the Lasso in an abstract setting (that is (7) when $\Phi_{\mathcal{G}}$ is replaced with any finite dimensional linear operator). Whereas Fuchs' result [23] characterizes the exact support recovery of the Lasso at low noise, our previous work [20] has pointed out that when Fuchs'criterion is not satisfied, the nonzero components of the solutions of the Basis-Pursuit at low noise are contained in the *extended support*, that is the saturation set of some *minimal norm dual certificate*. Theorem 1 states that under a sufficient non-degeneracy condition (hypothesis (20), which holds generically), all the components of the extended support are actually nonzero (with a prediction on the signs).

Section 3 applies this result to Problem (7) on thin grids. After recalling the convergence properties of Problem (7) towards (4), we show that, if the input measure $m_0 = m_{\alpha_0,x_0} = \sum_{\nu=1}^N \alpha_{0,\nu}\delta_{x_{0,\nu}}$ has support on the grid (*i.e.* $x_{0,\nu} \in \mathcal{G}$ for all ν), and if a non-degeneracy condition holds (the "Non-Degenerate Source Condition", see Definition 2), the methods actually reconstructs at low noise pairs of Dirac masses, i.e. solutions of the form

$$m_\lambda = \sum_{\nu=1}^{N} \left(\alpha_{\lambda,\nu}\delta_{x_{0,\nu}} + \beta_{\lambda,\nu}\delta_{x_{0,\nu}+\varepsilon_\nu h}\right), \quad \text{where} \quad \varepsilon_\nu \in \{-1,+1\}, \tag{9}$$

$$\text{and} \quad \operatorname{sign}(\alpha_{\lambda,\nu}) = \operatorname{sign}(\beta_{\lambda,\nu}) = \operatorname{sign}(\alpha_{0,\nu}). \tag{10}$$

The precise statement of this result can be found in Theorem 2. Compared to [20] where it is predicted that spikes could appear at most in pairs, this result states that all the pairs do appear, and it provides a closed-form expression for the shift ε. That closed-form expression does not vary as the grid is refined, so that the side on which each neighboring spike appears is in fact intrinsic to the measure, we call it the *natural shift*. Moreover, we characterize the low noise regime as $\frac{\|w\|_2}{\lambda} = O(1)$ and $\lambda = O(h)$.

It is worth emphasizing that, in this setting of spikes retrieval on thin grids, our contributions give important information about the structure of the recovered spikes when the noise w is small. This is especially important since, contrary to common belief, the spikes locations for Lasso are not stable: even for an arbitrary small noise w, neither methods retrieve the correct input spikes locations.

Eventually, we illustrate in Section 4 our abstract analysis of the Lasso problem (7) (as provided by Theorem 1) to characterize numerically the behavior of the Lasso for compressed sensing (CS) recovery (i.e. when one replaces the filtering $\Phi_{\mathcal{G}}$ appearing in (7) with a random matrix). The literature on CS only describes the regime where enough measurements are available so that the support is stable, or does not study support stability but rather ℓ^2 stability. Theorem 1 allows us to characterize numerically how much the support becomes unstable (in the sense that the extended support's size increases) as the number of measurements decreases (or equivalently the sparsity increases).

1.6. Notations and preliminaries

The set of Radon measures (resp. positive Radon measures) is denoted by $\mathcal{M}(\mathbb{T})$ (resp. $\mathcal{M}^+(\mathbb{T})$). Endowed with the total variation norm (5), $\mathcal{M}(\mathbb{T})$ is a Banach space. Another useful topology on $\mathcal{M}(\mathbb{T})$ is the weak* topology: a sequence of measures $(m_n)_{n\in\mathbb{N}}$ weak* converges towards $m \in \mathcal{M}(\mathbb{T})$ if and only if for all $\psi \in \mathscr{C}(\mathbb{T})$, $\lim_{n\to+\infty}\int_{\mathbb{T}}\psi \mathrm{d}m_n = \int_{\mathbb{T}}\psi \mathrm{d}m$. Any bounded subset of $\mathcal{M}(\mathbb{T})$ (for the total variation) is relatively sequentially compact for the weak* topology. Moreover the topology induced by the total variation is stronger than the weak* topology, and the total variation is sequentially lower semi-continuous for the weak* topology. Throughout the paper, given $\alpha \in \mathbb{R}^N$ and $x_0 \in \mathbb{T}^N$, the notation $m_{\alpha,x_0} \overset{\text{def.}}{=} \sum_{\nu=1}^N \alpha_\nu \delta_{x_{0,\nu}}$ hints that $\alpha_\nu \neq 0$ for all ν (contrary to the notation $m_{a,\mathcal{G}}$), and that the $x_{0,\nu}$'s are pairwise distinct.

Given a separable Hilbert space $\mathcal{H}$, the properties of $\Phi : \mathcal{M}(\mathbb{T}) \to \mathcal{H}$ and its adjoint are recalled in Proposition 1 in Appendix. The $\infty, 2$-operator norm of $\Phi^* : \mathcal{H} \to \mathscr{C}(\mathbb{T})$ is defined as $\|\Phi^*\|_{\infty,2} \overset{\text{def.}}{=} \sup \{\|\Phi^* w\|_\infty \; ; \; w \in \mathcal{H}, \|w\| \leqslant 1\}$ (and the $\infty, 2$ operator norm of a matrix is defined similarly). Given a vector $x_0 \in \mathbb{T}^N$, Φ_{x_0} refers to the linear operator $\mathbb{R}^N \to \mathcal{H}$, with

$$\forall \alpha \in \mathbb{R}^N, \quad \Phi_{x_0} \alpha \overset{\text{def.}}{=} \Phi(m_{\alpha,x_0}) = \sum_{\nu=1}^N \alpha_\nu \varphi(x_{0,\nu}).$$

It may also be seen as the restriction of Φ to measures supported on the set $\{x_{0,\nu} \; ; \; \nu \in [\![1, N]\!]\}$. A similar notation is adopted for Φ'_{x_0} (replacing $\varphi(x_{0,\nu})$ with $\varphi'(x_{0,\nu})$. The concatenation of Φ_{x_0} and Φ'_{x_0} is denoted by $\Gamma_{x_0} \overset{\text{def.}}{=} \begin{pmatrix} \Phi_{x_0} & \Phi'_{x_0} \end{pmatrix}$.

We shall rely on the notion of set convergence. Given a sequence $(C_n)_{n\in\mathbb{N}}$ of subsets of $\mathbb{T}$, we define

$$\limsup_{n\to+\infty} C_n = \left\{ x \in \mathbb{T} \; ; \; \liminf_{n\to+\infty} d(x, C_n) = 0 \right\} \tag{11}$$

$$\liminf_{n\to+\infty} C_n = \left\{ x \in \mathbb{T} \; ; \; \limsup_{n\to+\infty} d(x, C_n) = 0 \right\} \tag{12}$$

where d is defined by $d(x, C) = \inf_{x'\in C} |x' - x|$ and $|x - x'|$ refers to the distance between x and x' on the torus. If both sets are equal, let C be the corresponding set (then C is necessarily closed), we write

$$\lim_{n\to+\infty} C_n = C. \tag{13}$$

If the sequence $(C_n)_{n\in\mathbb{N}}$ is nondecreasing ($C_n \subset C_{n+1}$), then $\lim_{n\to\infty} C_n = \overline{\bigcup_{n\in\mathbb{N}} C_n}$, and if it is nonincreasing ($C_n \supset C_{n+1}$) then $\lim_{n\to\infty} C_n = \bigcap_{n\in\mathbb{N}} \overline{C_n}$ (where $\overline{C}$ denotes the closure of C). We refer the reader to [28] for more detail about set convergence. We shall also use this notion in Hilbert spaces, with obvious adaptations.

2. Abstract analysis of the Lasso

The aim of this section is to study the low noise regime of the Lasso problem in an abstract finite dimensional setting, regardless of the grid stepsize. In this framework, the columns of the (finite dimensional) degradation operator need not be the samples of a continuous (*e.g.* convolution) operator, and the provided analysis holds for any general Lasso problem. We extend the initial study of Fuchs of the basis pursuit method (see [23]) which gives the analytical expression of the solution when the noise is low and the support is stable. Here, provided we have access to a particular dual vector η_0, we give an explicit parametrization the solutions of the basis pursuit at low noise *even when the support is not stable.* This is especially relevant for the deconvolution problem since the support is not stable when the grid is thin enough.

2.1. Notations and optimality conditions

We consider in this section observations in an arbitrary separable Hilbert space $\mathcal{H}$, which might be for instance $L^2(\mathbb{T})$ (*e.g.* in the case of a convolution) or a finite dimensional vector space. The linear degradation operator is then denoted as $A : \mathbb{R}^G \to \mathcal{H}$. Let us emphasize that in this section, for $a \in \mathbb{R}^G$, $\|a\|_\infty \overset{\text{def.}}{=} \max_{0 \leqslant k \leqslant G-1} |a_k|$.

Given an observation $y_0 = Aa_0 \in \mathcal{H}$ (or $y = y_0 + w$, where $w \in \mathcal{H}$), we aim at reconstructing the vector $a_0 \in \mathbb{R}^G$ by solving the LASSO problem for $\lambda > 0$,

$$\min_{a \in \mathbb{R}^G} \frac{1}{2}\|y - Aa\|^2 + \lambda \|a\|_1 \tag{$\mathcal{P}_\lambda(y)$}$$

and for $\lambda = 0$ we consider the (Basis-Pursuit) problem

$$\min_{a \in \mathbb{R}^G} \|a\|_1 \text{ such that } Aa = y_0. \tag{$\mathcal{P}_0(y_0)$}$$

If $a \in \mathbb{R}^G$, we denote by $I(a)$, or I when the context is clear, the support of a, *i.e.* $I(a) \overset{\text{def.}}{=} \{i \in [\![0, G-1]\!] \; ; \; a_i \neq 0\}$. Also, we let $s_I \overset{\text{def.}}{=} \operatorname{sign}(a_I)$, and $\operatorname{supp}^\pm(a) \overset{\text{def.}}{=} \{(i, s_i) \; ; \; i \in I\}$ the *signed support* of a.

The optimality conditions for Problems ($\mathcal{P}_\lambda(y)$) and ($\mathcal{P}_0(y_0)$) are quite standard, as detailed in the following proposition.

Proposition 1. *Let $y \in \mathcal{H}$, and $a_\lambda \in \mathbb{R}^G$. Then a_λ is a solution to ($\mathcal{P}_\lambda(y)$) if and only if there exists $p_\lambda \in \mathcal{H}$ such that*

$$\|A^* p_\lambda\|_\infty \leqslant 1, \quad \textit{and} \quad (A^* p_\lambda)_I = \operatorname{sign}(a_{\lambda,I}), \tag{14}$$

$$\lambda A^* p_\lambda + A^*(Aa_\lambda - y) = 0. \tag{15}$$

Similarly, if $a_0 \in \mathbb{R}^G$, then a_0 is a solution to ($\mathcal{P}_0(y_0)$) if and only if $Aa_0 = y_0$ and there exists $p \in \mathcal{H}$ such that.

$$\|A^* p\|_\infty \leqslant 1 \quad \textit{and} \quad (A^* p)_I = \operatorname{sign}(a_{0,I}). \tag{16}$$

Conditions (14) and (16) merely express the fact that $\eta_\lambda \overset{\text{def.}}{=} A^* p_\lambda$ (resp. $\eta \overset{\text{def.}}{=} A^* p$) is in the subdifferential of the ℓ^1-norm at a_λ (resp. a_0). In that case we say that η_λ (resp. η) is a *dual certificate* for a_λ (resp. a_0). Condition (16) is also called the *source condition* in the literature [6].

The term *dual certificate* stems from the fact that p_λ (resp. p) is a solution to the dual problem to ($\mathcal{P}_\lambda(y)$) (resp. ($\mathcal{P}_0(y_0)$)),

$$\inf_{p \in C} \left\| \frac{y}{\lambda} - p \right\|_2^2, \tag{$\mathcal{D}_\lambda(y)$}$$

$$\text{resp.} \quad \sup_{p \in C} \langle y_0, p \rangle, \tag{$\mathcal{D}_0(y_0)$}$$

$$\text{where} \quad C \overset{\text{def.}}{=} \left\{ p \in \mathcal{H} \; ; \; \max_{k \in [\![0, G-1]\!]} |(A^* p)_k| \leqslant 1 \right\}. \tag{17}$$

If a is a solution to ($\mathcal{P}_\lambda(y)$) and p_λ is a solution to ($\mathcal{D}_\lambda(y)$), then (14) and (15) hold. Conversely, for any $a \in \mathbb{R}^P$ and any $p_\lambda \in \mathcal{H}$, if (14) and (15) hold, then a is a solution to ($\mathcal{P}_\lambda(y)$) and p_λ is a solution to ($\mathcal{D}_\lambda(y)$). A similar equivalence holds for ($\mathcal{P}_0(y_0)$) and ($\mathcal{D}_0(y_0)$).

Remark 1. In general, the solutions to ($\mathcal{P}_\lambda(y)$) and ($\mathcal{P}_0(y_0)$) need not be unique. However, the dual certificate $\eta_\lambda = A^* p_\lambda$ which appears in (14) and (15) is unique. On the contrary, the dual certificate $\eta = A^* p$ which appears in (16) is not unique in general.

We say that a vector a_0 is *identifiable* if it is the unique solution to ($\mathcal{P}_0(y_0)$) for the input $y = Aa_0$. The following classical result gives a sufficient condition for a_0 to be identifiable.

Proposition 2. *Let $a_0 \in \mathbb{R}^G$ such that A_I is injective and that there exists $p \in \mathcal{H}$ such that*

$$\|(A^*p)_{I^c}\|_\infty < 1 \quad \text{and} \quad (A^*p)_I = \operatorname{sign}(a_{0,I}), \tag{18}$$

where $I^c = [\![1, G]\!] \setminus I$. Then a_0 is identifiable.

Conversely, if a_0 is identifiable, there exists $p \in \mathcal{H}$ such that (18) holds and A_I is injective (see [24, Lemma 4.5]).

2.2. Extended support of the LASSO

From now on, we assume that the vector $a_0 \in \mathbb{R}^G$ is identifiable (*i.e.* a_0 is the unique solution to ($\mathcal{P}_0(y_0)$) where $y_0 = Aa_0$). We denote by $I = \operatorname{supp}(a_0)$ and $s_I = \operatorname{sign}(a_{0,I})$ the support and the sign of a_0.

It is well known that ($\mathcal{P}_0(y_0)$) is the limit of ($\mathcal{P}_\lambda(y)$) for $\lambda \to 0$ (see [11] for the noiseless case and [24] when the observation is $y = y_0 + w$ and the noise w tends to zero as a multiple of λ) at least in terms of the ℓ^2 convergence. In terms of the support of the solutions, the study in [20], which extends the one by Fuchs [23], emphasizes the role of a specific *minimal-norm certificate* η_0 which governs the behavior of the model at low noise regimes.

Definition 1 (Minimal-norm certificate and extended support)**.** *Let $a_0 \in \mathbb{R}^G$, and let p_0 be the solution to ($\mathcal{D}_0(y_0)$) with minimal norm. The minimal-norm certificate of a_0 is defined as $\eta_0 \stackrel{\text{def.}}{=} A^* p_0$. The set of indices $\operatorname{ext}(a_0) \stackrel{\text{def.}}{=} \{1 \leqslant j \leqslant G \,;\, |(\eta_0)_j| = 1\}$ is called the extended support of a_0, and the set $\operatorname{ext}^\pm(a_0) \stackrel{\text{def.}}{=} \{(j, (\eta_0)_j) \,;\, j \in \operatorname{ext}(a_0)\} \subset [\![0, G-1]\!] \times \{-1, 1\}$ is called the extended signed support of a_0.*

Remark 2. In the case where a_0 is a solution to ($\mathcal{P}_0(y_0)$) (which is the case here since we assume that a_0 is an identifiable vector for ($\mathcal{P}_0(y_0)$)), we have $(I, \operatorname{sign}(a_{0,I})) \subset \operatorname{ext}^\pm(a_0)$. The minimal norm certificate thus turns out to be

$$\eta_0 = A^* p_0 \quad \text{where} \quad p_0 = \underset{p \in \mathcal{H}}{\operatorname{argmin}} \left\{ \|p\|_2 \,;\, \|A^*p\|_\infty \leqslant 1 \quad \text{and} \quad A_I^* p = s_I \right\}. \tag{19}$$

The minimal norm certificate governs the (signed) support of the solution at low noise regimes insofar as the latter is contained in the extended signed support. The following new theorem, which is proved in Appendix C, shows that, in the generic case, both signed supports are equal.

Theorem 1. *Let $a_0 \in \mathbb{R}^G \setminus \{0\}$ be an identifiable signal , $J \overset{\text{def.}}{=} \text{ext}(a_0)$ such that A_J has full rank, and $v_J \overset{\text{def.}}{=} (A_J^* A_J)^{-1} \operatorname{sign}(\eta_{0,J})$. Assume that the following non-degeneracy condition holds*

$$\forall\, j \in J \setminus I, \quad v_j \neq 0. \tag{20}$$

Then, there exists constants $C^{(1)} > 0$, $C^{(2)} > 0$ (which depend only on A, J and $\operatorname{sign}(a_{0,J})$) such that for $0 < \lambda \leqslant C^{(1)} \left(\min_{i\in I} |a_{0,I}|\right)$ and all $w \in \mathcal{H}$ with $\|w\| \leqslant C^{(2)}\lambda$ the solution $\tilde{a}_\lambda$ of $(\mathcal{P}_\lambda(y))$ is unique, $\operatorname{supp}(\tilde{a}_\lambda) = J$ and it reads

$$\tilde{a}_{\lambda,J} = a_{0,J} + A_J^+ w - \lambda (A_J^* A_J)^{-1} \operatorname{sign}(\eta_{0,J}),$$

where $A_J^+ = (A_J^* A_J)^{-1} A_J^*$.

Remark 3 (Comparison with the analysis of Fuchs). When $J = I$, Theorem 1 recovers exactly the result of Fuchs [23]. Note that this result has been extended beyond the ℓ^1 setting,see in particular [34, 33] for a unified treatment of arbitrary partly smooth convex regularizers. For this result to hold, i.e. to obtain $I = J$, one needs to impose that the following pre-certificate

$$\eta_F \overset{\text{def.}}{=} A^* A_I^{+,*} s_I \tag{21}$$

is a valid certificate, i.e. one needs that $\|\eta_{F,I^c}\|_\infty < 1$. This condition is often called the *irrepresentability condition* in the statistics literature (see for instance [36]). It implies that the support I is stable for small noise. Unfortunately, it is easy to verify that for the deconvolution problem, in general, this condition does not hold when the grid stepsize is small enough (see [20, Section 5.3]), so that one cannot use the initial result. This motivates our additional study of the extended support $\text{ext}(a_0) \supset I$, which is always stable to small noise. While this new result is certainly very intuitive, to the best of our knowledge, it is the first time it is stated and proved, with explicit values of the stability constant involved.

Remark 4. Theorem 1 guarantees that the support of the reconstructed signal $\tilde{a}_\lambda$ at low noise is equal to the extended support. The required condition $v_j \neq 0$ in Theorem 1 is tight in the sense that if $v_j = 0$ for some $j \in J \setminus I$, then the saturation point of η_λ may be strictly included in J. Indeed, it is possible, using similar calculations as above, to construct w such that $\operatorname{supp} \tilde{a}_\lambda \subsetneq J$ with λ and $\|w\|_2/\lambda$ arbitrarily small.

3. Lasso on Thin Grids

In this section, we focus on inverse problems with smooth kernels, such as for instance the deconvolution problem. Our aim is to recover a measure $m_0 \in \mathcal{M}(\mathbb{T})$ from the observation $y_0 = \Phi m_0$ or $y = \Phi m_0 + w$, where $\varphi \in \mathscr{C}^k(\mathbb{T};\mathcal{H})$ ($k \geqslant 2$), $w \in \mathcal{H}$ and

$$\forall x \in \mathbb{T},\ \Phi m \overset{\text{def.}}{=} \int_{\mathbb{T}} \varphi(x)\mathrm{d}m(x), \tag{22}$$

so that $\Phi : \mathcal{M}(\mathbb{T}) \to \mathcal{H}$ is a bounded linear operator. Observe that Φ is in fact weak* to weak continuous and its adjoint is compact (see Lemma 1 in Appendix).

Typically, we assume that the unknown measure m_0 is sparse, in the sense that it is of the form $m_0 = \sum_{\nu=1}^{N} \alpha_{0,\nu}\delta_{x_{0,\nu}}$ for some $N \in \mathbb{N}^*$, here $\alpha_{0,\nu} \in \mathbb{R}^*$ and the $x_{0,\nu} \in \mathbb{T}$ are pairwise distinct.

The approach we study in this paper is the (discrete) Basis Pursuit. We look for measures that have support on a certain discrete grid $\mathcal{G} \subset \mathbb{T}$, and we want to recover the original signal by solving an instance of ($\mathcal{P}_0(y_0)$) or ($\mathcal{P}_\lambda(y)$) on that grid. Specifically, we aim at analyzing the behavior of the solutions at low noise regimes (*i.e.* when the noise w is small and λ well chosen) as the grid gets thinner and thinner. To this end, we take advantage of the characterizations given in Section 2, regardless of the grid, and we use the Beurling Lasso (4) as a limit of the discrete models.

3.1. Notations and Preliminaries

For the sake of simplicity we only study uniform grids, i.e. $\mathcal{G} \overset{\text{def.}}{=} \{ih\ ;\ i \in [\![0,\, G-1]\!]\}$ where $h \overset{\text{def.}}{=} \frac{1}{G}$ is the stepsize. Moreover, we shall consider sequences of grids $(\mathcal{G}_n)_{n\in\mathbb{N}}$ such that the stepsize vanishes ($h_n = \frac{1}{G_n} \to 0$ as $n \to +\infty$) and to ensure monotonicity, we assume that $\mathcal{G}_n \subset \mathcal{G}_{n+1}$. For instance, the reader may think of a dyadic grid (*i.e.* $h_n = \frac{h_0}{2^n}$). We shall identify in an obvious way measures with support in $\mathcal{G}_n$ (*i.e.* of the form $\sum_{k=0}^{G_n-1} a_k\delta_{kh_n}$) and vectors $a \in \mathbb{R}^{G_n}$.

The problem we consider is a particular instance of ($\mathcal{P}_\lambda(y)$) (or ($\mathcal{P}_0(y_0)$)) when choosing A as the restriction of Φ to measures with support in the grid $\mathcal{G}_n$,

$$A \overset{\text{def.}}{=} \Phi_{\mathcal{G}_n} = \begin{pmatrix} \varphi(0) & \dots & \varphi((G-1)h_n) \end{pmatrix}. \tag{23}$$

On the grid $\mathcal{G}_n$, we solve

$$\min_{a\in\mathbb{R}^{G_n}} \frac{1}{2}\|y - \Phi_{\mathcal{G}_n} a\|^2 + \lambda\|a\|_1, \tag{$\mathcal{P}_\lambda^n(y)$}$$

$$\text{and } \min_{a\in\mathbb{R}^{G_n}} \|a\|_1 \text{ such that } \Phi_{\mathcal{G}_n} a = y_0. \tag{$\mathcal{P}_0^n(y_0)$}$$

We say that a measure $m_0 = \sum_{\nu=1}^{N} \alpha_{0,\nu}\delta_{x_{0,\nu}}$ (with $\alpha_{0,\nu} \neq 0$ and the $x_{0,\nu}$'s pairwise distinct) is identifiable through ($\mathcal{P}_0^n(y_0)$) if it can be written as $m_0 = \sum_{k=0}^{G_n-1} a_i\delta_{ih_n}$ and that the vector a is identifiable using ($\mathcal{P}_0^n(y_0)$).

As before, given $a \in \mathbb{R}^{G_n}$, we shall write $I(a) \overset{\text{def.}}{=} \{i \in [\![0,\, G_n-1]\!]\ ;\ a_i \neq 0\}$ or simply I when the context is clear.

The optimality conditions (15) amount to the existence of some $p_\lambda \in \mathcal{H}$ such that

$$\max_{0\leqslant k\leqslant G_n-1} |(\Phi^* p_\lambda)(kh_n)| \leqslant 1, \quad \text{and} \quad (\Phi^* p_\lambda)(Ih_n) = \operatorname{sign}(a_{\lambda,I}), \tag{24}$$

$$\lambda\Phi^* p_\lambda + \Phi^*(\Phi a_\lambda - y) = 0. \tag{25}$$

Similarly the optimality condition (16) is equivalent to the existence of $p \in \mathcal{H}$ such that

$$\max_{0 \leqslant k \leqslant G_n - 1} |(\Phi^* p)(kh_n)| \leqslant 1 \quad \text{and} \quad (\Phi^* p)(Ih_n) = \operatorname{sign}(a_{0,I}). \tag{26}$$

Notice that the dual certificates are naturally given by the sampling of continuous functions $\eta = \Phi^* p : \mathbb{T} \to \mathbb{R}$, and that the notation $\eta(Ih_n)$ or $(\Phi^* p)(Ih_n)$ stands for $(\eta(ih_n))_{i\in I}$ where $I = I(a_0)$ (and similarly for $\eta_\lambda = \Phi^* p_\lambda$ and $I(a_\lambda)$).

If m_0 is identifiable through ($\mathcal{P}_0^n(y_0)$), the minimal norm certificate for the problem ($\mathcal{P}_0^n(y_0)$) (see Section 2) is denoted by η_0^n, whereas the extended support on $\mathcal{G}_n$ is defined as

$$\operatorname{ext}_n m_0 \overset{\text{def.}}{=} \{t \in \mathcal{G}_n \; ; \; \eta_0^n(t) = \pm 1\}. \tag{27}$$

From Section 2, we know that the extended support is the support of the solutions at low noise.

3.2. The Limit Problem: the Beurling Lasso

Problems ($\mathcal{P}_\lambda^n(y)$) and ($\mathcal{P}_0^n(y_0)$) have natural limits when the grid gets thin. Embedding those problems into the space $\mathcal{M}(\mathbb{T})$ of Radon measures, the present authors have studied in [20] their convergence towards the Beurling-LASSO used in [14, 9, 5, 30].

The idea is to recover the measure m_0 using the following variants of ($\mathcal{P}_\lambda(y)$) and ($\mathcal{P}_0(y_0)$):

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

$$\text{and} \quad \min_{m \in \mathcal{M}(\mathbb{T})} |m|(\mathbb{T}) \quad \text{such that} \quad \Phi m = y_0, \tag{$\mathcal{P}_0^\infty(y_0)$}$$

where $|m|(\mathbb{T})$ refers to the total variation of the measure m

$$|m|(\mathbb{T}) \overset{\text{def.}}{=} \sup \left\{ \int_{\mathbb{T}} \psi(x) \mathrm{d}m(x) \; ; \; \psi \in C(\mathbb{T}) \text{ and } \|\psi\|_\infty \leqslant 1 \right\}. \tag{28}$$

Observe that in this framework, the notation $\|\psi\|_\infty$ stands for $\sup_{t\in\mathbb{T}} |\psi(t)|$. When m is of the form $m = \sum_{\nu=1}^N \alpha_\nu x_\nu$ where $\alpha_\nu \in \mathbb{R}^*$ and $x_\nu \in \mathbb{T}$ (with the x_ν's pairwise distinct), $|m|(\mathbb{T}) = \sum_{\nu=1}^N |\alpha_\nu|$, so that those problems are natural extensions of ($\mathcal{P}_\lambda(y)$) and ($\mathcal{P}_0(y_0)$). This connection is emphasized in [20] by embedding ($\mathcal{P}_\lambda^n(y)$) and ($\mathcal{P}_0^n(y_0)$) in the space of Radon measures $\mathcal{M}(\mathbb{T})$, using the fact that

$$\sup \left\{ \int_{\mathbb{T}} \psi(x) \mathrm{d}m(x) \; ; \; \psi \in C(\mathbb{T}), \forall k \in [\![0,\, G_n - 1]\!] \; |\psi|(kh_n) \leqslant 1 \right\}$$
$$= \begin{cases} \|a\|_1 \text{ if } m = \sum_{k=0}^{G_n-1} a_k \delta_{kh_n}, \\ +\infty \text{ otherwise.} \end{cases}$$

We say that m_0 is identifiable through ($\mathcal{P}_0^\infty(y_0)$) if it is the unique solution of ($\mathcal{P}_0^\infty(y_0)$). A striking result of [9] is that when Φ is the ideal low-pass filter and that the spikes $m_0 = \sum_{\nu=1}^{N} \alpha_{0,\nu} x_{0,\nu}$ are sufficiently far from one another, the measure m_0 is identifiable through $\mathcal{P}_0^\infty(y_0)$.

The optimality conditions for ($\mathcal{P}_\lambda^\infty(y)$) and ($\mathcal{P}_0^\infty(y_0)$) are similar to those of the abstract LASSO (respectively (14), (15) and (16)). The corresponding dual problems are

$$\inf_{p \in C^\infty} \left\| \frac{y}{\lambda} - p \right\|_2^2, \qquad (\mathcal{D}_\lambda^\infty(y))$$

$$\text{resp.} \quad \sup_{p \in C^\infty} \langle y_0, p \rangle, \qquad (\mathcal{D}_0^\infty(y_0))$$

$$\text{where} \quad C^\infty \stackrel{\text{def.}}{=} \{ p \in \mathcal{H} \; ; \; \|\Phi^* p\|_\infty \leqslant 1 \}. \tag{29}$$

The source condition associated with ($\mathcal{P}_0^\infty(y_0)$) amounts to the existence of some $p \in \mathcal{H}$ such that

$$\|\Phi^* p\|_\infty \leqslant 1 \quad \text{and} \quad (\Phi^* p)(x_{0,\nu}) = \operatorname{sign}(\alpha_{0,\nu}) \text{ for all } \nu \in \{1, \ldots, N\}. \tag{30}$$

Here, $\|\Phi^* p\|_\infty = \sup_{t \in \mathbb{T}} |(\Phi^* p)(t)|$. Moreover, if such p exists and satisifies $|(\Phi^* p)(t)| < 1$ for all $t \in \mathbb{T} \setminus \{x_{0,1}, \ldots, x_{0,N}\}$, and Φ_{x_0} has full rank, then m_0 is the *unique* solution to ($\mathcal{P}_0^\infty(y_0)$) (*i.e.* m_0 is identifiable).

Observe that in this infinite dimensional setting, the source condition (30) implies the optimality of m_0 for ($\mathcal{P}_0^\infty(y_0)$) but the converse is not true (see [20]).

Remark 5. A simple but crucial remark made in [9] is that if m_0 is identifiable through ($\mathcal{P}_0^\infty(y_0)$) and that $\operatorname{supp} m_0 \subset \mathcal{G}_n$, then m_0 is identifiable for ($\mathcal{P}_0^n(y_0)$). Similarly, the source condition for ($\mathcal{P}_0^\infty(y_0)$) implies the source condition for ($\mathcal{P}_0^n(y_0)$).

If we are interested in noise robustness, a stronger assumption is the *Non Degenerate Source Condition* which relies on the notion of minimal norm certificate for ($\mathcal{P}_0^\infty(y_0)$). When there is a solution to ($\mathcal{D}_0^\infty(y_0)$), the one with minimal norm, p_0^∞, determines the *minimal norm certificate* $\eta_0^\infty \stackrel{\text{def.}}{=} \Phi^* p_0^\infty$. When m_0 is a solution to ($\mathcal{P}_0^\infty(y_0)$), the *minimal norm certificate* can be characterized as

$$\eta_0^\infty = \Phi^* p_0^\infty \quad \text{where} \tag{31}$$

$$p_0^\infty = \operatorname*{argmin}_{p \in \mathcal{H}} \; \{ \|p\| \; ; \; \|\Phi^* p\|_\infty \leqslant 1, \; (\Phi^* p)(x_{0,\nu}) = \operatorname{sign}(\alpha_{0,\nu}), 1 \leqslant \nu \leqslant N \}. \tag{32}$$

As with the discrete LASSO problem, a notion of extended (signed) support $\operatorname{ext}^\pm{}_\infty$ may be defined and the minimal norm certificate governs the behavior of the solutions at low noise (see [20] for more details).

Definition 2. *Let $m_0 = \sum_{\nu=1}^{N} \alpha_{0,\nu} \delta_{x_{0,\nu}}$ an identifiable measure for ($\mathcal{P}_0^\infty(y_0)$), and $\eta_0^\infty \in \mathscr{C}(\mathbb{T})$ its minimal norm certificate. We say that m_0 satisfies the Non-Degenerate Source Condition if*

- $|\eta_0^\infty(t)| < 1$ *for all* $t \in \mathbb{T} \setminus \{x_{0,1}, \ldots x_{0,N}\}$,
- $\eta_0^{\infty\prime\prime}(x_{0,\nu}) \neq 0$ *for all* $\nu \in \{1, \ldots, N\}$.

The Non Degenerate Source Condition might seem difficult to check in practice. It turns out that it is easy to check numerically by computing the vanishing derivatives precertificate.

Definition 3. *Let* $m_0 = \sum_{\nu=1}^N \alpha_{0,\nu}\delta_{x_{0,\nu}}$ *an identifiable measure for* ($\mathcal{P}_0^\infty(y_0)$) *such that* $\Gamma_{x_0} \stackrel{\text{def.}}{=} \begin{pmatrix}\Phi_{x_0} & \Phi'_{x_0}\end{pmatrix}$ *has full rank. We define the vanishing derivatives precertificate as* $\eta_V^\infty \stackrel{\text{def.}}{=} \Phi^* p_V^\infty$ *where*

$$p_V^\infty \stackrel{\text{def.}}{=} \underset{p\in\mathcal{H}}{\operatorname{argmin}} \ \{\|p\| \ ; \ (\Phi^* p)(x_{0,\nu}) = \operatorname{sign}(\alpha_{0,\nu}), (\Phi^* p)'(x_{0,\nu}) = 0, \ 1 \leqslant \nu \leqslant N\}. \tag{33}$$

This precertificate can be easily computed by solving a linear system in the least square sense.

Proposition 3 ([20])**.** *Let* $m_0 = \sum_{\nu=1}^N \alpha_{0,\nu}\delta_{x_{0,\nu}}$ *an identifiable measure for the problem* ($\mathcal{P}_0^\infty(y_0)$) *such that* Γ_{x_0} *has full rank.*

Then, the vanishing derivatives precertificate can be computed by

$$\eta_V^\infty \stackrel{\text{def.}}{=} \Phi^* p_V^\infty \quad \text{where} \quad p_V^\infty \stackrel{\text{def.}}{=} \Gamma_{x_0}^{+,*} \begin{pmatrix}\operatorname{sign}(\alpha_{0,\cdot}) \\ 0\end{pmatrix}, \tag{34}$$

and $\Gamma_{x_0}^{+,*} = \Gamma_{x_0}(\Gamma_{x_0}^*\Gamma_{x_0})^{-1}$*. Moreover, the following conditions are equivalent:*

(i) m_0 *satisfies the Non Degenerate Source Condition.*

(ii) The vanishing derivatives precertificate satisfies:
- $|\eta_V^\infty(t)| < 1$ *for all* $t \in \mathbb{T}\setminus\{x_{0,1},\dots x_{0,N}\}$,
- ${\eta_V^\infty}''(x_{0,\nu}) \neq 0$ *for all* $\nu \in \{1,\dots,N\}$.

And in that case, η_V^∞ *is equal to the minimal norm certificate* η_0^∞*.*

Remark 6. Using the block inversion formula in (34), it is possible to check that

$$p_V^\infty = \Phi_{x_0}^{+,*}\operatorname{sign}(\alpha_{0,\cdot}) - \Pi\Phi'_{x_0}({\Phi'_{x_0}}^*\Pi\Phi'_{x_0})^{-1}{\Phi'_{x_0}}^*\Phi_{x_0}^{+,*}\operatorname{sign}(\alpha_{0,\cdot}), \tag{35}$$

where Π is the orthogonal projector onto $(\operatorname{Im}\Phi_{x_0})^\perp$. If we denote by p_F^∞ the vector introduced by Fuchs (see (21)), which turns out to be

$$p_F^\infty = \underset{p\in\mathcal{H}}{\operatorname{argmin}} \ \{\|p\| \ ; \ (\Phi^* p)(x_{0,\nu}) = \operatorname{sign}(\alpha_{0,\nu}), \ 1 \leqslant \nu \leqslant N\},$$

we observe that $p_V^\infty = p_F^\infty - \Pi\Phi'_{x_0}({\Phi'_{x_0}}^*\Pi\Phi'_{x_0})^{-1}{\Phi'_{x_0}}^* p_F^\infty$.

Remark 7. At this stage, we see that two different minimal norm certificates appear: the one for the discrete problem ($\mathcal{P}_0^n(y_0)$) which should satisfy (26) on a discrete grid $\mathcal{G}_n$, and the one for gridless problem ($\mathcal{P}_0^\infty(y_0)$) which should satisfy (30). One should not mingle them.

3.3. *The* LASSO *on Thin Grids for Fixed* $\lambda > 0$

As hinted by the notation, Problem ($\mathcal{P}_\lambda^\infty(y)$) is the limit of Problem ($\mathcal{P}_\lambda^n(y)$) as the stepsize of the grid vanishes (*i.e.* $n \to +\infty$). Indeed, we may identify each vector $a \in \mathbb{R}^{G_n}$ with the measure $m_a = \sum_{k=0}^{G_n-1} a_k \delta_{kh_n}$ (so that $\|a\|_1 = |m_a|(\mathbb{T})$) and embed ($\mathcal{P}_\lambda^n(y)$) into the space of Radon measures. With this identification, the Problem ($\mathcal{P}_\lambda^n(y)$) Γ-converges towards Problem ($\mathcal{P}_\lambda^\infty(y)$) (see the definition below), and as a result, any accumulation point of the minimizers of ($\mathcal{P}_\lambda^n(y)$) is a minimizer of ($\mathcal{P}_\lambda^\infty(y)$).

Remark 8. The space $\mathcal{M}(\mathbb{T})$ endowed with the weak* topology is a topological vector space which does not satisfy the first axiom of countability (*i.e.* the existence of a countable base of neighborhoods at each point). However, each solution m_λ^n of ($\mathcal{P}_\lambda^n(y)$) (resp. m_λ^∞ of ($\mathcal{P}_\lambda^\infty(y)$)) satisfies

$$\lambda |m_\lambda^n|(\mathbb{T}) \leqslant \lambda |m_\lambda^n|(\mathbb{T}) + \frac{1}{2}\|\Phi m_\lambda^n - y\|^2 \leqslant \frac{1}{2}\|y\|^2. \tag{36}$$

Hence we may restrict those problems to the set

$$X \stackrel{\text{def.}}{=} \left\{ m \in \mathcal{M}(\mathbb{T})\ ;\ \lambda |m|(\mathbb{T}) \leqslant \frac{1}{2}\|y\|^2 \right\}$$

which is a metrizable space for the weak* topology. As a result, we shall work with the definition of Γ-convergence in metric spaces, which is more convenient than working with the general definition [13, Definition 4.1]). For more details about Γ-convergence, we refer the reader to the monograph [13].

Definition 4. *We say that the Problem* ($\mathcal{P}_\lambda^n(y)$) Γ*-converges towards Problem* ($\mathcal{P}_\lambda^\infty(y)$) *if, for all* $m \in X$*, the following conditions hold*

- *(Liminf inequality) for any sequence of measures* $(m^n)_{n\in\mathbb{N}} \in X^{\mathbb{N}}$ *such that* $\operatorname{supp}(m^n) \subset \mathcal{G}_n$ *and that* m^n *weakly* converges towards* m*,*
$$\liminf_{n\to+\infty} \left(\lambda |m^n|(\mathbb{T}) + \frac{1}{2}\|\Phi m^n - y\|^2 \right) \geqslant \lambda |m|(\mathbb{T}) + \frac{1}{2}\|\Phi m - y\|^2.$$
- *(Limsup inequality) there exists a sequence of measures* $(m^n)_{n\in\mathbb{N}} \in X^{\mathbb{N}}$ *such that* $\operatorname{supp}(m^n) \subset \mathcal{G}_n$*,* m^n *weakly* converges towards* m *and*
$$\limsup_{n\to+\infty} \left(\lambda |m^n|(\mathbb{T}) + \frac{1}{2}\|\Phi m^n - y\|^2 \right) \leqslant \lambda |m|(\mathbb{T}) + \frac{1}{2}\|\Phi m - y\|^2.$$

The following proposition shows the Γ-convergence of the discretized problems toward the Beurling LASSO problem. This ensures in particular the convergence of the minimizers, which was already proved in [31]. Notice that this Γ-convergence can be seen as a consequence of the study in [25], where discrete vectors a are embedded in $\mathcal{M}(\mathbb{T})$ using $m_a = \sum_{k=0}^{G_n-1} a_k \mathbb{1}_{[kh_n,(k+1)h_n)}$ (as opposed to $\sum_{k=0}^{G_n-1} a_k \delta_{kh_n}$). While that other discretization yields the same convergence of the primal problems, it seems less convenient to interpret the convergence of dual certificates, so that we propose a direct proof (using our discretization) in Appendix D.1.

Proposition 4 ([25]). *The Problem* $(\mathcal{P}_\lambda^n(y))$ Γ*-converges towards* $(\mathcal{P}_\lambda^\infty(y))$*, and*

$$\lim_{n\to+\infty} \inf(\mathcal{P}_\lambda^n(y)) = \inf(\mathcal{P}_\lambda^\infty(y)). \tag{37}$$

Each sequence $(m_\lambda^n)_{n\in\mathbb{N}}$ *such that* m_λ^n *is a minimizer of* $(\mathcal{P}_\lambda^n(y))$ *has accumulation points (for the weak*) topology, and each of these accumulation point is a minimizer of* $(\mathcal{P}_\lambda^\infty(y))$*.*

In particular, if the solution m_λ to $(\mathcal{P}_\lambda^\infty(y))$ is unique, the minimizers of $(\mathcal{P}_\lambda^n(y))$ converge towards m_λ.

Here, we propose to describe the convergence of the minimizers of $(\mathcal{P}_\lambda^n(y))$ more accurately than the plain weak-* by studying the dual certificates p_λ and looking at the support of the solutions m_λ^n to $(\mathcal{P}_\lambda^n(y))$ (see [20, Section 5.4]). One may prove that m_λ^n is generally composed of at most one pair of Dirac masses in the neighborhood of each Dirac mass of the solution $m_\lambda^\infty = \sum_{\nu=1}^{N_\lambda} \alpha_{\lambda,\nu}\delta_{x_{\lambda,\nu}}$ to $(\mathcal{P}_\lambda^\infty(y))$. More precisely,

Proposition 5. *Let* $\lambda > 0$*, and assume that there exists a solution to* $(\mathcal{P}_\lambda^\infty(y))$ *which is a sum of a finite number of Dirac masses:* $m_\lambda^\infty = \sum_{\nu=1}^{N_\lambda} \alpha_{\lambda,\nu}\delta_{x_{\lambda,\nu}}$ *(where* $\alpha_\nu \neq 0$*). Assume that the corresponding dual certificate* $\eta_\lambda^\infty = \Phi^* p_\lambda^\infty$ *satisfies* $|\eta_\lambda^\infty(t)| < 1$ *for all* $t \in \mathbb{T} \setminus \{x_1, \ldots, x_N\}$*.*

Then any sequence of solution $m_\lambda^n = \sum_{i=0}^{G_n-1} a_{\lambda,i}^n \delta_{ih_n}$ *to* $(\mathcal{P}_\lambda^n(y))$ *satisfies*

$$\limsup_{n\to+\infty} (\operatorname{supp}(m_\lambda^n)) \subset \{x_1, \ldots x_N\}.$$

If, moreover, m_λ^∞ *is the unique solution to* $(\mathcal{P}_\lambda^\infty(y))$*,*

$$\lim_{n\to+\infty} (\operatorname{supp}(m_\lambda^n)) = \{x_1, \ldots x_N\}. \tag{38}$$

If, additionally, $(\eta_\lambda^\infty)''(x_\nu) \neq 0$ *for some* $\nu \in \{1, \ldots, N\}$*, then for all* n *large enough, the restriction of* m_λ^n *to* $(x_\nu - r, x_\nu + r)$ *(with* $0 < r < \frac{1}{2}\min_{\nu-\nu'} |x_{\lambda,\nu} - x_{\lambda,\nu'}|$*) is a sum of Dirac masses of the form* $a_{\lambda,i}\delta_{ih_n} + a_{\lambda,i+\varepsilon_{i,n}}\delta_{(i+\varepsilon_{i,n})h_n}$ *with* $\varepsilon_{i,n} \in \{-1, 1\}$*,* $a_{\lambda,i} \neq 0$ *and* $\operatorname{sign}(a_{\lambda,i}) = \operatorname{sign}(\alpha_{\lambda,\nu})$*. Moreover, if* $a_{\lambda,i+\varepsilon_{i,n}} \neq 0$*,* $\operatorname{sign}(a_{\lambda,i+\varepsilon_{i,n}}) = \operatorname{sign}(\alpha_{\lambda,\nu})$*.*

We skip the proof as it is very close to the arguments of [20, Section 5.4]. Moreover the proof of Proposition TODO in the companion paper [21] for the C-BP is quite similar.

3.4. Convergence of the Extended Support

Now, we focus on the study of low noise regimes. The convergence of the extended support for $(\mathcal{P}_0^n(y_0))$ towards the extended support of $(\mathcal{P}_0^\infty(y_0))$ is analyzed by the following proposition.

From now on, we assume that the source condition for $(\mathcal{P}_0^\infty(y_0))$ holds, and that $\operatorname{supp} m_0 \subset \mathcal{G}_n$ for n large enough (in other words, $y_0 = \Phi_{\mathcal{G}_n} a_0$ for some $a_0 \in \mathbb{R}^{G_n}$), so that $m_0 = \sum_{\nu=1}^N \alpha_{0,i}\delta_{x_{0,\nu}}$ is a solution of $(\mathcal{P}_0^n(y_0))$. Moreover we assume that n is large enough so that $|x_{0,\nu} - x_{0,\nu'}| > 2h_n$ for $\nu' \neq \nu$.

Proposition 6 ([20])**.** *The following result holds:*

$$\lim_{n\to+\infty} \eta_0^n = \eta_0^\infty, \tag{39}$$

in the sense of the uniform convergence (which also holds for the first and second derivatives). Moreover, if m_0 satisfies the Non Degenerate Source Condition, for n large enough, there exists $\varepsilon^n \in \{-1,0,+1\}^N$ such that

$$\mathrm{ext}^{\pm n}(m_0) = \mathrm{supp}^{\pm}(m_0) \cup \left(\mathrm{supp}^{\pm}(m_0) + \varepsilon^n h_n\right), \tag{40}$$

where $\mathrm{supp}^{\pm}(m_0) + \varepsilon^n h_n \overset{\text{def.}}{=} \{(x_{0,\nu} + \varepsilon^n_\nu h_n, \eta_0^\infty(x_{0,\nu}))\ ;\ 1 \leqslant \nu \leqslant N\}$.

That result ensures that on thin grids, there is a low noise regime for which the solutions are made of the same spikes as the original measure, plus possibly one immediate neighbor of each spike with the same sign. However, it does not predict which neighbors may appear and where (is it at the left or at the right of the original spike?).

The following new theorem, whose proof can be found in Appendix D.2, refines that result by giving a sufficient condition for the spikes to appear in pairs (*i.e.* $\varepsilon_\nu = \pm 1$ for $1 \leqslant \nu \leqslant N$). Moreover, it shows that the value of ε^n does not depend on n, and it gives the explicit positions of the added spikes ε_ν, for $1 \leqslant \nu \leqslant N$.

Theorem 2. *Assume that the operator $\Gamma_{x_0} = \begin{pmatrix}\Phi_{x_0} & \Phi'_{x_0}\end{pmatrix}$ has full rank, and that m_0 satisfies the Non-Degenerate Source Condition. Moreover, assume that all the components of the natural shift*

$$\rho \overset{\text{def.}}{=} (\Phi'^*_{x_0}\Pi\Phi'_{x_0})^{-1}\Phi'^*_{x_0}\Phi^{+,*}_{x_0}\,\mathrm{sign}(m_0(x_0)) \tag{41}$$

are nonzero, where Π is the orthogonal projector onto $(\mathrm{Im}\,\Phi_{x_0})^\perp$.

Then, for n large enough, the extended signed support of m_0 on $\mathcal{G}_n$ has the form

$$\mathrm{ext}^{\pm n}(m_0) = \{(x_\nu, \mathrm{sign}(\alpha_{0,\nu}))\}_{1\leqslant\nu\leqslant N} \cup \{(x_\nu + \varepsilon_\nu h_n, \mathrm{sign}(\alpha_{0,\nu})\}_{1\leqslant\nu\leqslant N} \tag{42}$$

$$\text{where } \varepsilon = \mathrm{sign}\left(\mathrm{diag}(\mathrm{sign}(\alpha_0))\rho\right). \tag{43}$$

In the above theorem, observe that $\Phi'^*_{x_0}\Pi\Phi'_{x_0}$ is indeed invertible since Γ_{x_0} has full rank.

Corollary 1. *Under the hypotheses of Theorem 2, for n large enough, there exists constants $C_n^{(1)} > 0$, $C_n^{(2)} > 0$ such that for $\lambda \leqslant C_n^{(1)} \min_{1\leqslant\nu\leqslant N}|\alpha_{0,\nu}|$, and for all $w \in \mathcal{H}$ such that $\|w\| \leqslant C_n^{(2)}\lambda$, the solution to ($\mathcal{P}_\lambda^n(y)$) is unique, and reads $m_\lambda = \sum_{\nu=1}^N(\alpha_{\lambda,\nu}\delta_{x_{0,\nu}} + \beta_{\lambda,\nu}\delta_{x_{0,\nu}+\varepsilon h_n})$, where*

$$\begin{pmatrix}\alpha_\lambda\\ \beta_\lambda\end{pmatrix} = \begin{pmatrix}\alpha_0\\ 0\end{pmatrix} + \Phi^+_{\mathrm{ext}_n} w - \lambda(\Phi^*_{\mathrm{ext}_n}\Phi_{\mathrm{ext}_n})^{-1}\,\mathrm{sign}\begin{pmatrix}\alpha_0\\ \alpha_0\end{pmatrix},$$

$$\begin{aligned}\text{where}\quad \mathrm{ext}_n(m_0) &= \{x_\nu\}_{1\leqslant\nu\leqslant N} \cup \{x_\nu + \varepsilon_\nu h_n\}_{1\leqslant\nu\leqslant N},\\ \varepsilon &= \mathrm{sign}\left(\mathrm{diag}(\mathrm{sign}(\alpha_0))\rho\right),\\ \mathrm{sign}(\alpha_{\lambda,\nu}) &= \mathrm{sign}(\beta_{\lambda,\nu}) = \mathrm{sign}(\alpha_{0,\nu}).\end{aligned}$$

3.5. Asymptotics of the Constants

To conclude this section, we examine the decay of the constants $C_n^{(1)}$, $C_n^{(2)}$ in Corollary 1 as $n \to +\infty$. For this we look at the values of $c_1, \ldots, c_5$ given in the proof of Theorem 1.

By Lemma 3 applied to $\Phi_{\mathrm{ext}_n(m_0)} = \begin{pmatrix}\Phi_{x_0} & \Phi_{x_0} + h_n(\Phi'_{x_0} + O(h_n))\end{pmatrix}$, we see that

$$c_{1,n} \overset{\text{def.}}{=} \|R_I \Phi^+_{\mathrm{ext}_n(m_0)}\|_{\infty,2} \sim \frac{1}{h_n}\|(\Phi'^*_{x_0}\Pi\Phi'^*_{x_0})^{-1}\Phi'^*_{x_0}\Pi\|_{\infty,2}, \tag{44}$$

$$c_{2,n} \overset{\text{def.}}{=} \|v_I\|_\infty = \left\| R_I(\Phi^*_{\mathrm{ext}_n(m_0)}\Phi_{\mathrm{ext}_n(m_0)})^{-1}\begin{pmatrix} s_I \\ s_I\end{pmatrix}\right\|_\infty \sim \frac{1}{h_n}\|\rho\|_\infty, \tag{45}$$

$$c_{3,n} \overset{\text{def.}}{=} (\|R_K \Phi^+_{\mathrm{ext}_n(m_0)}\|_{\infty,2})^{-1}\left(\min_{k\in K}|v_k|\right) \sim \frac{\min_{k\in K}|\rho_k|}{\|(\Phi'^*_{x_0}\Pi\Phi'^*_{x_0})^{-1}\Phi'^*_{x_0}\Pi\|_{\infty,2}}. \tag{46}$$

However, the expressions of c_4 and c_5 lead to an overly pessimistic bound on the signal-to-noise ratio. Indeed the majorization used in (Appendix C.2) is too rough in this framework: it does not distinguish between neighborhoods of $x_{0,\nu}$'s, where the certificate is close to 1, and the rest of the domain. The following proposition, whose proof can be found in Appendix D.3, gives a more refined asymptotic.

Proposition 7. *The constants $C_n^{(1)}, C_n^{(2)}$ in Corollary 1 can be chosen as $C_n^{(1)} = O(h_n)$ and $C_n^{(2)} = O(1)$, and one has*

$$\left\|\begin{pmatrix}\alpha_\lambda \\ \beta_\lambda\end{pmatrix} - \begin{pmatrix}\alpha_0 \\ 0\end{pmatrix}\right\|_\infty = O\left(\frac{w}{h_n}, \frac{\lambda}{h_n}\right). \tag{47}$$

4. Numerical illustrations on Compressed sensing

In this section, we illustrate the "abstract" support analysis of the LASSO problem provided in Section 2, in the context of ℓ^1 recovery for compressed sensing. Let us mention that more experiments, illustrating the doubling of the support for sparse spikes recovery on thin grids are described in the companion paper, in a comparison of the LASSO and the C-BP. Compressed sensing corresponds to the recovery of a high dimensional (but hopefully sparse) vector $a_0 \in \mathbb{R}^P$ from low resolution, possibly noisy, randomized observations $y = Ba_0 + w \in \mathbb{R}^Q$, see for instance [7] for an overview of the literature on this topic. For simplicity, we assume that there is no noise ($w = 0$) and we consider here the case where $B \in \mathbb{R}^{Q\times P}$ is a realization from the Gaussian matrix ensemble, where the entries are independent and uniformly distributed according to a Gaussian $\mathcal{N}(0,1)$ distribution. This setting is particularly well documented, and it has been shown, assuming that a_0 is s-sparse (meaning that $|\operatorname{supp}(a_0)| = s$), that there are roughly three regimes:

- If $s < s_0 \overset{\text{def.}}{=} \frac{Q}{2\log(P)}$, then a_0 is with "high probability" the unique solution of ($\mathcal{P}_0(y_0)$) (it is identifiable), and the support is stable to small noise, because η_F (as defined in (21)) is a valid certificate, $\|\eta_F\|_\infty \leqslant 1$. This is shown for instance in [35, 19].

– If $s < s_1 \overset{\text{def.}}{=} \frac{Q}{2\log(P/Q)}$, then a_0 is with "high probability" the unique solution of $(\mathcal{P}_0(y_0))$, but the support is not stable, meaning that η_F is not a valid certificate. This phenomena is precisely analyzed in [10, 1] using tools from random matrix theory and so-called Gaussian width computations.
– If $s > s_1$, then a_0 with "high probability" is not the solution of $(\mathcal{P}_0(y_0))$.

We do not want to give details here on the precise meaning of with "high probability", but this can be precisely quantified in term of probability of success (with respect to the random draw of B) and one can show that a phase transition occurs, meaning that for large (P, Q) the transition between these regimes is sharp.

While the regime $s < s_0$ is easy to understand, a precise analysis of the intermediate regime $s_0 < s < s_1$ in term of support stability is still lacking. Figure 2 shows how Theorem 1 allows us to compute numerically the size of the recovered support, hence providing a quantification of the degree of "instability" of the support when a small noise w contaminates the observations. The simulation is done with $(P, Q) = (400, 100)$.

The left part of the figure shows, as a function of s (in abscissa), the probability (with respect to a random draw of Φ and a_0 a s-sparse vector) of the event that a_0 is identifiable (plain curve) and of the event that η_F is a valid certificate (dashed curve). This clearly highlights the phase transition phenomena between the three different regimes, and one roughly gets that $s_0 \approx 6$ and $s_1 \approx 20$, which is consistent with the theoretical asymptotic bounds found in the literature.

The right part of the figure, shows, for three different sparsity levels $s \in \{14, 16, 18\}$, the histogram of the repartition of $|J|$ where J is the extended support, as defined in Theorem 1. According to Theorem 1, this histogram thus shows the repartition of the sizes of the supports of the solutions to $(\mathcal{P}_\lambda(y))$ when the noise w contaminating the observations $y = Ba_0 + w$ is small and λ is chosen in accordance to the noise level. As one could expect, this histogram is more and more concentrated around the minimum possible value s (since we are in the regime $s < s_1$ so that the support I of size s is included in the extended support J) as s approaches s_0 (for smaller values, the histogram being only concentrated at s since $J = I$ and the support is stable). Analyzing theoretically this numerical observation is an interesting avenue for future work that would help to better understand the performance of compressed sensing.

Conclusion

In this work, we have provided a precise analysis of the properties of the solution path of ℓ^1 regularization in the low-noise regime. A particular attention has been paid to the support set of this path, which in general cannot be expected to match the one of the sought after solution. Two striking examples support the relevance of this approach. For imaging problems (when the observations depend smoothly on the spikes locations), we showed theoretically that in general this support is not stable, and we were able to derive in closed form the solution of the "extended support" that is twice larger, but is stable. In the compressed sensing scenario (i.e. when the operator of the inverse problem

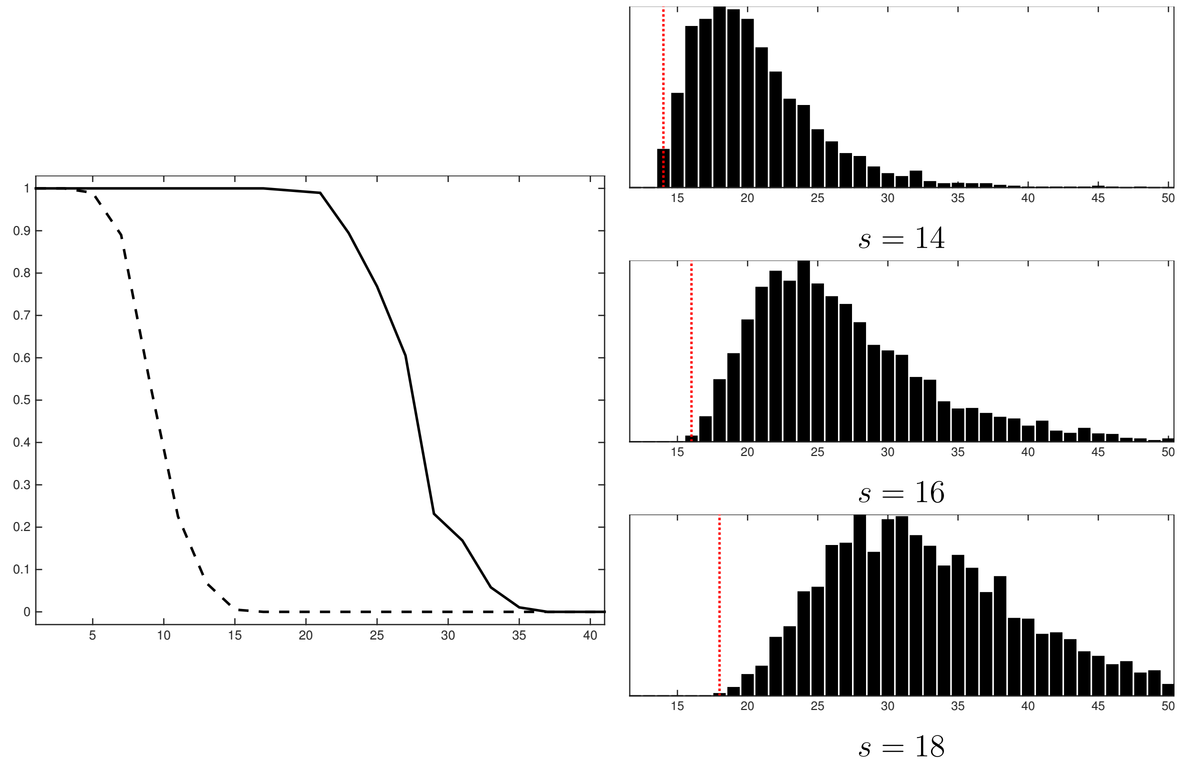


Figure 2: *Left:* probability as a function of s of the event that a_0 is identifiable (plain curve) and of the even that its support is stable (dashed curve). *Right:* for several value of s, display of histogram of repartition of the sizes $|J|$ of the extended support J.

is random), we showed numerically how to leverage our theoretical findings and analyze the growth of the extended support size as the number of measurements diminishes. This analysis opens the doors for many new developments to better understand this extended support, both for deterministic operators (e.g. Radon transform in medical imaging) and random ones.

Acknowledgements

We would like to thank Charles Dossal, Jalal Fadili and Samuel Vaiter for stimulating discussions on the notion of extended support. This work has been supported by the European Research Council (ERC project SIGMA-Vision).

Appendix A. Useful properties of the integral transform

Lemma 1. *Let $k_0 \in \mathbb{N}^*$ and assume that $\varphi \in \mathscr{C}^{k_0}(\mathbb{T}, \mathcal{H})$. Then $\Phi^{(k)} : \mathcal{M}(\mathbb{T}) \to \mathcal{H}$, $m \mapsto \int_{\mathbb{T}} \varphi^{(k)}(t)\mathrm{d}m(t)$ is weak-* to weak continuous, and its adjoint operator $\Phi^{(k),*} : \mathcal{H} \to \mathscr{C}(\mathbb{T})$ is compact and given by $(\Phi^{(k),*}q)(t) = \langle q, \varphi^{(k)}(t)\rangle$ for all $q \in \mathcal{H}$, $t \in \mathbb{T}$.*

Eventually, $\frac{\mathrm{d}^k}{\mathrm{d}t^k}(\Phi^* q)(t) = (\Phi^{(k),*}q)(t)$.

Proof. By continuity and bilinearity of the inner product, we see that

$$\forall q \in \mathcal{H}, \quad \langle q, \Phi^{(k)} m \rangle = \int_{\mathbb{T}} \langle q, \varphi^{(k)}(t) \rangle \mathrm{d}m(t).$$

Since $t \mapsto \langle q, \varphi^{(k)}(t) \rangle$ is in $\mathscr{C}(\mathbb{T})$ we obtain the weak-* to weak continuity and the expression of the adjoint operator. Its compactness, namely that $\{\Phi^* p;\ p \in \mathcal{H}, \|p\| \leqslant 1\}$ is relatively compact in $\mathscr{C}(\mathbb{T})$, follows from the Ascoli-Arzelà theorem. The last assertion is simply that $\frac{\mathrm{d}^k}{\mathrm{d}t^k} \langle q, \varphi(t) \rangle = \langle q, \frac{\mathrm{d}^k}{\mathrm{d}t^k} \varphi(t) \rangle$ □

The compactness mentioned above yields the following property. Given any bounded sequence $\{p_n\}_{n \in \mathbb{N}}$ in $\mathcal{H}$, we may extract a subsequence $\{p_{n'}\}_{n' \in \mathbb{N}}$ which converges weakly towards some $\tilde{p} \in \mathcal{H}$. Then, the (sub)sequence $\Phi^* p_{n'}$ converges towards $\Phi^* \tilde{p}$ for the (strong) uniform topology, and its derivatives $\Phi^{(k),*} p_{n'}$ also converge towards $\Phi^{(k),*} \tilde{p}$ for that topology.

Appendix B. Asymptotic expansion of the inverse of a Gram matrix

In this Appendix, we gather some useful lemmas on the asymptotic behavior of inverse Gram matrices.

Lemma 2. *Let $A \colon \mathbb{R}^N \to \mathcal{H}$, $B : \mathbb{R}^N \to \mathbb{R}^N$ be linear operators such that A has full rank and B is invertible. Then the Moore-Penrose pseudo-inverse of AB is $(AB)^+ = B^{-1} A^+$.*

Proof. Since AB has full rank, the classical formula of the pseudo-inverse yields

$$\left((AB)^*(AB)\right)^{-1}(AB)^*) = B^{-1}(A^*A)^{-1} B^{-1,*} B^* A^* = B^{-1} A^+.$$

□

Lemma 3. *Let $A, B, B_h \colon \mathbb{R}^N \to \mathcal{H}$ be linear operators such that $B_h = B + O(h)$ for $h \to 0^+$, and that $\begin{pmatrix} A & B \end{pmatrix}$ has full rank. Let Π be the orthogonal projector onto $(\operatorname{Im} A)^\perp$, and let*

$$G_h \stackrel{\text{def.}}{=} \begin{pmatrix} A^* \\ A^* + h B_h^* \end{pmatrix} \begin{pmatrix} A & A + h B_h \end{pmatrix}$$

and $s \in \mathbb{R}^N$. Then for $h > 0$ small enough, G_h and $B^* \Pi B$ are invertible, and

$$G_h^{-1} \begin{pmatrix} s \\ s \end{pmatrix} = \frac{1}{h} \begin{pmatrix} (B^* \Pi B)^{-1} B^* A^{+,*} s \\ -(B^* \Pi B)^{-1} B^* A^{+,*} s \end{pmatrix} + O(1), \tag{B.1}$$

$$\begin{pmatrix} A & A + h B_h \end{pmatrix}^+ = \frac{1}{h} \begin{pmatrix} (B^* \Pi B)^{-1} B^* \Pi \\ -(B^* \Pi B)^{-1} B^* \Pi \end{pmatrix} + O(1), \tag{B.2}$$

$$\text{but } \begin{pmatrix} A & A + h B_h \end{pmatrix}^{+,*} \begin{pmatrix} s \\ s \end{pmatrix} = A^{+,*} s - \Pi B (B^* \Pi B)^{-1} B^* A^{+,*} s + O(h). \tag{B.3}$$

Proof. Observe that $\begin{pmatrix} A & A+hB_h \end{pmatrix} = \begin{pmatrix} A & B_h \end{pmatrix} \begin{pmatrix} I_N & I_N \\ 0 & hI_N \end{pmatrix}$ so that

$$G_h = \begin{pmatrix} I_N & 0 \\ I_N & hI_N \end{pmatrix} \begin{pmatrix} A^*A & A^*B_h \\ B_h^*A & B_h^*B_h \end{pmatrix} \begin{pmatrix} I_N & I_N \\ 0 & hI_N \end{pmatrix}.$$

Since $\begin{pmatrix} A & B \end{pmatrix}$ has full rank, the middle matrix is invertible for h small enough, and

$$G_h^{-1} = \begin{pmatrix} I_N & -\frac{1}{h}I_N \\ 0 & \frac{1}{h}I_N \end{pmatrix} \begin{pmatrix} A^*A & A^*B_h \\ B_h^*A & B_h^*B_h \end{pmatrix}^{-1} \begin{pmatrix} I_N & 0 \\ -\frac{1}{h}I_N & \frac{1}{h}I_N \end{pmatrix}.$$

Writing $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \stackrel{\text{def.}}{=} \begin{pmatrix} A^*A & A^*B_h \\ B_h^*A & B_h^*B_h \end{pmatrix}$, the block inversion formula yields

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \begin{pmatrix} a^{-1} + a^{-1}bS^{-1}ca^{-1} & -a^{-1}bS^{-1} \\ -S^{-1}ca^{-1} & S^{-1} \end{pmatrix},$$

$$\text{where} \quad S \stackrel{\text{def.}}{=} d - ca^{-1}b = B_h^*B_h - B_h^*A(A^*A)^{-1}A^*B_h = B_h^*\Pi B_h$$

is indeed invertible for small h since $\begin{pmatrix} A & B \end{pmatrix}$ has full rank. Moreover, $a^{-1}bS^{-1} = A^+B_h(B_h^*\Pi B_h)^{-1}$, and $S^{-1}ca^{-1} = (B_h^*\Pi B_h)^{-1}B_h^*A^{+,*}$.

Now, we evaluate $G_h^{-1}\begin{pmatrix} s \\ s \end{pmatrix} = \begin{pmatrix} I_N & -\frac{1}{h}I_N \\ 0 & \frac{1}{h}I_N \end{pmatrix} \begin{pmatrix} a^{-1}s + a^{-1}bS^{-1}ca^{-1}s \\ -S^{-1}ca^{-1}s \end{pmatrix}$. We obtain

$$G_h^{-1}\begin{pmatrix} s \\ s \end{pmatrix} = \frac{1}{h}\begin{pmatrix} S^{-1}ca^{-1}s \\ -S^{-1}ca^{-1}s \end{pmatrix} + O(1) = \frac{1}{h}\begin{pmatrix} (B^*\Pi B)^{-1}B^*A^{+,*}s \\ -(B^*\Pi B)^{-1}B^*A^{+,*}s \end{pmatrix} + O(1).$$

Eventually, by Lemma 2, $\begin{pmatrix} A & A+hB_h \end{pmatrix}^+ = \begin{pmatrix} I_N & -\frac{1}{h}I_N \\ 0 & \frac{1}{h}I_N \end{pmatrix} \begin{pmatrix} A^*A & A^*B_h \\ B_h^*A & B_h^*B_h \end{pmatrix}^{-1} \begin{pmatrix} A^* \\ B_h^* \end{pmatrix}$. We obtain

$$\begin{pmatrix} A & A+hB_h \end{pmatrix}^+ = \begin{pmatrix} I_N & -\frac{1}{h}I_N \\ 0 & \frac{1}{h}I_N \end{pmatrix} \begin{pmatrix} A^+ - A^+B_h(B_h^*\Pi B_h)^{-1}B_h^*\Pi \\ -(B_h^*\Pi B_h)^{-1}B_h^*\Pi \end{pmatrix}$$

and we deduce

$$\begin{pmatrix} A & A+hB_h \end{pmatrix}^+ = \frac{1}{h}\begin{pmatrix} (B^*\Pi B)^{-1}B^*\Pi \\ -(B^*\Pi B)^{-1}B^*\Pi \end{pmatrix} + O(1),$$

and
$$\begin{aligned} \begin{pmatrix} A & A+hB_h \end{pmatrix}^{+,*}\begin{pmatrix} s \\ s \end{pmatrix} &= \begin{pmatrix} A^{+,*} - \Pi B_h(B_h\Pi B_h)^{-1}B_h^*A^{+,*} & \Pi B_h(B_h^*\Pi B_h)^{-1} \end{pmatrix} \\ &\qquad \begin{pmatrix} I_N & 0 \\ -\frac{1}{h}I_N & \frac{1}{h}I_N \end{pmatrix}\begin{pmatrix} s \\ s \end{pmatrix} \\ &= A^{+,*}s - \Pi B(B^*\Pi B)^{-1}B^*A^{+,*}s + O(h). \end{aligned}$$

□

Appendix C. Proofs for Section 2

Appendix C.1. Characterization of η_0

It is shown in [20] that there exists a low noise regime where the (signed) support of any solution $\tilde{a}_\lambda$ of $\mathcal{P}_\lambda(y_0+w)$ is included in $\text{ext}^\pm(a_0)$, $\text{supp}^\pm \tilde{a}_\lambda \subset \text{ext}^\pm(a_0)$. It is therefore crucial to understand precisely the behavior of η_0 and the structure of the extended (signed) support $\text{ext}^\pm(a_0)$. Before detailing the proof of Theorem 1, we thus detail in the following lemma a (new) result giving a characterization of η_0.

Lemma 4. *Let $(J, s_J) \subset [\![0,\, G-1]\!] \times \{-1, 1\}$ such that $(I, \text{sign}((a_0)_I)) \subset (J, s_J)$ and A_J has full rank. Define $v_J = (A_J^* A_J)^{-1} s_J$.*

Then (J, s_J) is the extended signed support of a_0, i.e. $(J, s_J) = \text{ext}^\pm(a_0)$, if and only if the following two conditions hold:

- *for all $j \in J \setminus I$, $v_j = 0$ or $s_j = -\text{sign}(v_j)$,*
- *$\|A_{J^c}^* A_J v_J\|_\infty < 1$.*

In that case, the minimal norm certificate is given by $\eta_0 = A^* A_J^{+,*} s_J$.

Proof. Writing the optimality conditions for (19), we see that $p \in \mathcal{H}$ is equal to p_0 if and only if $\|A^* p\|_\infty \leqslant 1$, $A_I^* p = \text{sign}(a_{0,I})$, and there exists $u_+ \in (\mathbb{R}^+)^G$ and $u_- \in (\mathbb{R}^+)^G$ such that:

$$2p + Au_+ - Au_- = 0, \tag{C.1}$$

where for $i \in I^c$, $u_{+,i}$ (resp. $u_{-,i}$) is a Lagrange multiplier for the constraint $(A^*p)_i \leqslant 1$ (resp. $(A^*p)_i \geqslant -1$) which satisfies the complementary slackness condition: $u_{+,i}((A^*p)_i - 1) = 0$ (resp $u_{-,i}((A^*p)_i + 1) = 0$), and for $i \in I$, $(u_{+,i} - u_{-,i})$ is the Lagrange multiplier for the constraint $(A^*p)_i = \text{sign}(a_0)_i$.

First, let $(J, s_J) = \text{ext}^\pm(a_0)$ (so that J determines the set of active constraints) and $p = p_0$. Using the complementary slackness condition we may reformulate (C.1) as

$$p_0 - A_J v_J = 0, \tag{C.2}$$

for some $v \in \mathbb{R}^G$, where $v_j = 0$ or $\text{sign}\, v_j = -(A^* p_0)_j$ for $j \in J \setminus I$, and $v_j = 0$ for $j \in [\![0,\, G-1]\!] \setminus J$. Inverting this relation, we obtain $v_J = (A_J^* A_J)^{-1} (\eta_0)_J$, and the stated conditions hold.

Conversely, let $(J, s_J) \subset [\![0,\, G-1]\!] \times \{-1, 1\}$ (not necessarily equal to $\text{ext}^\pm(a_0)$) such that $(I, \text{sign}((a_0)_I)) \subset (J, s_J)$ and that the conditions of the lemma hold, with $v_J = (A_J^* A_J)^{-1} s_J$. Then, setting $p = -A_J v_J$, we see that $\|A^* p\|_\infty \leqslant 1$, $A_I^* p = \text{sign}(a_0)_I$, and (C.1) holds with the complementary slackness when setting $u_{+,j} = \frac{1}{2}\max(v_j, 0)$, $u_{-,j} = \frac{1}{2}\max(-v_j, 0)$ for $j \in J$ and $u_{\pm,j} = 0$ for $j \notin J$. Then $p = p_0$ and the equivalence is proved. □

Appendix C.2. Proof of Theorem 1

We define a candidate solution $\hat{a}$ by

$$\hat{a}_J = a_{0,J} + A_J^+ w - \lambda v_J, \quad \hat{a}_{J^c} = 0 \tag{C.3}$$

and we prove that $\hat{a}$ is the unique solution to $(\mathcal{P}_\lambda(y_0 + w))$ using the optimality conditions (14) and (15).

We first exhibit a condition for $\operatorname{sign}(\hat{a}_J) = \operatorname{sign}(\eta_{0,J})$. To shorten the notation, we write $s_J \overset{\text{def.}}{=} \operatorname{sign}(\eta_{0,J})$. Since for $i \in I$, $a_{0,i} \neq 0$, the constraint $\operatorname{sign}(\hat{a}_I) = s_I$ is implied by

$$\|R_I A_J^+\|_{\infty,2}\|w\| + \|v_I\|_\infty \lambda < T, \quad \text{where} \quad T = \min_{i \in I} |a_{0,I}| > 0,$$

and $R_I : u \mapsto u_I$ is the restriction operator. As for $K = J \setminus I$, for all $k \in K$ $a_{0,k} = 0$ but we know from Lemma 4 that $\operatorname{sign}(v_k) = -s_k$. The constraint $\operatorname{sign}(\hat{a}_K) = s_K$ is thus implied by

$$\|R_K A_J^+\|_{\infty,2}\|w\| \leqslant \lambda \underbrace{\left(\min_{k \in K} |v_k|\right)}_{>0}.$$

Hence, we have $\operatorname{sign} \hat{a}_J = \operatorname{sign} \eta_{0,J} = s_J$, and by construction $\operatorname{supp}(\hat{a}) = J$ with

$$A_J^*(y - A\hat{a}) = \lambda s_J. \tag{C.4}$$

To ensure that $\hat{a}$ is the unique solution to $(\mathcal{P}_\lambda(y))$ with $y = y_0 + w$, it remains to check that

$$\|A_{J^c}^*(y - A\hat{a})\|_\infty < \lambda. \tag{C.5}$$

From (C.3) and (C.2),

$$\begin{aligned} y - A\hat{a} &= y_0 + w - Aa_0 - AA_J^+ w - \lambda A v \\ &= \left(\mathrm{I}_{\mathcal{H}} - A_J A_J^+\right) w - \lambda p_0 \\ &= P_{\ker(A_J^*)} w - \lambda p_0, \end{aligned}$$

and we see that (C.5) is implied by

$$\|A_{J^c}^* P_{\ker(A_J^*)}\|_{2,\infty}\|w\| - \lambda(1 - \|\eta_{0,J^c}\|_\infty) < 0$$

where by construction $\|\eta_{0,J^c}\|_\infty < 1$.

Putting everything together, one sees that $\hat{a}$ is the unique solution of $(\mathcal{P}_\lambda(y))$ if the following affine inequalities hold simultaneously

$$c_1\|w\| + c_2\lambda < T \quad \text{where} \quad \begin{cases} c_1 \overset{\text{def.}}{=} \|R_I A_J^+\|_{\infty,2}, \\ c_2 \overset{\text{def.}}{=} \|v_I\|_\infty, \end{cases} \tag{C.6}$$

$$\|w\| \leqslant c_3\lambda \quad \text{where} \quad c_3 \overset{\text{def.}}{=} (\|R_K A_J^+\|_{\infty,2})^{-1}\left(\min_{k \in K} |v_k|\right) > 0, \tag{C.7}$$

$$c_4\|w\| - c_5\lambda < 0 \quad \text{where} \quad \begin{cases} c_4 \overset{\text{def.}}{=} \|A_{J^c}^* P_{\ker(A_J^*)}\|_{2,\infty}, \\ c_5 \overset{\text{def.}}{=} 1 - \|\eta_{0,J^c}\|_\infty > 0. \end{cases} \tag{C.8}$$

Hence, for $\|w\| < \min(c_3, \frac{c_5}{c_4})\lambda$ and $\left(\frac{c_1 c_5}{c_4} + c_2\right)\lambda < T$, the first order optimality conditions hold.

Appendix D. Proofs for Section 3

Appendix D.1. Proof of Proposition 4

The liminf inequality of Definition 4 is a consequence of the (weak) lower semi-continuity of the total variation and the norm in $\mathcal{H}$ (since Φ is weak* to weak continuous, $\Phi_{\mathcal{G}_n} m^n - y$ weakly converges towards $\Phi m - y$):

$$\begin{aligned}\liminf_{n\to+\infty} &\left(\lambda |m^n|(\mathbb{T}) + \frac{1}{2}\|\Phi m^n - y\|^2\right)\\ &\geqslant \lambda \liminf_{n\to+\infty}(|m^n|(\mathbb{T})) + \frac{1}{2}\liminf_{n\to+\infty}\left(\|\Phi m^n - y\|^2\right)\\ &\geqslant \lambda|m|(\mathbb{T}) + \frac{1}{2}\|\Phi m - y\|^2.\end{aligned}$$

As for the limsup inequality, we approximate m with the measure $m^n = \sum_{k=0}^{G_n-1} b_k \delta_{kh_n}$, where $b_k = m([kh_n, (k+1)h_n))$. Then, for any $\psi \in C(\mathbb{T})$,

$$\begin{aligned}\left|\int_{\mathbb{T}} \psi \mathrm{d}m - \int_{\mathbb{T}} \psi \mathrm{d}m^n\right| &= \left|\sum_{k=0}^{G_n-1}\int_{[kh_n,(k+1)h_n)} (\psi(x) - \psi(kh_n))\mathrm{d}m\right|\\ &\leqslant \omega_\psi(h_n)|m|(\mathbb{T}),\end{aligned}$$

where $\omega_\psi : t \mapsto \sup_{|x'-x|\leqslant t} |\psi(x) - \psi(x')|$ is the modulus of continuity of ψ. Therefore, $\lim_{n\to+\infty}\langle m^n, \psi\rangle = \langle m, \psi\rangle$, and m^n weakly* converges towards m. Incidentally, observe that $|m^n|(\mathbb{T}) \leqslant |m|(\mathbb{T})$, so that using the liminf inequality we get $\lim_{n\to+\infty} |m^n|(\mathbb{T}) = |m|(\mathbb{T})$. Moreover, by similar majorizations,

$$\|\Phi m^n - \Phi m\| = \left\|\int_{\mathbb{T}} \varphi \mathrm{d}m - \int_{\mathbb{T}} \varphi \mathrm{d}m^n\right\| \leqslant \omega_\varphi(h_n)|m|(\mathbb{T}),$$

so that Φm^n converges strongly in $L^2(\mathbb{T})$ towards Φm. As a result $\lim_{n\to+\infty}\|\Phi m^n - y\|^2 = \|\Phi m - y\|^2$, and the limsup inequality is proved.

Eventually, from (36) we deduce the compactness of X, hence the existence of accumulation points, and [13, Theorem 7.8] implies that accumulation points of $(m_\lambda^n)_{n\in\mathbb{N}}$ are minimizers of $(\mathcal{P}_\lambda^\infty(y))$, as well as (37).

Appendix D.2. Proof of Theorem 2

We define a good candidate for η_0^n and using Lemma 4 we prove that it is indeed equal to η_0^n when the grid is thin enough.

To comply with the notations of Section 2, we write

$$\sum_{\nu=1}^{N} \alpha_{0,i}\delta_{x_{0,\nu}} = \sum_{k=0}^{G_n-1} a_{0,k}\delta_{kh_n},$$

and we let $I \stackrel{\text{def.}}{=} \{i \in [\![0,\, G_n - 1]\!] \;;\; a_{0,i} \neq 0\}$. Moreover, for any choice of sign $(\varepsilon_i)_{i\in I} \in \{-1,+1\}^N$, we set $J \stackrel{\text{def.}}{=} \bigcup_{i\in I}\{i, i+\varepsilon_i\}$ and $s_J = (s_j)_{j\in J}$ where $s_i \stackrel{\text{def.}}{=} s_{i+\varepsilon_i} \stackrel{\text{def.}}{=} \operatorname{sign}(a_{0,i})$ for $i \in I$. Since $|x_{0,\nu} - x_{0,\nu'}| > 2h_n$ for $\nu' \neq \nu$, we have $\operatorname{Card} J = 2 \times \operatorname{Card} I = 2N$.

Recalling that $A = \begin{pmatrix}\varphi(0) & \dots & \varphi((G_n-1)h_n)\end{pmatrix}$, we consider the submatrices

$$A_I \stackrel{\text{def.}}{=} \begin{pmatrix}\varphi(ih_n)\end{pmatrix}_{i\in I} = \begin{pmatrix}\varphi(x_{0,1}) & \dots & \varphi(x_{0,N})\end{pmatrix} \quad \text{and} \quad A_{J\setminus I} \quad \stackrel{\text{def.}}{=} \begin{pmatrix}\varphi((i+\varepsilon_i)h_n)\end{pmatrix}_{i\in I}$$

so that up to a reordering of the columns $A_J = \begin{pmatrix}A_I & A_{J\setminus I}\end{pmatrix}$. In order to apply Lemma 4, we shall exhibit a choice of $(\varepsilon_i)_{i\in I}$ such that A_J has full rank, that $v \stackrel{\text{def.}}{=} (A_J^* A_J)^{-1} s_J$ satisfies $\operatorname{sign}(v_j) = -s_j$ for $j \in J \setminus I$ and $\|A_{J^c}^* A_J v\|_\infty < 1$.

The following Taylor expansion holds for $A_{J\setminus I}$ as $n \to \infty$:

$$\begin{aligned} A_{J\setminus I} &= A_0 + h_n(B_0 + O(h_n)), \qquad \text{with } A_0 = A_I = \Phi_{x_0}\\ \text{and } B_0 &= \begin{pmatrix}\varphi'(x_{0,1}) & \dots & \varphi'(x_{0,N})\end{pmatrix} \operatorname{diag}\left((\varepsilon_{i_1}), \dots (\varepsilon_{i_N})\right)\\ &= \Phi'_{x_0} \operatorname{diag}\left((\varepsilon_{i_1}), \dots (\varepsilon_{i_N})\right). \end{aligned}$$

By Lemma 3 in Appendix, the Gram matrix $A_J^* A_J$ is invertible for n large enough, and

$$(A_J^* A_J)^{-1} \begin{pmatrix} s_I \\ s_I \end{pmatrix} = \frac{1}{h_n} \begin{pmatrix} (\operatorname{diag}(\varepsilon_{i_1}, \dots, \varepsilon_{i_N}))^{-1}\rho \\ -(\operatorname{diag}(\varepsilon_{i_1}, \dots, \varepsilon_{i_N}))^{-1}\rho \end{pmatrix} + O(1),$$

where ρ is defined in (41), where Π is the orthogonal projector onto $(\operatorname{Im}\Phi_{x_0})^\perp$, and for $\nu \in [\![1,\, N]\!]$, i_ν refers to the index $i \in I$ such that $ih_n = x_{0,\nu}$. Therefore, $v_{J\setminus I}$ has the sign of $-\operatorname{diag}(\varepsilon_{i_1}, \dots \varepsilon_{i_N})\rho$, and it is sufficient to choose $\varepsilon_{i_\nu} = s_{i_\nu} \times \operatorname{sign}(\rho_\nu)$ to ensure that $\operatorname{sign} v_{J\setminus I} = -s_{J\setminus I}$ for n large enough.

With that choice of ε, it remains to prove that $\|A_{J^c}^* A_J v\|_\infty < 1$. Let us write $\tilde{p}_n \stackrel{\text{def.}}{=} A_J v = A_J^{+,*} \begin{pmatrix} s_I \\ s_I \end{pmatrix}$. It is equivalent to prove that for $k \in J^c$, $|\Phi^* \tilde{p}_n(kh_n)| < 1$. Using the above Taylor expansion and Lemma 3 in Appendix, we obtain that

$$\begin{aligned} \lim_{n\to+\infty} \tilde{p}_n &= A_0^{+,*} s_I - \Pi B_0 (B_0^* \Pi B_0)^{-1} B_0^* A_0^{+,*} s_I\\ &= \Phi_{x_0}^{+,*} \operatorname{sign}(\alpha_{0,\cdot}) - \Pi \Phi'_{x_0} ({\Phi'_{x_0}}^* \Pi \Phi'_{x_0})^{-1} {\Phi'_{x_0}}^* \Phi_{x_0}^{+,*} \operatorname{sign}(\alpha_{0,\cdot})\\ &= p_V^\infty \text{ (by (35))}. \end{aligned}$$

Hence, $\Phi^* \tilde{p}_n$ and its derivatives converge to those of $\eta_V^\infty = \eta_0^\infty$, and there exists $r > 0$ such that for all n large enough, for all $1 \leqslant \nu \leqslant N$, $\Phi^* \tilde{p}_n$ is strictly concave (or stricly convex, depending on the sign of $\eta_0^{\infty\prime\prime}(x_{0,\nu})$) in $(x_{0,\nu} - r, x_{0,\nu} + r)$. Hence, for $t \in (x_{0,\nu} - r, x_{0,\nu} + r) \setminus [x_{0,\nu}, x_{0,\nu} + \varepsilon_{i(\nu)} h_n]$, we have $|\Phi^* \tilde{p}_n(t)| < 1$. Since by compactness

$$\max\left\{ |\eta_0^\infty(t)| \;;\; t \in \mathbb{T} \setminus \bigcup_{\nu=1}^N (x_{0,\nu} - r, x_{0,\nu} + r) \right\} < 1$$

we also see that for n large enough

$$\max\left\{|\Phi^*\tilde{p}_n(t)|\ ;\ t\in\mathbb{T}\setminus\bigcup_{\nu=1}^{N}(x_{0,\nu}-r,x_{0,\nu}+r)\right\}<1.$$

As a consequence, for $k\in J^c$, $|\Phi^*\tilde{p}_n(kh_n)|<1$, and from Lemma 4, we obtain that $\Phi^*\tilde{p}_n=\eta_0^n$ and $\bigcup_{\nu=1}^N\{x_{0,\nu},x_{0,\nu}+\varepsilon_{i(\nu)}h_n\}$ is the extended support on $\mathcal{G}_n$.

Appendix D.3. Proof of Proposition 7

The proof of (47) follows from applying (44) and (45) in the expression for α_λ and β_λ provided by Corollary 1. Let $\omega\overset{\text{def.}}{=}\Phi^*\Pi w$, where Π is the orthogonal projector onto $(\operatorname{Im}\Phi_{x_0})^\perp=\ker\Phi_{x_0}^*$. In order to ensure (C.5) we may ensure that :

$$|\omega(jh_n)+\lambda\eta_0^n(jh_n)|-\lambda<0, \tag{D.1}$$

for all $j\in J^c$ (that is $(jh_n\notin\operatorname{ext}_n(m_0))$.

By the Non-Degenerate Source Condition, there exists $r>0$ such that for all $\nu\in\{1,\ldots,N\}$,

$$\forall t\in(x_{0,\nu}-r,x_{0,\nu}+r),\ |\eta_0^\infty(t)|>0.95\quad\text{and}\quad|(\eta_0^\infty)''(t)|>\frac{3}{4}|(\eta_0^\infty)''(x_{0,\nu})|,$$

and by compactness $\sup_{\mathbb{T}\setminus\bigcup_{\nu=1}^N(x_{0,\nu}-r,x_{0,\nu}+r)}|\eta_0^\infty|<1$. Since $\eta_0^n\to\eta_0^\infty$ (with uniform convergence of all the derivatives), for n large enough,

$$\forall\nu\in\{1,\ldots,N\},\ \forall t\in(x_{0,\nu}-r,x_{0,\nu}+r),\ |\eta_0^n(t)|>0.9\quad\text{and}\quad|(\eta_0^n)''(t)|>\frac{1}{2}|(\eta_0^\infty)''(x_{0,\nu})|,$$

(with equality of the signs) and

$$\sup_{\mathbb{T}\setminus\bigcup_{\nu=1}^N(x_{0,\nu}-r,x_{0,\nu}+r)}|\eta_0^n|\leqslant k\overset{\text{def.}}{=}\frac{1}{2}\left(\sup_{\mathbb{T}\setminus\bigcup_{\nu=1}^N(x_{0,\nu}-r,x_{0,\nu}+r)}|\eta_0^\infty|+1\right)<1.$$

First, for j such that $jh_n\in\mathbb{T}\setminus\bigcup_{\nu=1}^N(x_{0,\nu}-r,x_{0,\nu}+r)$, we see that it is sufficient to assume $\|\Phi^*\|_{\infty,2}\|w\|_2<(1-k)\lambda$ to obtain (D.1).

Now, let $\nu\in\{1,\ldots,N\}$ and assume that $\eta_0^\infty(x_{0,\nu})=1$ (so that $(\eta_0^\infty)''(x_{0,\nu})<0$) and that $\varepsilon_\nu=1$, the other cases being similar. We make the following observation: if a function $f\colon(-r,+r)\to\mathbb{R}$ satisfies $f''(t)\leqslant C$ for some $C<0$ and $f(0)=f(h_n)=0$, then $f(t)\leqslant\frac{C}{2}t(t-h_n)<0$ for $t\in(-r,0]\cup[h_n,r)$.

Notice that $\omega=\Phi^*\Pi w$ is a $\mathscr{C}^2$ function which vanishes on $\operatorname{ext}_n(m_0)$ (hence at $x_{0,\nu}$ and $x_{0,\nu}+h_n$), and that its second derivative is bounded by $\|(\Phi'')^*\|_{\infty,2}\|w\|$. Moreover, $\eta_0^n(x_{0,\nu})=\eta_0^n(x_{0,\nu}+h_n)=1$ and $\sup_{(x_{0,\nu}-r,x_{0,\nu}+r)}(\eta_0^n)''\leqslant\frac{1}{2}(\eta_0^\infty)''(x_{0,\nu})<0$. Thus, for $\frac{\|w\|}{\lambda}<\frac{|(\eta_0^\infty)''(x_{0,\nu})|}{2\|(\Phi'')^*\|_{\infty,2}}$, we may apply the observation to $\omega(\cdot-x_{0,\nu})+\lambda(\eta_0^n(\cdot-x_{0,\nu})-1)$ so as to get

$$\omega(t)+\lambda(\eta_0^n(t)-1)\leqslant\left(\|(\Phi'')^*\|_{\infty,2}\|w\|+\lambda\frac{1}{2}(\eta_0^\infty)''(x_{0,\nu})\right)(t-x_{0,\nu})(t-x_{0,\nu}-h_n)<0$$

for $t \in (x_{0,\nu} - r, x_{0,\nu}] \cup [x_{0,\nu} + h_n, x_{0,\nu} + r)$.

On the other hand, the inequality $-\omega(t) - \lambda(\eta_0^n(t) + 1) < 0$ holds for $\|\Phi^*\|_{\infty,2}\|w\|_2 < 1.9\lambda$. As a result (D.1) holds for all j such that $jh_n \in (x_{0,\nu} - r, x_{0,\nu} + r)$, provided that the signal-to-noise ratio satisfies $\frac{\|w\|_2}{\lambda} \leqslant c$, where $c > 0$ is a constant which only depends on $\min_\nu |(\eta_0^\infty)''(x_{0,\nu})|$, $\|\Phi^*\|_{\infty,2}$, $\|(\Phi'')^*\|_{\infty,2}$ and $\sup_{\mathbb{T}\setminus\bigcup_{\nu=1}^N (x_{0,\nu}-r, x_{0,\nu}+r)} |\eta_0^\infty|$. In other words, including the condition involving $c_{3,n}$, we may choose $C_n^{(2)} = \min(c_{3,n}, c) = O(1)$.

References

[1] D. Amelunxen et al. "Living on the edge: A geometric theory of phase transitions in convex optimization". In: *arXiv preprint arXiv:1303.6672* (2013).

[2] J-M. Azais, Y. De Castro, and F. Gamboa. *Spike detection from inaccurate samplings*. Tech. rep. Preprint hal-00780808, Jan. 2013.

[3] 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*. 2011, pp. 261–268.

[4] T. Blu et al. "Sparse Sampling of Signal Innovations". In: *IEEE Signal Processing Magazine* 25.2 (2008), pp. 31–40.

[5] K. Bredies and H.K. Pikkarainen. "Inverse problems in spaces of measures". In: *ESAIM: Control, Optimisation and Calculus of Variations* 19.1 (2013), pp. 190–218.

[6] M. Burger and S. Osher. "Convergence rates of convex variational regularization". In: *Inverse Problems* 20.5 (2004), pp. 1411–1421.

[7] E. Candes and M. Wakin. "An Introduction to Compressive Sensing". In: *IEEE Signal Processing Magazine* 25.2 (2008), pp. 21–30.

[8] E. J. Candès and C. Fernandez-Granda. "Super-Resolution from Noisy Data". In: *Journal of Fourier Analysis and Applications* 19.6 (2013), pp. 1229–1254.

[9] E. J. Candès and C. Fernandez-Granda. "Towards a Mathematical Theory of Super-Resolution". In: *Communications on Pure and Applied Mathematics* 67.6 (2013), pp. 906–956.

[10] V. Chandrasekaran et al. "The Convex Geometry of Linear Inverse Problems". In: *Foundations of Computational Mathematics* 12.6 (2012), pp. 805–849.

[11] S.S. Chen, D.L. Donoho, and M.A. Saunders. "Atomic decomposition by basis pursuit". In: *SIAM journal on scientific computing* 20.1 (1999), pp. 33–61.

[12] J. F. Claerbout and F. Muir. "Robust modeling with erratic data". In: *Geophysics* 38.5 (1973), pp. 826–844.

[13] G. Dal Maso. *An Introduction to Γ-convergence*. Vol. 8. Progress in nonlinear differential equations and their applications. Boston, MA: Birkhauser, 1993.

[14] Y. de Castro and F. Gamboa. “Exact reconstruction using Beurling minimal extrapolation”. In: *Journal of Mathematical Analysis and Applications* 395.1 (2012), pp. 336–354.

[15] Q. Denoyelle, V. Duval, and G. Peyré. “Support Recovery for Sparse Deconvolution of Positive Measures”. In: 1506.08264 (2015). URL: http://arxiv.org/abs/1506.08264.

[16] D. L. Donoho. “Superresolution via sparsity constraints”. In: *SIAM J. Math. Anal.* 23.5 (1992), pp. 1309–1331.

[17] C. Dossal. “A necessary and sufficient condition for exact recovery by ℓ^1 minimization”. In: *Comptes Rendus Mathématique* 350.1 (2011), pp. 117–120.

[18] C. Dossal and S. Mallat. “Sparse spike deconvolution with minimum scale”. In: *Proceedings of SPARS*. Nov. 2005, pp. 123–126.

[19] C. Dossal et al. “Sharp Support Recovery from Noisy Random Measurements by L1 minimization”. In: *Applied and Computational Harmonic Analysis* 33.1 (2012), pp. 24–43.

[20] V. Duval and G. Peyré. “Exact Support Recovery for Sparse Spikes Deconvolution”. In: *to appear in Foundation of Computational Mathematics* (2015).

[21] V. Duval and G. Peyré. *Sparse Spikes Retrieval on Thin Grids II: the Continuous Basis Pursuit*. Tech. rep. 2016.

[22] C. Fernandez-Granda. “Support detection in super-resolution”. In: *Proc. Proceedings of the 10th International Conference on Sampling Theory and Applications* (2013), pp. 145–148.

[23] J.J. Fuchs. “On sparse representations in arbitrary redundant bases”. In: *IEEE Transactions on Information Theory* 50.6 (2004), pp. 1341–1344.

[24] M. Grasmair, O. Scherzer, and M. Haltmeier. “Necessary and sufficient conditions for linear convergence of ℓ_1-regularization”. In: *Communications on Pure and Applied Mathematics* 64.2 (2011), pp. 161–182.

[25] P. Heins. In: *PhD thesis* (2014).

[26] J. Min et al. “FALCON: Fast and Unbiased Reconstruction of High-Density Super-Resolution Microscopy Data”. In: *Scientific Reports* 4577 (2014), pp. 1–9.

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

[28] R. T. Rockafellar, R. J-B. Wets, and M. Wets. *Variational analysis*. Grundlehren der mathematischen Wissenschaften. Berlin, Heidelberg, New York: Springer, 1998.

[29] G. Still. “Discretization in semi-infinite programming: the rate of convergence”. In: *Mathematical Programming* 91.1 (2001), pp. 53–69.

[30] G. Tang, B. Narayan Bhaskar, and B. Recht. “Near Minimax Line Spectral Estimation”. In: *CoRR* abs/1303.4348 (2013).

[31] G. Tang, B. Narayan Bhaskar, and B. Recht. “Sparse recovery over continuous dictionaries-just discretize”. In: *Proc. ACSSC*. IEEE, 2013, pp. 1043–1047.

[32] R. Tibshirani. “Regression shrinkage and selection via the Lasso”. In: *Journal of the Royal Statistical Society. Series B. Methodological* 58.1 (1996), pp. 267–288.

[33] S. Vaiter, G. Peyré, and J. Fadili. *Model Consistency of Partly Smooth Regularizers*. Tech. rep. Preprint Hal-00987293, 2014.

[34] S. Vaiter et al. “Model Selection with Low Complexity Priors”. In: *to appear in Information and Inference* (2015).

[35] M. J. Wainwright. “Sharp Thresholds for High-Dimensional and Noisy Sparsity Recovery Using ℓ_1-Constrained Quadratic Programming (Lasso)”. In: *IEEE Transactions on Information Theory* 55.5 (2009), pp. 2183–2202.

[36] P. Zhao and B. Yu. “On Model Selection Consistency of Lasso”. In: *J. Mach. Learn. Res.* 7 (Dec. 2006), pp. 2541–2563.