



**HAL**  
open science

## Sparse Spikes Deconvolution on Thin Grids

Vincent Duval, Gabriel Peyré

► **To cite this version:**

| Vincent Duval, Gabriel Peyré. Sparse Spikes Deconvolution on Thin Grids. 2015. hal-01135200v1

**HAL Id: hal-01135200**

**<https://hal.science/hal-01135200v1>**

Preprint submitted on 24 Mar 2015 (v1), last revised 31 Oct 2016 (v2)

**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.

# Sparse Spikes Deconvolution on Thin Grids

Vincent Duval<sup>1</sup>, Gabriel Peyré<sup>2</sup>

<sup>1</sup> INRIA, MOKAPLAN

E-mail: [vincent.duval@inria.fr](mailto:vincent.duval@inria.fr)

<sup>2</sup> CNRS and CEREMADE, Université Paris Dauphine

E-mail: [gabriel.peyre@ceremade.dauphine.fr](mailto:gabriel.peyre@ceremade.dauphine.fr)

**Abstract.** This article analyzes the recovery performance of two popular finite dimensional approximations of the sparse spikes deconvolution problem over Radon measures. We examine in a unified framework both the  $\ell^1$  regularization (often referred to as LASSO or Basis-Pursuit) and the Continuous Basis-Pursuit (C-BP) methods. The LASSO is the de-facto standard for the sparse regularization of inverse problems in imaging. It performs a nearest neighbor interpolation of the spikes locations on the sampling grid. The C-BP method, introduced by Ekanadham, Tranchina and Simoncelli, uses a linear interpolation of the locations to perform a better approximation of the infinite-dimensional optimization problem, for positive measures. We show that, in the small noise regime, both methods estimate twice the number of spikes as the number of original spikes. Indeed, we show that they both detect two neighboring spikes around the locations of an original spikes. These results for deconvolution problems are based on an abstract analysis of the so-called extended support of the solutions of  $\ell^1$ -type problems (including as special cases the LASSO and C-BP for deconvolution), which are of an independent interest. They precisely characterize the support of the solutions when the noise is small and the regularization parameter is selected accordingly. We illustrate these findings to analyze for the first time the support instability of compressed sensing recovery when the number of measurements is below the critical limit (well documented in the literature) where the support is provably stable.

## 1. Introduction

We consider the problem of estimating an unknown Radon measure  $m_0 \in \mathcal{M}(\mathbb{T})$  from low-resolution noisy observations

$$y = \Phi(m_0) + w \in L^2(\mathbb{T}) \tag{1}$$

where  $w \in L^2(\mathbb{T})$  is some measurement noise, and  $\Phi : \mathcal{M}(\mathbb{T}) \rightarrow L^2(\mathbb{T})$  is an integral transform with smooth kernel  $\varphi \in C^2(\mathbb{T} \times \mathbb{T})$ , i.e.

$$\forall x \in \mathbb{T}, \quad (\Phi m)(x) = \int_{\mathbb{T}} \varphi(x, y) dm(y). \tag{2}$$

A typical example of such an operation is the convolution, where  $\varphi(x, y) = \tilde{\varphi}(x - y)$  for some smooth function  $\tilde{\varphi}$  defined on the torus  $\mathbb{T} = \mathbb{R}/\mathbb{Z}$  (i.e. an interval with periodic boundary conditions). We focus our attention here for simplicity on the compact 1-D domain  $\mathbb{T}$ , but the algorithms considered (LASSO and C-BP) as well as our theoretical analysis can be extended to higher dimensional settings (see Section 1.4).

### 1.1. Sparse Regularization

The problem of inverting (1) is severely ill-posed. A particular example is when  $\Phi$  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 [13]. Sparse regularization has later been studied by David Donoho and co-workers, see for instance [17].

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

$$\min_{m \in \mathcal{M}(\mathbb{T})} \frac{1}{2} \|y - \Phi(m)\|^2 + \lambda |m|(\mathbb{T}) \quad (3)$$

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

$$|m|(\mathbb{T}) \stackrel{\text{def.}}{=} \sup \left\{ \int_{\mathbb{T}} \psi(x) dm(x) ; \psi \in C(\mathbb{T}), \|\psi\|_{\infty} \leq 1 \right\}. \quad (4)$$

This formulation of the recovery of sparse Radon measures has recently received lots of attention in the literature, see for instance the works of [6, 15, 10].

### 1.2. LASSO

The optimization problem (3) is convex but infinite dimensional, and while there exists solvers when  $\Phi$  is measuring a finite number of Fourier frequency (see [10]), they do not scale well with the number of frequencies. Furthermore, the case of an arbitrary linear operator  $\Phi$  is still difficult to handle, see [6] for an iterative scheme. The vast majority of practitioners thus approximate (3) by a finite dimensional problem computed over a finite grid  $\mathcal{G} \stackrel{\text{def.}}{=} \{z_i ; i \in \llbracket 0, P - 1 \rrbracket\} \subset \mathbb{T}$ , by restricting their attention to measures of the form

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

For such a discrete measure, one has  $|m|(\mathbb{T}) = \sum_{i=0}^{|\mathcal{G}|-1} |a_i| = \|a\|_1$ , which can be interpreted as the fact that  $|\cdot|(\mathbb{T})$  is the natural extension of the  $\ell^1$  norm from finite dimensional vectors to the infinite dimensional space of measures. Inserting this

parametrization in (3) leads to the celebrated Basis-Pursuit problem [12], which is also known as the LASSO method in statistics [33],

$$\min_{a \in \mathbb{R}^N} \frac{1}{2} \|y - \Phi_{\mathcal{G}} a\|^2 + \lambda \|a\|_1 \quad (5)$$

where in the following we make use of the notations

$$\Phi_{\mathcal{G}} a \stackrel{\text{def.}}{=} \Phi(m_{a,\mathcal{G}}) = \sum_{i=0}^{P-1} a_i \varphi(\cdot, z_i), \quad (6)$$

$$\Phi'_{\mathcal{G}} b \stackrel{\text{def.}}{=} \Phi'(m_{b,\mathcal{G}}) = \sum_{i=0}^{P-1} b_i \partial_2 \varphi(\cdot, z_i), \quad (7)$$

and  $\partial_2 \varphi$  is the derivative with respect to the second variable. One can understand (5) as performing a nearest neighbor interpolation of the Dirac's locations.

This approximation is however quite crude, and we recently showed in [21] that it leads to imperfect estimation of both the number of spikes and their locations. Indeed, this problem typically recovers up to twice as many spikes as the input measures, because spikes of  $m_0$  gets duplicated as the two nearest neighbors on the grid  $\mathcal{G}$ .

Note that while we focus in this paper on convex recovery method, and in particular  $\ell^1$ -type regularization, there is a vast literature on the subject, which makes use of alternative algorithms, see for instance [28, 5] and the references therein.

### 1.3. Continuous Basis-Pursuit (C-BP)

To obtain a better approximation of the infinite dimensional problem, [23] proposes to perform a first order approximation of the kernel. This method assumes that the unknown measure is positive. To ease the exposition, we consider a uniform grid  $\mathcal{G} \stackrel{\text{def.}}{=} \{i/P ; i \in \llbracket 0, P-1 \rrbracket\}$  of  $P$  points, so that the grid size is  $h \stackrel{\text{def.}}{=} 1/P$ . The C-BP method of [23] solves

$$\min_{(a,b) \in \mathbb{R}^P \times \mathbb{R}^P} \frac{1}{2} \|y - \Phi_{\mathcal{G}} a - \Phi'_{\mathcal{G}} b\|^2 + \lambda \|a\|_1 \quad \text{subject to} \quad |b| \leq \frac{h}{2} a, \quad (8)$$

where the inequality should be understood component-wise. Note also that the obtained  $a$  is always nonnegative, hence the C-BP method is tailored for the recovery of positive measures. This is a convex optimization problem, which can be solved using traditional conic optimization methods. As detailed in Section 4.2, this problem can also be re-cast as a LASSO in dimension  $2P$  with positivity constraints (see Section 4.2). Hence it can be solved using a large variety of first order proximal method, the most simple one being the Forward-Backward, see [3] and the references therein.

If  $(a^*, b^*)$  are solutions of (8), one recovers an output discrete measure defined by

$$m^* = \sum_{a_i^* \neq 0} a_i^* \delta_{x_i^*} \quad \text{where} \quad x_i^* \stackrel{\text{def.}}{=} ih + \frac{b_i^*}{a_i^*}, \quad (9)$$

where we set  $\frac{b_i^*}{a_i^*} = 0$  whenever  $a_i^* = 0$ . The rationale behind (8) is to perform a first order Taylor approximation of the operator  $\Phi$ , where the variable  $\tau_i \stackrel{\text{def.}}{=} b_i/a_i \in [-h/2, h/2]$  encodes the horizontal shift of the Dirac location with respect to the grid sample  $ih$ . The landmark idea introduced in [23] is that, while the optimization is non-convex with respect to the pair  $(a, \tau)$ , it is convex with respect to the pair  $(a, b)$ .

#### 1.4. Extensions

While we restrict here the exposition to 1-D problems, the C-BP formulation (8) can be extended to cope with measures in arbitrary dimension  $d \geq 1$ , i.e. to consider  $m_0 \in \mathcal{M}(\mathbb{T}^d)$ . This requires to define at each sampling grid point indexed by  $i$  a vector  $b_i = (b_{i,k})_{k=1}^d \in \mathbb{R}^d$  together with the constraint  $\|b_i\|_\infty \leq \frac{h}{2} a_i$ , and also to use a matrix  $\Phi'_G$  defined as

$$\Phi'_G b \stackrel{\text{def.}}{=} \sum_{i \in \mathcal{G}} \sum_{k=1}^d b_{i,k} \partial_k \varphi(\cdot, x_i) \in L^2(\mathbb{T}^d)$$

where  $\partial^k$  denote the differential operator with respect to the  $k^{\text{th}}$  direction in  $\mathbb{R}^d$ . Our analysis carries over to this setting without major difficulties.

The paper [23] also proposes other interpolation schemes than a first order Taylor expansion at the grid points. In particular, they develop a ‘‘polar’’ interpolation which makes use of two adjacent grid points. This method seems to outperform the linear interpolation in practice, and has been employed to perform spikes sorting in neuronal recordings [22].

Extending the results we propose in the present paper to these higher dimensional settings and alternative interpolation schemes is an interesting avenue for future work.

Let us also mention that an important problem is to extend the C-BP method (9) to measures with arbitrary signs and that can even be complex-valued. Unfortunately, the corresponding constraint  $|b| \leq |a|$  is then non-convex, which makes the mathematical analysis apparently much more involved. A non-convex and non-smooth optimization solver is proposed for this problem in [25], and shows promising practical performance for spectrum estimation.

#### 1.5. 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 (5). 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 [17] study the Lipschitz behavior of the inverse map  $y \mapsto a^*$ , where  $a^*$  is a solution of (5), 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 [19]. This work uses the analysis of  $\ell^1$  regularization introduced by Fuchs in [26]. 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 (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  $\ell^1$  methods can be found in the work of Dossal [18] who focusses on noiseless recovery and stability in term of  $\ell^2$  error.

Recently, a few works have studied the theoretical properties of the recovery over measures (3). Candès and Fernandez-Granda show in [10] that this convex program does recover exactly the initial sparse measure when  $w = 0$  and  $\lambda \rightarrow 0$ , if the spikes are well-separated. The robustness to noisy measurements is analyzed by the same authors in [9] using an Hilbertian norm, and in [24, 2] in terms of spikes localization. The work of [31] analyzes the reconstruction error. Lastly, [21] provides a condition ensuring that (3) 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.

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. 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 [32], where a speed of convergence is shown using tools from semi-infinite programming [30]. The same authors show in [4] that the discretized problem achieves a similar prediction  $L^2$  error as the grid-free method. In [21], 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.

### 1.6. Contributions

Our first contribution is an improvement over the known analysis of the LASSO in an abstract setting (that is (5) when  $\Phi_{\mathcal{G}}$  is replaced with any linear operator  $\mathbb{R}^P \rightarrow L^2(\mathbb{R})$ ). Whereas Fuchs’ result [26] characterizes the support recovery of the LASSO at low noise, our previous work [21] 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*. In this work, we provide a characterization of this minimal norm certificate, and we give a sufficient condition which holds generically and which ensures that all the components of the extended support are actually nonzero (with a prediction on the signs). Our main result in this direction is Theorem 1.

Our second contribution applies this result to Problem (5) on thin grids. After recalling the convergence properties of Problem (5) towards (3), we show that under some assumption, 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  $\nu$ ), the model at low noise actually reconstructs pairs of

Dirac masses, i.e. solutions of the form

$$m_\lambda = \sum_{\nu=1}^N (\alpha_{\lambda,\nu} \delta_{x_{0,\nu}} + \beta_{\lambda,\nu} \delta_{x_{0,\nu} + \varepsilon_\nu h}), \quad \text{where } \varepsilon_\nu \in \{-1, +1\}, \quad (10)$$

$$\text{and } \text{sign}(\alpha_{\lambda,\nu}) = \text{sign}(\beta_{\lambda,\nu}) = \text{sign}(\alpha_{0,\nu}). \quad (11)$$

The precise statement of this result can be found in Theorem 2. Compared to [21] 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  $\varepsilon$ . 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  $\lambda = O(h)$ .

Then, we turn to the Continuous Basis-Pursuit (8). We first study this problem in an abstract setting, where it is reformulated as a LASSO with positivity constraints. We derive similar noise robustness properties as for the LASSO, and we characterize the extended support, see Theorem 3. Working on a thin grid, we show the  $\Gamma$ -convergence of Problem (8) towards the Beurling LASSO (3) with positivity constraints and we give a fine analysis of the support of the solutions as the grid stepsize tends to zero. We also study the low noise behavior when the measure has support on the grid: under a suitable assumption, the recovered spikes appear again in pairs,

$$m_\lambda = \sum_{\nu=1}^N (\alpha_{\lambda,\nu} \delta_{x_{0,\nu} + t_\nu} + \beta_{\lambda,\nu} \delta_{x_{0,\nu} + \varepsilon_\nu h/2}) \quad \text{where } \begin{cases} \varepsilon_\nu \in \{-1, +1\}, \\ -h/2 < t_\nu < h/2, \end{cases}$$

see Theorem 4. A closed form expression for  $\varepsilon$  is given, which depends on some corresponding *natural shift* intrinsic to the measure (which differs from the one of the LASSO). The corresponding low noise regime is characterized by  $\frac{\|w\|_2}{\lambda} = O(1)$  and  $\lambda = O(h^3)$ .

It is important to realize that, in this setting of convolution 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, on contrary to common belief, the spikes locations for LASSO and C-BP are not stable: even for an arbitrary small noise  $w$ , neither methods retrieve the correct input spikes locations.

Eventually, we illustrate in Section 6 these theoretical results with numerical experiments. We first display the evolution of the solution path  $\lambda \mapsto a_\lambda$  (a solution of (5)) and  $\lambda \mapsto (a_\lambda, b_\lambda)$  (a solution of (8)). These paths are piecewise-affine, and our contributions (Theorems 2 and 4) precisely characterize the first affine segment of these paths, which perfectly matches the numerical observations. We then illustrate our abstract analysis of the LASSO problem (5) (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  $\Phi_G$  appearing in (5) 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  $\ell^2$  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).

### 1.7. 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 (4),  $\mathcal{M}(\mathbb{T})$  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  $\psi \in C(\mathbb{T})$ ,  $\lim_{n \rightarrow +\infty} \int_{\mathbb{T}} \psi dm_n = \int_{\mathbb{T}} \psi dm$ . 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  $m_{\alpha, x_0} \stackrel{\text{def.}}{=} \sum_{\nu=1}^N \alpha_\nu \delta_{x_{0,\nu}}$  hints that  $\alpha_\nu \neq 0$  for all  $\nu$  (contrary to the notation  $m_{a,G}$ ), and that the  $x_{0,\nu}$ 's are pairwise distinct.

The properties of  $\Phi : \mathcal{M}(\mathbb{T}) \rightarrow L^2(\mathbb{T})$  and its adjoint are recalled in Proposition 6 in Appendix. The  $\infty, 2$ -operator norm of  $\Phi^* : L^2(\mathbb{T}) \rightarrow C(\mathbb{T})$  is defined as  $\|\Phi^*\|_{\infty, 2} \stackrel{\text{def.}}{=} \sup \{ \|\Phi^* w\|_\infty ; w \in L^2(\mathbb{T}), \|w\|_{L^2} \leq 1 \}$  (and the  $\infty, 2$  operator norm of a matrix is defined similarly). Given a vector  $x_0 \in \mathbb{T}^N$ ,  $\Phi_{x_0}$  refers to the linear operator  $\mathbb{R}^N \rightarrow L^2(\mathbb{T})$ , with

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

It may also be seen as the restriction of  $\Phi$  to measures supported on the set  $\{x_{0,\nu} ; \nu \in [1, N]\}$ . A similar notation is adopted for  $\Phi'_{x_0}$  (replacing  $\varphi(\cdot, x_{0,\nu})$  with  $\partial_2 \varphi(\cdot, x_{0,\nu})$ ). The concatenation of  $\Phi_{x_0}$  and  $\Phi'_{x_0}$  is denoted by  $\Gamma_{x_0} \stackrel{\text{def.}}{=} \begin{pmatrix} \Phi_{x_0} & \Phi'_{x_0} \end{pmatrix}$ .

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

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

$$\liminf_{n \rightarrow +\infty} C_n = \left\{ x \in \mathbb{T} ; \limsup_{n \rightarrow +\infty} d(x, C_n) = 0 \right\} \quad (13)$$

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

$$\lim_{n \rightarrow +\infty} C_n = C. \quad (14)$$

If the sequence  $(C_n)_{n \in \mathbb{N}}$  is nondecreasing ( $C_n \subset C_{n+1}$ ), then  $\lim_{n \rightarrow \infty} C_n = \overline{\bigcup_{n \in \mathbb{N}} C_n}$ , and if it is nonincreasing ( $C_n \supset C_{n+1}$ ) then  $\lim_{n \rightarrow \infty} C_n = \bigcap_{n \in \mathbb{N}} \overline{C_n}$  (where  $\overline{C}$  denotes the closure of  $C$ ). We refer the reader to [29] 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 [26]) 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  $\eta_0$ , 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 Hilbert space  $\mathcal{H}$ , which might be for instance  $L^2(\mathbb{T})$  (as in the previous section) or a finite dimensional vector space. The linear degradation operator is then denoted as  $A : \mathbb{R}^P \rightarrow \mathcal{H}$ . Let us emphasize that in this section, for  $a \in \mathbb{R}^P$ ,  $\|a\|_\infty \stackrel{\text{def.}}{=} \max_{0 \leq k \leq P-1} |a_k|$ .

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  $a_0 \in \mathbb{R}^P$  by solving the LASSO problem for  $\lambda > 0$ ,

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

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

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

If  $a \in \mathbb{R}^P$ , we denote by  $I(a)$ , or  $I$  when the context is clear, the support of  $a$ , *i.e.*  $I(a) \stackrel{\text{def.}}{=} \{i \in \llbracket 0, P-1 \rrbracket ; a_i \neq 0\}$ . Also, we let  $s_I \stackrel{\text{def.}}{=} \text{sign}(a_I)$ , and  $\text{supp}^\pm(a) \stackrel{\text{def.}}{=} \{(i, s_i) ; i \in I\}$  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  $a_\lambda \in \mathbb{R}^P$ . Then  $a_\lambda$  is a solution to  $(\mathcal{P}_\lambda(y))$  if and only if there exists  $p_\lambda \in \mathcal{H}$  such that*

$$\|A^*p_\lambda\|_\infty \leq 1, \quad \text{and} \quad (A^*p_\lambda)_I = \text{sign}(a_{\lambda,I}), \quad (15)$$

$$\lambda A^*p_\lambda + A^*(Aa_\lambda - y) = 0. \quad (16)$$

*Similarly, if  $a_0 \in \mathbb{R}^P$ , 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.*

$$\|A^*p\|_\infty \leq 1 \quad \text{and} \quad (A^*p)_I = \text{sign}(a_{0,I}). \quad (17)$$

Conditions (15) and (17) merely express the fact that  $\eta_\lambda \stackrel{\text{def.}}{=} A^*p_\lambda$  (resp.  $\eta \stackrel{\text{def.}}{=} A^*p$ ) is in the subdifferential of the  $\ell^1$ -norm at  $a_\lambda$  (resp.  $a_0$ ). In that case we say that  $\eta_\lambda$  (resp.  $\eta$ ) is a *dual certificate* for  $a_\lambda$  (resp.  $a_0$ ). Condition (17) is also called the *source condition* in the literature [7].

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

$$\begin{aligned} & \inf_{p \in C} \left\| \frac{y}{\lambda} - p \right\|_2^2, & (\mathcal{D}_\lambda(y)) \\ \text{resp.} & \sup_{p \in C} \langle y_0, p \rangle, & (\mathcal{D}_0(y_0)) \\ & \text{where } C \stackrel{\text{def.}}{=} \left\{ p \in \mathcal{H} ; \max_{k \in \llbracket 0, P-1 \rrbracket} |(A^*p)_k| \leq 1 \right\}. & (18) \end{aligned}$$

If  $a$  is a solution to  $(\mathcal{P}_\lambda(y))$  and  $p_\lambda$  is a solution to  $(\mathcal{D}_\lambda(y))$ , then (15) and (16) hold. Conversely, for any  $a \in \mathbb{R}^P$  and any  $p_\lambda \in \mathcal{H}$ , if (15) and (16) hold, then  $a$  is a solution to  $(\mathcal{P}_\lambda(y))$  and  $p_\lambda$  is a solution to  $(\mathcal{D}_\lambda(y))$ . A similar equivalence hold 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 (15) and (16) is unique. On the contrary, the dual certificate  $\eta = A^*p$  which appears in (17) 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  $a_0 \in \mathbb{R}^P$  such that  $A_I$  is injective and that there exists  $p \in \mathcal{H}$  such that*

$$\|(A^*p)_{I^c}\|_\infty < 1 \quad \text{and} \quad (A^*p)_I = \text{sign}(a_{0,I}), \quad (19)$$

where  $I^c = \llbracket 1, P \rrbracket \setminus I$ . Then  $a_0$  is identifiable.

Conversely, if  $a_0$  is identifiable, there exists  $p \in L^2(\mathbb{T})$  such that (19) holds and  $A_I$  is injective (see [27, Lemma 4.5]).

## 2.2. Extended support of the LASSO

From now on, we assume that the vector  $a_0 \in \mathbb{R}^P$  is identifiable (*i.e.*  $a_0$  is the unique solution to  $(\mathcal{P}_0(y_0))$  where  $y_0 = Aa_0$ ). We denote by  $I = \text{supp}(a_0)$  and  $s_I = \text{sign}(a_{0,I})$  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  $\lambda \rightarrow 0$  (see [12] for the noiseless case and [27] when the observation is  $y = y_0 + w$  and the noise  $w$  tends to zero as a multiple of  $\lambda$ ) at least in terms of the  $\ell^2$  convergence. In terms of the support of the solutions, the study in [21], which extends the one by Fuchs [26], emphasizes the role of a specific *minimal-norm certificate*  $\eta_0$  which governs the behavior of the model at low noise regimes.

**Definition 1** (Minimal-norm certificate and extended support). *Let  $a_0 \in \mathbb{R}^P$ , and let  $p_0$  be the solution to  $(\mathcal{D}_0(y_0))$  with minimal  $L^2$ -norm. The minimal-norm certificate of  $a_0$  is defined as  $\eta_0 \stackrel{\text{def.}}{=} A^*p_0$ . The set of indices  $\text{ext}(a_0) \stackrel{\text{def.}}{=} \{1 \leq j \leq P ; |(\eta_0)_j| = 1\}$  is called the extended support of  $a_0$ , and the set  $\text{ext}^\pm(a_0) \stackrel{\text{def.}}{=} \{(j, (\eta_0)_j) ; j \in \text{ext}(a_0)\} \subset \llbracket 0, P-1 \rrbracket \times \{-1, 1\}$  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  $(I, \text{sign}(a_{0,I})) \subset \text{ext}^\pm(a_0)$ . The minimal norm certificate thus turns out to be

$$\eta_0 = A^*p_0 \quad \text{where} \quad p_0 = \underset{p \in \mathcal{H}}{\text{argmin}} \{ \|p\|_2 ; \|A^*p\|_\infty \leq 1 \quad \text{and} \quad A_I^*p = s_I \}. \quad (20)$$

It is shown in [21] 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  $\eta_0$  and the structure of the extended (signed) support  $\text{ext}^\pm(a_0)$ . The following (new) result gives a characterization of  $\eta_0$ .

**Lemma 1.** *Let  $(J, s_J) \subset \llbracket 0, P-1 \rrbracket \times \{-1, 1\}$  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 (20), we see that  $p \in L^2(\mathbb{T})$  is equal to  $p_0$  if and only if  $\|A^*p\|_\infty \leq 1$ ,  $A_J^*p = \text{sign}(a_{0,I})$ , and there exists  $u_+ \in (\mathbb{R}^+)^P$  and  $u_- \in (\mathbb{R}^+)^P$  such that:

$$2p + Au_+ - Au_- = 0, \quad (21)$$

where for  $i \in I^c$ ,  $u_{+,i}$  (resp.  $u_{-,i}$ ) is a Lagrange multiplier for the constraint  $(A^*p)_i \leq 1$  (resp.  $(A^*p)_i \geq -1$ ) 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  $(A^*p)_i = \text{sign}(a_0)_i$ .

Now, 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 (21) as

$$p_0 - A_J v_J = 0,$$

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

Conversely, let  $(J, s_J) \subset \llbracket 0, P-1 \rrbracket \times \{-1, 1\}$  (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  $\|A^* p\|_\infty \leq 1$ ,  $A_J^* p = \text{sign}(a_0)_I$ , and (21) 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.  $\square$

As mentioned above, 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 theorem shows that, in the generic case, both signed supports are equal.

**Theorem 1.** *Let  $a_0 \in \mathbb{R}^P \setminus \{0\}$  be an identifiable signal,  $J \stackrel{\text{def.}}{=} \text{ext}(a_0)$  such that  $A_J$  has full rank, and  $v_J \stackrel{\text{def.}}{=} (A_J^* A_J)^{-1} \text{sign}(\eta_{0,J})$ . Assume that for all  $j \in J \setminus I$ ,  $v_j \neq 0$ . Then, there exists constants  $C^{(1)} > 0$ ,  $C^{(2)} > 0$  (which depend only on  $A$ ,  $I$  and  $\text{sign}(a_{0,I})$ ) such that for  $\lambda \leq C^{(1)} \left( \min_{i \in I} |a_{0,I}| \right)$  and all  $w \in \mathcal{H}$  with  $\|w\| \leq C^{(2)} \lambda$  the solution  $\tilde{a}_\lambda$  of  $(\mathcal{P}_\lambda(y))$  is unique,  $\text{supp}(\tilde{a}_\lambda) = J$  and it reads*

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

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

*Proof.* We define a candidate solution  $\hat{a}$  by

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

and we prove that  $\hat{a}$  is the unique solution to  $(\mathcal{P}_\lambda(y_0 + w))$  using the optimality conditions (15) and (16).

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

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

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 1 that  $\text{sign}(v_k) = -s_k$ . The constraint  $\text{sign}(\hat{a}_K) = s_K$  is thus implied by

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

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

$$A_J^*(y - A\hat{a}) = \lambda s_J. \tag{22}$$

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

$$\|A_{J^c}^*(y - A\hat{a})\|_\infty < \lambda. \tag{23}$$

Since we have

$$y - A\hat{a} = w - A_J(A_J^*A_J)^{-1}A_J^*w + \lambda A_J(A_J^*A_J)^{-1}s_J = P_{\ker(A_J^*)}w + \lambda A_J^{+,*}s_J,$$

we see that (23) 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

$$c_1\|w\| + c_2\lambda < T \quad \text{where} \quad \begin{cases} c_1 \stackrel{\text{def.}}{=} \|R_I A_J^+\|_{\infty,2}, \\ c_2 \stackrel{\text{def.}}{=} \|v_I\|_\infty, \end{cases} \quad (24)$$

$$\|w\| \leq c_3\lambda \quad \text{where} \quad c_3 \stackrel{\text{def.}}{=} (\|R_K A_J^+\|_{\infty,2})^{-1} \left( \min_{k \in K} |v_k| \right) > 0, \quad (25)$$

$$c_4\|w\| - c_5\lambda < 0 \quad \text{where} \quad \begin{cases} c_4 \stackrel{\text{def.}}{=} \|A_{J^c}^*P_{\ker(A_J^*)}\|_{2,\infty}, \\ c_5 \stackrel{\text{def.}}{=} 1 - \|\eta_{0,J^c}\|_\infty > 0. \end{cases} \quad (26)$$

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

*Remark 3* (Comparison with the analysis of Fuchs). When  $J = I$ , Theorem 1 recovers exactly the result of Fuchs [26]. Note that this result has been extended beyond the  $\ell^1$  setting, see in particular [35, 34] 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

$$\eta_F \stackrel{\text{def.}}{=} A^* A_I^{+,*} s_I \quad (27)$$

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 [37]). 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 [21, 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  $\eta_\lambda$  may be strictly included in  $J$ . Indeed, it is possible, using similar calculations as above, to construct  $w$  such that  $\text{supp } \tilde{a}_\lambda \subsetneq J$  with  $\lambda$  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  $\varphi \in C^k(\mathbb{T} \times \mathbb{T})$  ( $k \geq 2$ ),  $w \in L^2(\mathbb{T})$  and

$$\forall x \in \mathbb{T}, (\Phi m)(x) \stackrel{\text{def.}}{=} \int_{\mathbb{T}} \varphi(x, y) dm(y), \quad (28)$$

so that  $\Phi : \mathcal{M}(\mathbb{T}) \rightarrow L^2(\mathbb{T})$  is a bounded linear operator. Observe that  $\Phi$  is in fact weak\* to weak continuous and its adjoint is compact (see Lemma 6 in Appendix).

Typically, we assume that the unknown measure  $m_0$  is sparse, in the sense that it is of the form  $m_0 = \sum_{\nu=1}^N \alpha_{0,\nu} x_{0,\nu}$  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 first approach we study is the one of 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  $\lambda$  well chosen) as the grid gets thinner and thinner. To this end, we take advantage of the characterizations given in Section 2 with  $\mathcal{H} \stackrel{\text{def.}}{=} L^2(\mathbb{T})$ , regardless of the grid, and we use the Beurling LASSO (3) as a limit of the discrete models.

#### 3.1. Notations and preliminaries

For the sake of simplicity we only study uniform grids, *i.e.*  $\mathcal{G} \stackrel{\text{def.}}{=} \{ih ; i \in \llbracket 0, P-1 \rrbracket\}$  where  $h \stackrel{\text{def.}}{=} \frac{1}{P}$  is the stepsize. Moreover, we shall consider sequences of grids  $(\mathcal{G}_n)_{n \in \mathbb{N}}$  such that the stepsize vanishes ( $h_n = \frac{1}{P_n} \rightarrow 0$  as  $n \rightarrow +\infty$ ) and to ensure monotonicity, we assume that  $\mathcal{G}_{n+1} \subset \mathcal{G}_n$ . 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  $\sum_{k=0}^{P_n-1} a_k \delta_{kh_n}$ ) and vectors  $a \in \mathbb{R}^{P_n}$ .

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  $\Phi$  to measures with support in the grid  $\mathcal{G}_n$ ,

$$A \stackrel{\text{def.}}{=} \Phi_{\mathcal{G}_n} = \left( \varphi(\cdot, 0), \dots, \varphi(\cdot, (P-1)h_n) \right). \quad (29)$$

More explicitly, on the grid  $\mathcal{G}_n$ , we solve

$$\begin{aligned} \min_{a \in \mathbb{R}^{P_n}} \frac{1}{2} \|y - \Phi_{\mathcal{G}_n} a\|^2 + \lambda \|a\|_1, & \quad (\mathcal{P}_\lambda^n(y)) \\ \text{and } \min_{a \in \mathbb{R}^{P_n}} \|a\|_1 \text{ such that } \Phi_{\mathcal{G}_n} a = y_0. & \quad (\mathcal{P}_0^n(y_0)) \end{aligned}$$

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  $m_0 = \sum_{k=0}^{P_n-1} a_i \delta_{ih_n}$  and that the vector  $a$  is identifiable using  $(\mathcal{P}_0^n(y_0))$ .

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

The optimality conditions (16) amount to the existence of some  $p_\lambda \in L^2(\mathbb{T})$  such that

$$\max_{0 \leq k \leq P_n - 1} |(\Phi^* p_\lambda)(kh_n)| \leq 1, \quad \text{and} \quad (\Phi^* p_\lambda)(Ih_n) = \text{sign}(a_{\lambda, I}), \quad (30)$$

$$\lambda \Phi^* p_\lambda + \Phi^*(\Phi a_\lambda - y) = 0. \quad (31)$$

Similarly the optimality condition (17) is equivalent to the existence of  $p \in L^2(\mathbb{T})$  such that

$$\max_{0 \leq k \leq P_n - 1} |(\Phi^* p)(kh_n)| \leq 1 \quad \text{and} \quad (\Phi^* p)(Ih_n) = \text{sign}(a_{0, I}). \quad (32)$$

Notice that the dual certificates are naturally given by the sampling of continuous functions  $\eta = \Phi^* p : \mathbb{T} \rightarrow \mathbb{R}$ , 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  $\eta_0^n$ , whereas the extended support on  $\mathcal{G}_n$  is defined as

$$\text{ext}_n m_0 \stackrel{\text{def.}}{=} \{t \in \mathcal{G}_n ; \eta_0^n(t) = \pm 1\}. \quad (33)$$

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

It turns out that 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 [21] their convergence towards the Beurling-LASSO used in [15, 10, 6, 31].

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

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

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

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

$$|m|(\mathbb{T}) \stackrel{\text{def.}}{=} \sup \left\{ \int_{\mathbb{T}} \psi(x) dm(x) ; \psi \in C(\mathbb{T}) \text{ and } \|\psi\|_\infty \leq 1 \right\}. \quad (34)$$

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  $x_\nu$ '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 [21] 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

$$\begin{aligned} & \sup \left\{ \int_{\mathbb{T}} \psi(x) dm(x) ; \psi \in C(\mathbb{T}), \forall k \in \llbracket 0, P_n - 1 \rrbracket |\psi|(kh_n) \leq 1 \right\} \\ & = \begin{cases} \|a\|_1 & \text{if } m = \sum_{k=0}^{P_n-1} a_k \delta_{kh_n}, \\ +\infty & \text{otherwise.} \end{cases} \end{aligned}$$

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 [10] is that when  $\Phi$  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 (15), (16) and (17)). The corresponding dual problems are

$$\begin{aligned} & \inf_{p \in C^\infty} \left\| \frac{y}{\lambda} - p \right\|_2^2, & (\mathcal{D}_\lambda^\infty(y)) \\ \text{resp.} & \sup_{p \in C^\infty} \langle y_0, p \rangle, & (\mathcal{D}_0^\infty(y_0)) \\ & \text{where } C^\infty \stackrel{\text{def.}}{=} \{p \in L^2(\mathbb{T}) ; \|\Phi^* p\|_\infty \leq 1\}. & (35) \end{aligned}$$

The source condition associated with  $(\mathcal{P}_0^\infty(y_0))$  is of particular interest. It amounts to the existence of some  $p \in L^2(\mathbb{T})$  such that

$$\|\Phi^* p\|_\infty \leq 1 \quad \text{and} \quad (\Phi^* p)(x_{0,\nu}) = \text{sign}(\alpha_{0,\nu}) \text{ for all } \nu \in \{1, \dots, N\}. \quad (36)$$

Here,  $\|\Phi^* p\|_\infty = \sup_{t \in \mathbb{T}} |(\Phi^* p)(t)|$ . Moreover, if such  $p$  exists and satisfies  $|(\Phi^* p)(t)| < 1$  for all  $t \in \mathbb{T} \setminus \{x_{0,1}, \dots, x_{0,N}\}$ , and  $\Phi_{x_0}$  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 (36) implies the optimality of  $m_0$  for  $(\mathcal{P}_0^\infty(y_0))$  but the converse is not true (see [21]).

*Remark 5.* A simple but crucial remark made in [10] is that if  $m_0$  is identifiable through  $(\mathcal{P}_0^\infty(y_0))$  and that  $\text{supp } m_0 \subset \mathcal{G}_n$ , then  $m_0$  is identifiable for  $(\mathcal{P}_0^n(y_0))$ . Similarly, observe that the source condition for  $(\mathcal{P}_0^\infty(y_0))$  implies the source condition for the  $(\mathcal{P}_0^n(y_0))$ .

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  $L^2$  norm,  $p_0^\infty$ , 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

$$\eta_0^\infty = \Phi^* p_0^\infty \quad \text{where} \quad (37)$$

$$p_0^\infty = \underset{p \in L^2(\mathbb{T})}{\text{argmin}} \{ \|p\|_2 ; \|\Phi^* p\|_\infty \leq 1, (\Phi^* p)(x_{0,\nu}) = \text{sign}(\alpha_{0,\nu}), 1 \leq \nu \leq N \}. \quad (38)$$



As with the discrete LASSO problem, a notion of extended (signed) support  $\text{ext}^\pm_\infty$  may be defined and the minimal norm certificate governs the behavior of the solutions at low noise (see [21] 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  $\eta_0^\infty \in C(\mathbb{T})$  its minimal norm certificate. We say that  $m_0$  satisfies the Non Degenerate Source Condition if

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

The Non Degenerate Source Condition might seem difficult to check in practice. The following proposition shows that it is in fact easy to check numerically on 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

$$p_V^\infty \stackrel{\text{def.}}{=} \underset{p \in L^2(\mathbb{T})}{\text{argmin}} \{ \|p\|_2 ; (\Phi^* p)(x_{0,\nu}) = \text{sign}(\alpha_{0,\nu}), (\Phi^* p)'(x_{0,\nu}) = 0, 1 \leq \nu \leq N \}. \quad (39)$$

The following proposition shows that this precertificate is easily computed by solving a linear system in the least square sense.

**Proposition 3** ([21]). 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  $\Gamma_{x_0}$  has full rank.

Then, the vanishing derivatives precertificate can be computed by

$$\eta_V^\infty \stackrel{\text{def.}}{=} \Phi^* p_V^\infty \quad \text{where} \quad p_V^\infty \stackrel{\text{def.}}{=} \Gamma_{x_0}^{+,*} \begin{pmatrix} \text{sign}(\alpha_{0,\cdot}) \\ 0 \end{pmatrix}, \quad (40)$$

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  $t \in \mathbb{T} \setminus \{x_{0,1}, \dots, x_{0,N}\}$ ,
- $\eta_V^{\infty\prime\prime}(x_{0,\nu}) \neq 0$  for all  $\nu \in \{1, \dots, N\}$ .

And in that case,  $\eta_V^\infty$  is equal to the minimal norm certificate  $\eta_0^\infty$ .

*Remark 6.* Using the block inversion formula in (40), it is possible to check that

$$p_V^\infty = \Phi_{x_0}^{+,*} \text{sign}(\alpha_{0,\cdot}) - \Pi \Phi'_{x_0} (\Phi'_{x_0} \Pi \Phi'_{x_0})^{-1} \Phi'_{x_0} \Phi_{x_0}^{+,*} \text{sign}(\alpha_{0,\cdot}), \quad (41)$$

where  $\Pi$  is the orthogonal projector onto  $(\text{Im } \Phi_{x_0})^\perp$ . If we denote by  $p_F^\infty$  the vector introduced by Fuchs (see (27)), which turns out to be

$$p_F^\infty = \underset{p \in L^2(\mathbb{T})}{\text{argmin}} \{ \|p\|_2 ; (\Phi^* p)(x_{0,\nu}) = \text{sign}(\alpha_{0,\nu}), 1 \leq \nu \leq N \},$$

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

*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 (32) on a discrete grid  $\mathcal{G}_n$ , and the one for gridless problem  $(\mathcal{P}_0^\infty(y_0))$  which should satisfy (36). 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.*  $n \rightarrow +\infty$ ). Indeed, we may identify each vector  $a \in \mathbb{R}^{P_n}$  with the measure  $m_a = \sum_{k=0}^{P_n-1} a_k \delta_{kh_n}$  (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  $(\mathcal{P}_\lambda^n(y))$   $\Gamma$ -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  $m_\lambda^n$  of  $(\mathcal{P}_\lambda^n(y))$  (resp.  $m_\lambda^\infty$  of  $(\mathcal{P}_\lambda^\infty(y))$ ) satisfies

$$\lambda |m_\lambda^n|(\mathbb{T}) \leq \lambda |m_\lambda^n|(\mathbb{T}) + \frac{1}{2} \|\Phi m_\lambda^n - y\|^2 \leq \frac{1}{2} \|y\|^2. \quad (42)$$

Hence we may restrict those problems to the set

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

which is a metrizable space for the weak\* topology. As a result, we shall work with the definition of  $\Gamma$ -convergence in metric spaces, which is more convenient than working with the general definition [14, Definition 4.1]). For more details about  $\Gamma$ -convergence, we refer the reader to the monograph [14].

**Definition 4.** We say that the Problem  $(\mathcal{P}_\lambda^n(y))$   $\Gamma$ -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  $\text{supp}(m^n) \subset \mathcal{G}_n$  and that  $m^n$  weakly\* converges towards  $m$ ,

$$\liminf_{n \rightarrow +\infty} \left( \lambda |m^n|(\mathbb{T}) + \frac{1}{2} \|\Phi m^n - y\|^2 \right) \geq \lambda |m|(\mathbb{T}) + \frac{1}{2} \|\Phi m - y\|^2.$$

- (*Limsup inequality*) there exists a sequence of measures  $(m^n)_{n \in \mathbb{N}} \in X^{\mathbb{N}}$  such that  $\text{supp}(m^n) \subset \mathcal{G}_n$ ,  $m^n$  weakly\* converges towards  $m$  and

$$\limsup_{n \rightarrow +\infty} \left( \lambda |m^n|(\mathbb{T}) + \frac{1}{2} \|\Phi m^n - y\|^2 \right) \leq \lambda |m|(\mathbb{T}) + \frac{1}{2} \|\Phi m - y\|^2.$$

The following proposition shows the  $\Gamma$ -convergence of the discretized problems toward the Beurling LASSO problem. This ensures in particular the convergence of the minimizers, which was already proved in [32].

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

$$\liminf_{n \rightarrow +\infty} (\mathcal{P}_\lambda^n(y)) = \inf (\mathcal{P}_\lambda^\infty(y)). \quad (43)$$

Each sequence  $(m_\lambda^n)_{n \in \mathbb{N}}$  such that  $m_\lambda^n$  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  $m_\lambda$  to  $(\mathcal{P}_\lambda^\infty(y))$  is unique, the minimizers of  $(\mathcal{P}_\lambda^n(y))$  converge towards  $m_\lambda$ .

*Proof.* The liminf inequality of Definition (4) is a consequence of the lower semi-continuity of the total variation and the  $L^2$  norm (since  $\Phi$  is weak\* to weak continuous,  $\Phi_{\mathcal{G}_n} m^n - y$  weakly converges towards  $\Phi m - y$ ):

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

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

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

where  $\omega_\psi : t \mapsto \sup_{|x'-x| \leq t} |\psi(x) - \psi(x')|$  is the modulus of continuity of  $\psi$ . Therefore,  $\lim_{n \rightarrow +\infty} \langle m^n, \psi \rangle = \langle m, \psi \rangle$ , and  $m^n$  weakly\* converges towards  $m$ . Incidentally, observe that  $|m^n|(\mathbb{T}) \leq |m|(\mathbb{T})$ , so that using the liminf inequality we get  $\lim_{n \rightarrow +\infty} |m^n|(\mathbb{T}) = |m|(\mathbb{T})$ . Moreover, by similar majorizations, one may prove that  $\Phi m^n$  converges strongly in  $L^2(\mathbb{T})$  towards  $\Phi m$ . As a result  $\lim_{n \rightarrow +\infty} \|\Phi m^n - y\|^2 = \|\Phi m - y\|^2$ , and the limsup inequality is proved.

Eventually, from (42) we deduce the compactness of  $X$ , hence the existence of accumulation points, and [14, 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 (43).  $\square$

The weak\* convergence of the minimizers of  $(\mathcal{P}_\lambda^n(y))$  can be described more accurately by studying the dual certificates  $p_\lambda$  and looking at the support of the solutions  $m_\lambda^n$  to  $(\mathcal{P}_\lambda^n(y))$  (see [21, Section 5.4]). One may prove that  $m_\lambda^n$  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  $t \in \mathbb{T} \setminus \{x_1, \dots, x_N\}$ .*

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

$$\limsup_{n \rightarrow +\infty} (\text{supp}(m_\lambda^n)) \subset \{x_1, \dots, x_N\}.$$

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

$$\lim_{n \rightarrow +\infty} (\text{supp}(m_\lambda^n)) = \{x_1, \dots, x_N\}. \quad (44)$$

*If, additionally,  $(\eta_\lambda^\infty)''(x_\nu) \neq 0$  for some  $\nu \in \{1, \dots, N\}$ , then for all  $n$  large enough, the restriction of  $m_\lambda^n$  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  $\text{sign}(a_{\lambda,i}) = \text{sign}(\alpha_{\lambda,\nu})$ . Moreover, if  $a_{\lambda,i+\varepsilon_{i,n}} \neq 0$ ,  $\text{sign}(a_{\lambda,i+\varepsilon_{i,n}}) = \text{sign}(\alpha_{\lambda,\nu})$ .*

We skip the proof as it is very close to the arguments of [21, Section 5.4]. Moreover the proof of Proposition 13 below 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  $\text{supp } m_0 \subset \mathcal{G}_n$  for  $n$  large enough (in other words,  $y_0 = \Phi_{\mathcal{G}_n} a_0$  for some  $a_0 \in \mathbb{R}^{P_n}$ ), so that  $m_0 = \sum_{\nu=1}^N \alpha_{0,\nu} \delta_{x_{0,\nu}}$  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** ([21]). *The following result holds:*

$$\lim_{n \rightarrow +\infty} \eta_0^n = \eta_0^\infty, \quad (45)$$

*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*

$$\text{ext}^{\pm n}(m_0) = \text{supp}^\pm(m_0) \cup (\text{supp}^\pm(m_0) + \varepsilon^n h_n), \quad (46)$$

*where  $\text{supp}^\pm(m_0) + \varepsilon^n h_n \stackrel{\text{def.}}{=} \{(x_{0,\nu} + \varepsilon_\nu^n h_n, \eta_0^\infty(x_{0,\nu})) ; 1 \leq \nu \leq N\}$ .*

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 Theorem refines that result by giving a sufficient condition for the spikes to appear in pairs (*i.e.*  $\varepsilon_\nu = \pm 1$  for  $1 \leq \nu \leq N$ ). Moreover, it shows that the value of  $\varepsilon^n$  does not depend on  $n$ , and it gives the explicit positions of the added spikes  $\varepsilon_\nu$ , for  $1 \leq \nu \leq N$ .

**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*

$$\rho \stackrel{\text{def.}}{=} (\Phi_{x_0}^* \Pi \Phi'_{x_0})^{-1} \Phi_{x_0}^* \Phi_{x_0}^{+,*} \text{sign}(m_0(x_0)) \quad (47)$$

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

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

$$\text{ext}^{\pm n}(m_0) = \{(x_\nu, \text{sign}(\alpha_{0,\nu}))\}_{1 \leq \nu \leq N} \cup \{(x_\nu + \varepsilon_\nu h_n, \text{sign}(\alpha_{0,\nu}))\}_{1 \leq \nu \leq N} \quad (48)$$

$$\text{where } \varepsilon = \text{sign}(\text{diag}(\text{sign}(\alpha_0))\rho). \quad (49)$$

In the above theorem, observe that  $\Phi_{x_0}^* \Pi \Phi'_{x_0}$  is indeed invertible since  $\Gamma_{x_0}$  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  $\lambda \leq C_n^{(1)} \min_{1 \leq \nu \leq N} |\alpha_{0,\nu}|$ , and for all  $w \in L^2(\mathbb{T})$  such that  $\|w\|_2 \leq C_n^{(2)} \lambda$ , 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*

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

$$\text{where } \text{ext}_n(m_0) = \{x_\nu\}_{1 \leq \nu \leq N} \cup \{x_\nu + \varepsilon_\nu h_n\}_{1 \leq \nu \leq N},$$

$$\varepsilon = \text{sign}(\text{diag}(\text{sign}(\alpha_0))\rho),$$

$$\text{sign}(\alpha_{\lambda,\nu}) = \text{sign}(\beta_{\lambda,\nu}) = \text{sign}(\alpha_{0,\nu}).$$

*Proof of Theorem 2.* We define a good candidate for  $\eta_0^n$  and using Lemma 1 we prove that it is indeed equal to  $\eta_0^n$  when the grid is thin enough.

To comply with the notations of Section 2, we write

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

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

Recalling that  $A = \left( \varphi(\cdot, 0), \dots, \varphi(\cdot, (P_n - 1)h_n) \right)$ , we consider the submatrices  $A_I \stackrel{\text{def.}}{=} \left( \varphi(\cdot, ih_n) \right)_{i \in I} = \left( \varphi(\cdot, x_{0,1}), \dots, \varphi(\cdot, x_{0,N}) \right)$  and  $A_{J \setminus I} \stackrel{\text{def.}}{=} \left( \varphi(\cdot, (i + \varepsilon_i)h_n) \right)_{i \in I}$  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 1, 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  $\text{sign}(v_j) = -s_j$  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  $n \rightarrow \infty$ :

$$\begin{aligned} A_{J \setminus I} &= A_0 + h_n(B_0 + O(h_n)), \quad \text{with } A_0 = A_I = \Phi_{x_0} \\ \text{and } B_0 &= \begin{pmatrix} (\partial_2)\varphi(\cdot, x_{0,1}) & \dots & (\partial_2)\varphi(\cdot, x_{0,N}) \end{pmatrix} \text{diag}((\varepsilon_{i_1}), \dots, (\varepsilon_{i_N})) \\ &= \Phi'_{x_0} \text{diag}((\varepsilon_{i_1}), \dots, (\varepsilon_{i_N})). \end{aligned}$$

By Lemma 8 in Appendix, the Gram matrix  $A_J^* A_J$  is invertible for  $n$  large enough, and

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

where  $\rho$  is defined in (47), where  $\Pi$  is the orthogonal projector onto  $(\text{Im } \Phi_{x_0})^\perp$ , and for  $\nu \in \llbracket 1, N \rrbracket$ ,  $i_\nu$  refers to the index  $i \in I$  such that  $ih_n = x_{0,\nu}$ . Therefore,  $v_{J \setminus I}$  has the sign of  $-\text{diag}(\varepsilon_{i_1}, \dots, \varepsilon_{i_N})\rho$ , and it is sufficient to choose  $\varepsilon_{i_\nu} = s_{i_\nu} \times \text{sign}(\rho_\nu)$  to ensure that  $\text{sign } v_{J \setminus I} = -s_{J \setminus I}$  for  $n$  large enough.

With that choice of  $\varepsilon$ , 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 8 in Appendix, we obtain that

$$\begin{aligned} \lim_{n \rightarrow +\infty} \tilde{p}_n &= A_0^{+,*} s_I - \Pi B_0 (B_0^* \Pi B_0)^{-1} B_0^* A_0^{+,*} s_I \\ &= \Phi_{x_0}^{+,*} \text{sign}(\alpha_{0,\cdot}) - \Pi \Phi'_{x_0} (\Phi'_{x_0}{}^* \Pi \Phi'_{x_0})^{-1} \Phi'_{x_0}{}^* \Phi_{x_0}^{+,*} \text{sign}(\alpha_{0,\cdot}) \\ &= p_V^\infty \text{ (by (41)).} \end{aligned}$$

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  $1 \leq \nu \leq N$ ,  $\Phi^* \tilde{p}_n$  is strictly concave (or strictly convex, depending on the sign of  $\eta_0^{\infty''}(x_{0,\nu})$ ) 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 1, 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  $\mathcal{G}_n$ .  $\square$

## 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  $n \rightarrow +\infty$ . For this we look at the values of  $c_1, \dots, c_5$  given in the proof of Theorem 1.

By Lemma 8 applied to  $\Phi_{\text{ext}_n(m_0)} = \left( \Phi_{x_0} \quad \Phi_{x_0} + h_n(\Phi'_{x_0} + O(h_n)) \right)$ , we see that

$$c_{1,n} \stackrel{\text{def.}}{=} \|R_I \Phi_{\text{ext}_n(m_0)}^+\|_{\infty,2} \sim \frac{1}{h_n} \|(\Phi_{x_0}^* \Pi \Phi_{x_0}^*)^{-1} \Phi_{x_0}^* \Pi\|_{\infty,2}, \quad (50)$$

$$c_{2,n} \stackrel{\text{def.}}{=} \|v_I\|_{\infty} = \left\| R_I (\Phi_{\text{ext}_n(m_0)}^* \Phi_{\text{ext}_n(m_0)})^{-1} \begin{pmatrix} s_I \\ s_I \end{pmatrix} \right\|_{\infty} \sim \frac{1}{h_n} \|\rho\|_{\infty}, \quad (51)$$

$$c_{3,n} \stackrel{\text{def.}}{=} (\|R_K \Phi_{\text{ext}_n(m_0)}^+\|_{\infty,2})^{-1} \left( \min_{k \in K} |v_k| \right) \sim \frac{\min_{k \in K} |\rho_k|}{\|(\Phi_{x_0}^* \Pi \Phi_{x_0}^*)^{-1} \Phi_{x_0}^* \Pi\|_{\infty,2}}. \quad (52)$$

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 (2.2) 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.

**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*

$$\left\| \begin{pmatrix} \alpha_{\lambda} \\ \beta_{\lambda} \end{pmatrix} - \begin{pmatrix} \alpha_0 \\ 0 \end{pmatrix} \right\|_{\infty} = O\left(\frac{w}{h_n}, \frac{\lambda}{h_n}\right). \quad (53)$$

*Proof.* The proof of (53) follows from applying (50) and (51) in the expression for  $\alpha_{\lambda}$  and  $\beta_{\lambda}$  provided by Corollary 1. Let  $\omega \stackrel{\text{def.}}{=} \Phi^* \Pi w$ , where  $\Pi$  is the orthogonal projector onto  $(\text{Im } \Phi_{x_0})^{\perp} = \ker \Phi_{x_0}^*$ . In order to ensure (23) we may ensure that :

$$|\omega(jh_n) + \lambda \eta_0^n(jh_n)| - \lambda < 0, \quad (54)$$

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

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

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

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

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

(with equality of the signs) and

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

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 (54).

Now, let  $\nu \in \{1, \dots, N\}$  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  $f: (-r, +r) \rightarrow \mathbb{R}$  satisfies  $f''(t) \leq C$  for some  $C < 0$  and  $f(0) = f(h_n) = 0$ , then  $f(t) \leq \frac{C}{2}t(t - h_n) < 0$  for  $t \in (-r, 0] \cup [h_n, r)$ .

Notice that  $\omega = \Phi^*\Pi w$  is a  $C^2$  function which vanishes on  $\text{ext}_n(m_0)$  (hence at  $x_{0,\nu}$  and  $x_{0,\nu} + h_n$ ), and that its second derivative is bounded by  $\|(\Phi'')^*\|_{\infty,2}\|w\|_2$ . Moreover,  $\eta_0^n(x_{0,\nu}) = \eta_0^n(x_{0,\nu} + h_n) = 1$  and  $\sup_{(x_{0,\nu}-r, x_{0,\nu}+r)}(\eta_0^n)'' \leq \frac{1}{2}(\eta_0^\infty)''(x_{0,\nu}) < 0$ . Thus, for  $\frac{\|w\|_2}{\lambda} < \frac{|(\eta_0^\infty)''(x_{0,\nu})|}{2\|(\Phi'')^*\|_{\infty,2}}$ , we may apply the observation to  $\omega(\cdot - x_{0,\nu}) + \lambda(\eta_0^n(\cdot - x_{0,\nu}) - 1)$  so as to get

$$\omega(t) + \lambda(\eta_0^n(t) - 1) \leq \left( \|(\Phi'')^*\|_{\infty,2}\|w\|_2 + \lambda\frac{1}{2}(\eta_0^\infty)''(x_{0,\nu}) \right) (t - x_{0,\nu})(t - x_{0,\nu} - h_n) < 0$$

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 (54) 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  $\frac{\|w\|_2}{\lambda} \leq c$ , 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  $C_n^{(2)} = \min(c_{3,n}, c) = O(1)$ .  $\square$

#### 4. Abstract analysis of the Lasso with cone constraint

This section studies a simple variant of the LASSO with cone constraint in an abstract setting. The results stated here shall be useful in Section 5, since this variant turns out to be the Continuous Basis-Pursuit when the degradation operator is a convolution with an impulse response and its derivative. Similarly to Section 2, we consider in this section observations in an arbitrary Hilbert space  $\mathcal{H}$ .

##### 4.1. Notations

Given a parameter  $h > 0$ , we consider the cone generated by the vectors  $(1, \frac{h}{2})$  and  $(1, -\frac{h}{2})$ ,

$$\mathcal{C}_h \stackrel{\text{def.}}{=} \left\{ (c, d) \in \mathbb{R} \times \mathbb{R} ; c \geq 0 \quad \text{and} \quad -c\frac{h}{2} + |d| \leq 0 \right\}. \quad (55)$$

We also define the cone  $\mathcal{C}_h^P$  as the set of vectors  $(a, b) \in \mathbb{R}^P \times \mathbb{R}^P$  such that for all  $k \in \llbracket 0, P-1 \rrbracket$   $(a_k, b_k) \in \mathcal{C}_h$ .

Now, given a vector  $(a_0, b_0) \in \mathcal{C}_h^P$  (i.e.  $\forall k \in \llbracket 0, P-1 \rrbracket, a_{0,k} \geq \frac{2}{h}|b_{0,k}|$ ), we observe  $y_0 = Aa_0 + Bb_0$ , where  $A: \mathbb{R}^P \rightarrow \mathcal{H}$  and  $B: \mathbb{R}^P \rightarrow \mathcal{H}$  are linear operators, or its noisy version  $y = y_0 + w$  where  $w \in \mathcal{H}$ . To recover  $(a_0, b_0)$  from  $y$  or  $y_0$ , we consider the following reconstruction problems:

$$\min_{(a,b) \in \mathcal{C}_h^P} \frac{1}{2} \|y - Aa - Bb\|_2^2 + \lambda \|a\|_1, \quad (\mathcal{Q}_\lambda(y))$$



and for  $\lambda = 0$ ,

$$\min_{(a,b) \in \mathcal{C}_h^P} \|a\|_1 \text{ such that } Aa + Bb = y_0. \quad (\mathcal{Q}_0(y_0))$$

Our main focus is on the support recovery properties of  $(\mathcal{Q}_\lambda(y))$ . Precisely, we split the “support” of  $(a, b) \in \mathcal{C}_h^P$  into several parts:

$$I \stackrel{\text{def.}}{=} \text{supp}(a) \stackrel{\text{def.}}{=} \{i \in \llbracket 0, P-1 \rrbracket ; a_i > 0\} \quad (56)$$

$$= I^{(\mathbf{r})} \cup I^{(1)} \quad (57)$$

$$\text{where } I^{(\mathbf{r})} \stackrel{\text{def.}}{=} \left\{ i \in I ; a_i + \frac{2}{h}b_i > 0 \right\}, \quad I^{(1)} \stackrel{\text{def.}}{=} \left\{ i \in I ; a_i - \frac{2}{h}b_i > 0 \right\}. \quad (58)$$

Observe that in general  $I^{(\mathbf{r})} \cap I^{(1)} \neq \emptyset$ . If  $(a_\lambda, b_\lambda)$  is a solution of  $(\mathcal{Q}_\lambda(y))$ , we say that we have exact support recovery provided that  $I^{(\mathbf{r})}(a_\lambda, b_\lambda) = I^{(\mathbf{r})}(a_0, b_0)$  and  $I^{(1)}(a_\lambda, b_\lambda) = I^{(1)}(a_0, b_0)$ .

*Remark 9.* The notation  $I^{(\mathbf{r})}$ ,  $I^{(1)}$ , which might seem a bit obscure at this point, shall become clearer in the next section. It turns out that when considering the Continuous Basis-Pursuit on a grid with stepsize  $h > 0$ , points  $i$  in  $I^{(\mathbf{r})}$  correspond to Dirac masses which “tend to be on the right”, that is they do not coincide with the left half-grid point  $ih - \frac{h}{2}$ . Similarly, points in  $I^{(1)}$  correspond to Dirac masses which “tend to be on the left”, as they do not coincide with the right half-grid point  $ih + \frac{h}{2}$ . In fact, if  $i \in I^{(\mathbf{r})} \setminus I^{(1)}$ , it corresponds to a Dirac mass at the right half-grid point:  $\delta_{ih + \frac{h}{2}}$ , and if  $i \in I^{(1)} \setminus I^{(\mathbf{r})}$ , it corresponds to a Dirac mass at the left half-grid point:  $\delta_{ih - \frac{h}{2}}$ . If  $i \in I^{(\mathbf{r})} \cap I^{(1)}$ , it corresponds to a Dirac mass which may belong “freely” to the interval  $(ih - \frac{h}{2}, ih + \frac{h}{2})$ .

#### 4.2. Parametrization as a positive LASSO

To characterize the solutions of  $(\mathcal{Q}_\lambda(y))$  and  $(\mathcal{Q}_0(y_0))$ , it is convenient to reparametrize the problem as a LASSO with positivity constraint. Indeed, let us write for all  $i \in \llbracket 0, P-1 \rrbracket$ ,

$$\begin{pmatrix} a_i \\ b_i \end{pmatrix} \stackrel{\text{def.}}{=} \begin{pmatrix} 1 & 1 \\ \frac{h}{2} & -\frac{h}{2} \end{pmatrix} \begin{pmatrix} r_i \\ l_i \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} r_i \\ l_i \end{pmatrix} = \frac{1}{2} \begin{pmatrix} a_i + \frac{2}{h}b_i \\ a_i - \frac{2}{h}b_i \end{pmatrix}. \quad (59)$$

In the following, we define the linear map

$$H_h : \begin{pmatrix} r \\ l \end{pmatrix} \mapsto \begin{pmatrix} a \\ b \end{pmatrix} \quad (60)$$

It is clear that  $(a_i, b_i) \in \mathcal{C}_h$  if and only if  $r_i \geq 0$  and  $l_i \geq 0$ . Moreover, given  $(a, b) \in \mathcal{C}_h^P$ ,

$$I^c = \{i \in \llbracket 0, P-1 \rrbracket ; (r_i, l_i) = (0, 0)\}, \quad I^{(\mathbf{r})} = \{i \in \llbracket 0, P-1 \rrbracket ; r_i > 0\},$$

and  $I^{(1)} = \{i \in \llbracket 0, P-1 \rrbracket ; l_i > 0\}.$

Therefore, Problems  $(\mathcal{Q}_\lambda(y))$  and  $(\mathcal{Q}_0(y_0))$  are respectively equivalent to the LASSO and Basis Pursuit with positivity constraint:

$$\min_{(r,l) \in (\mathbb{R}_+)^P \times (\mathbb{R}_+)^P} \lambda \left\| \begin{pmatrix} r \\ l \end{pmatrix} \right\|_1 + \frac{1}{2} \left\| y - \mathcal{A}_h \begin{pmatrix} r \\ l \end{pmatrix} \right\|_2^2 \quad (\tilde{\mathcal{Q}}_\lambda(y))$$

$$\text{and } \min_{(r,l) \in (\mathbb{R}_+)^P \times (\mathbb{R}_+)^P} \left\| \begin{pmatrix} r \\ l \end{pmatrix} \right\|_1 \quad \text{such that } \mathcal{A}_h \begin{pmatrix} r \\ l \end{pmatrix} = y_0, \quad (\tilde{\mathcal{Q}}_0(y_0))$$

where  $\mathcal{A}_h \stackrel{\text{def.}}{=} \begin{pmatrix} A + \frac{h}{2}B & A - \frac{h}{2}B \end{pmatrix} : \mathbb{R}^{2P} \rightarrow \mathcal{H}$ .

Observe that there is ‘‘support recovery’’ of  $(a_0, b_0)$  through  $(\mathcal{Q}_\lambda(y))$  if and only if there is support recovery of  $(r_0, l_0)$  through  $(\tilde{\mathcal{Q}}_\lambda(y))$ . But precisely, as we shall explain below, the characterization of minimizers and the support recovery properties of the LASSO with positivity constraint  $(\tilde{\mathcal{Q}}_\lambda(y))$  are quite similar to those exposed in Section 2.

The regularization term may be written as  $J : \mathbb{R}^P \rightarrow \mathbb{R} \cup \{+\infty\}$ , where for all  $(r, l) \in \mathbb{R}^P \times \mathbb{R}^P$ ,

$$\begin{aligned} J(r, l) &\stackrel{\text{def.}}{=} \begin{cases} \sum_{i=0}^{P-1} (r_i + l_i) & \text{if } r_i \geq 0 \text{ and } l_i \geq 0 \text{ for all } i \in \llbracket 0, P-1 \rrbracket, \\ +\infty & \text{otherwise.} \end{cases} \\ &= \sum_{i=0}^{P-1} j(r_i) + \sum_{i=0}^{P-1} j(l_i), \end{aligned}$$

$$\text{with } j(x) \stackrel{\text{def.}}{=} \sup \{qx ; q \leq 1\} = \begin{cases} x & \text{if } x \geq 0, \\ +\infty & \text{otherwise.} \end{cases}$$

Hence, the subdifferential of  $J$  is the product of the subdifferentials  $\partial j(r_i)$  and  $\partial j(l_i)$  for  $1 \leq i \leq P-1$ , where

$$\partial j(x) = \begin{cases} \{1\} & \text{if } x > 0, \\ (-\infty, 1] & \text{if } x = 0. \end{cases}$$

That is quite similar to the subdifferential of  $|\cdot|$  at  $x \in \mathbb{R}$  which is  $-1$ ,  $[-1, 1]$  or  $1$  if  $x < 0$ ,  $x = 0$  or  $x > 0$  respectively, and one may adapt all the results of Section 2 to the LASSO with positivity constraint. It essentially amounts to replacing the conditions  $\|\eta\|_\infty \leq 1$  with  $\max \eta \leq 1$  (and similarly for strict inequalities) wherever they appear. We leave the detail to the reader, and in the following, we use those results freely to derive the properties of the LASSO with cone constraint  $(\mathcal{Q}_\lambda(y))$ .

### 4.3. Optimality conditions

Applying the results (or their straightforward adaptations) of Section 2 to  $(\tilde{\mathcal{Q}}_\lambda(y))$  and  $(\tilde{\mathcal{Q}}_0(y_0))$ , then composing by  $H_h$ , we immediately get the following results.

**Proposition 8.** *Let  $y \in \mathcal{H}$ ,  $(a_\lambda, b_\lambda) \in \mathcal{C}_h^P$ , and  $I = I(a_\lambda, b_\lambda)$ . Then  $(a_\lambda, b_\lambda)$  is a solution to  $(\mathcal{Q}_\lambda(y))$  if and only if there exists  $q_\lambda \in \mathcal{H}$  such that*

$$\max \left( (A^* + \frac{h}{2}B^*)q_\lambda \right) \leq 1, \quad \text{and} \quad \max \left( (A^* - \frac{h}{2}B^*)q_\lambda \right) \leq 1, \quad (61)$$

$$(A_{I^{(r)}}^* + \frac{h}{2}B_{I^{(r)}}^*)q_\lambda = \mathbf{1}_{I^{(r)}}, \quad \text{and} \quad (A_{I^{(l)}}^* - \frac{h}{2}B_{I^{(l)}}^*)q_\lambda = \mathbf{1}_{I^{(l)}}, \quad (62)$$

$$\lambda \begin{pmatrix} A^* \\ B^* \end{pmatrix} q_\lambda + \begin{pmatrix} A^* \\ B^* \end{pmatrix} (Aa_\lambda + Bb_\lambda - y) = 0. \quad (63)$$

Similarly,  $(a_0, b_0) \in \mathcal{C}_h^P$  is a solution to  $(\mathcal{Q}_0(y_0))$  if and only if  $Aa_0 + Bb_0 = y_0$  and there exists  $q \in \mathcal{H}$  such that

$$\max \left( (A^* + \frac{h}{2}B^*)q \right) \leq 1, \quad \text{and} \quad \max \left( (A^* - \frac{h}{2}B^*)q \right) \leq 1, \quad (64)$$

$$(A_{I^{(r)}}^* + \frac{h}{2}B_{I^{(r)}}^*)q = \mathbf{1}_{I^{(r)}}, \quad \text{and} \quad (A_{I^{(l)}}^* - \frac{h}{2}B_{I^{(l)}}^*)q \leq \mathbf{1}, \quad (65)$$

where  $I = I(a_0, b_0)$ .

The corresponding dual problems are given by

$$\inf_{q \in D} \left\| \frac{y}{\lambda} - q \right\|^2 \quad (\mathcal{E}_\lambda(y))$$

$$\sup_{q \in D} \langle y, q \rangle \quad (\mathcal{E}_0(y))$$

$$\text{where } D \stackrel{\text{def.}}{=} \left\{ q \in \mathcal{H} ; \max_{k \in \llbracket 0, P_n - 1 \rrbracket} (A^*q)_k + \frac{h}{2}|(B^*q)_k| \leq 1 \right\}, \quad (66)$$

Again, if the inequalities outside the support are strict, it is possible to ensure the uniqueness of the solution.

**Proposition 9.** *Under the hypotheses of Proposition 8, if  $\left( (A + \frac{h}{2}B)_{I^{(r)}} \quad (A - \frac{h}{2}B)_{I^{(l)}} \right)$  has full rank and if  $q_\lambda$  (resp.  $q$ ) satisfies*

$$\forall k \in I^c, \quad (A^*q_\lambda)_k + \frac{h}{2}|(B^*q_\lambda)_k| < 1, \quad (67)$$

$$\forall i \in I^{(l)} \setminus I^{(r)}, \quad ((A^* + \frac{h}{2}B^*)q_\lambda)_i < 1, \quad (68)$$

$$\forall i \in I^{(r)} \setminus I^{(l)}, \quad ((A^* - \frac{h}{2}B^*)q_\lambda)_i < 1, \quad (69)$$

then  $(a_\lambda, b_\lambda)$  (resp.  $(a_0, b_0)$ ) is the unique solution to  $(\mathcal{Q}_\lambda(y))$  (resp.  $(\mathcal{Q}_0(y_0))$ ).

#### 4.4. Low noise behavior of C-BP

The Theorem of Fuchs [26] for the Lasso (see Remark 3) extends to this setting as follows.

**Proposition 10.** Let  $(a_0, b_0) \in \mathcal{C}_h^P \setminus \{0\}$  such that

$$\hat{\mathcal{A}}_h \stackrel{\text{def.}}{=} \begin{pmatrix} (A + \frac{h}{2}B)_{I^{(r)}} & (A - \frac{h}{2}B)_{I^{(l)}} \end{pmatrix}$$

has full rank, and let

$$T \stackrel{\text{def.}}{=} \min \left\{ b_{0,i} + \frac{2}{h}a_{0,i} ; i \in I^{(r)}(b_0, a_0) \right\} \cup \left\{ b_{0,i} - \frac{2}{h}a_{0,i} ; i \in I^{(l)}(b_0, a_0) \right\}. \quad (70)$$

Then there exists constants  $C^{(1)} > 0, C^{(2)} > 0$  such that for  $\lambda \leq C^{(1)}T, \|w\| < C^{(2)}\lambda$ , the solution to  $(\mathcal{Q}_\lambda(y_0 + w))$  is unique, satisfies  $I^{(r)}(b_\lambda, a_\lambda) = I^{(r)}(b_0, a_0), I^{(l)}(b_\lambda, a_\lambda) = I^{(l)}(b_0, a_0)$ , and it reads:

$$\begin{pmatrix} a_\lambda \\ b_\lambda \end{pmatrix} = \begin{pmatrix} a_0 \\ b_0 \end{pmatrix} + H_h \hat{\mathcal{A}}_h^+ w - \lambda H_h (\hat{\mathcal{A}}_h^* \hat{\mathcal{A}}_h)^{-1} s,$$

where  $H_h$  is defined in (60).

In general, the conditions of Proposition 10 do not hold, and the support at low noise is strictly larger than  $(I^{(r)}(a_0, b_0), I^{(l)}(a_0, b_0))$ . This support is governed by the minimal norm certificate.

**Definition 5** (Minimal norm certificate). Let  $(a_0, b_0) \in \mathcal{C}_h^P$ . Its minimal norm certificate is  $\zeta_0 \stackrel{\text{def.}}{=} \begin{pmatrix} A^* + \frac{h}{2}B^* \\ A^* - \frac{h}{2}B^* \end{pmatrix} q_0$  where  $q_0$  is the solution to  $(\mathcal{E}_0(y))$  with minimal  $L^2$  norm. The extended support is  $\text{ext}^{(r,l)}(a_0, b_0) = (\text{ext}^{(r)}(a_0, b_0), \text{ext}^{(l)}(a_0, b_0))$ , where

$$\text{ext}^{(r)}(a_0, b_0) = \left\{ j \in \llbracket 0, P-1 \rrbracket ; ((A^* + \frac{h}{2}B^*)q_0)_j = 1 \right\}, \quad (71)$$

$$\text{ext}^{(l)}(a_0, b_0) = \left\{ j \in \llbracket 0, P-1 \rrbracket ; ((A^* - \frac{h}{2}B^*)q_0)_j = 1 \right\}. \quad (72)$$

From the optimality conditions, if  $(a_0, b_0)$  is a solution of  $(\mathcal{Q}_0(y_0))$  then  $I^{(r)} \subset \text{ext}^{(r)}(a_0, b_0)$  and  $I^{(l)} \subset \text{ext}^{(l)}(a_0, b_0)$  (where  $I = I(a_0, b_0)$ ), and  $q_0$  can be characterized as

$$q_0 = \underset{q \in \mathcal{H}}{\text{argmin}} \left\{ \|q\|_2 ; \begin{pmatrix} A^* + \frac{h}{2}B^* \\ A^* - \frac{h}{2}B^* \end{pmatrix} q \in \partial J(r_0, l_0) \right\}. \quad (73)$$

**Lemma 2.** Let  $J^{(r)}, J^{(l)} \subset \llbracket 0, P-1 \rrbracket$ , and  $(a_0, b_0) \in \mathcal{C}_h^P$ . Assume that  $(I^{(r)}, I^{(l)}) \stackrel{\text{def.}}{=} (I^{(r)}(a_0, b_0), I^{(l)}(a_0, b_0))$  is such that

$$I^{(r)} \subset J^{(r)}, \quad I^{(l)} \subset J^{(l)} \quad \text{and} \quad \hat{\mathcal{A}}_h \stackrel{\text{def.}}{=} \begin{pmatrix} (A + \frac{h}{2}B)_{J^{(r)}} & (A - \frac{h}{2}B)_{J^{(l)}} \end{pmatrix}$$

has full rank. Define  $\begin{pmatrix} u_{J^{(r)}} \\ v_{J^{(l)}} \end{pmatrix} \stackrel{\text{def.}}{=} -(\hat{\mathcal{A}}_h^* \hat{\mathcal{A}}_h)^{-1} s$  where  $s \stackrel{\text{def.}}{=} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^{|J^{(r)}|+|J^{(l)}|}$ . Then

$(J^{(r)}, J^{(l)})$  is the extended support of  $(a_0, b_0)$  if and only if the following two conditions hold:

- for all  $j \in J^{(\mathbf{r})} \setminus I^{(\mathbf{r})}$ ,  $u_j \geq 0$ , and for all  $j \in J^{(1)} \setminus I^{(1)}$ ,  $v_j \geq 0$ .
- $\max \left[ \begin{pmatrix} (A + \frac{h}{2}B)_{J^{(\mathbf{r})}^c}^* \\ (A - \frac{h}{2}B)_{J^{(1)}^c}^* \end{pmatrix} \hat{\mathcal{A}}_h \begin{pmatrix} u_{J^{(\mathbf{r})}} \\ v_{J^{(1)}} \end{pmatrix} \right] < 1$ .

Moreover, in that case, the minimal norm certificate is given by

$$\zeta_0 = -\mathcal{A}_h^* (\hat{\mathcal{A}}_h^* \hat{\mathcal{A}}_h)^{-1} s.$$

The proof is identical to the one of Lemma 1, therefore we omit it. We are now in position to describe the behavior of  $(\mathcal{Q}_\lambda(y))$  at low noise in the generic case:

**Theorem 3.** Let  $(a_0, b_0) \in (\mathbb{R}_+)^2 \setminus \{0\}$  be an identifiable signal,  $(J^{(\mathbf{r})}, J^{(1)}) = \text{ext}^{(\mathbf{r}1)}(a_0, b_0)$  such that  $\hat{\mathcal{A}}_h \stackrel{\text{def.}}{=} \begin{pmatrix} (A + \frac{h}{2}B)_{J^{(\mathbf{r})}} & (A - \frac{h}{2}B)_{J^{(1)}} \end{pmatrix}$  has full rank. Let

$\begin{pmatrix} u_{J^{(\mathbf{r})}} \\ v_{J^{(1)}} \end{pmatrix} \stackrel{\text{def.}}{=} -(\hat{\mathcal{A}}_h^* \hat{\mathcal{A}}_h)^{-1} s$  where  $s \stackrel{\text{def.}}{=} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^{|J^{(\mathbf{r})}|+|J^{(1)}|}$ , and assume that for all  $j \in J^{(\mathbf{r})} \setminus I^{(\mathbf{r})}$ ,  $u_j > 0$ , and that for all  $j \in J^{(1)} \setminus I^{(1)}$ ,  $v_j > 0$ .

Then, there exists constants  $C^{(1)} > 0$ ,  $C^{(2)} > 0$  such that for

$$\lambda \leq C^{(1)} \min \left\{ b_{0,i} + \frac{2}{h} a_{0,i}; i \in I^{(\mathbf{r})}(a_0, b_0) \right\} \cup \left\{ b_{0,i} - \frac{2}{h} a_{0,i}; i \in I^{(1)}(a_0, b_0) \right\} \quad (74)$$

and  $\|w\|_2 \leq C^{(2)} \lambda$ , the solution  $(a_\lambda, b_\lambda)$  to  $(\mathcal{Q}_\lambda(y))$  is unique,  $I^{(\mathbf{r})}(a_\lambda, b_\lambda) = J^{(\mathbf{r})}$ ,  $I^{(1)}(a_\lambda, b_\lambda) = J^{(1)}$ , and it reads

$$\begin{pmatrix} a_\lambda \\ b_\lambda \end{pmatrix} = \begin{pmatrix} a_0 \\ b_0 \end{pmatrix} + H_h \hat{\mathcal{A}}_h^+ w - \lambda H_h (\hat{\mathcal{A}}_h^* \hat{\mathcal{A}}_h)^{-1} s,$$

where  $H_h$  is defined in (60).

## 5. Continuous-Basis Pursuit on thin grids

Facing the same inverse problem as in Section 3, but this time assuming that each  $\alpha_\nu$  ( $1 \leq \nu \leq N$ ) is positive, we aim at recovering  $m_0$  using the Continuous Basis-Pursuit proposed in [23]. Given a grid  $\mathcal{G}_n$  as in Section 3, the goal is to reconstruct a measure  $m = \sum_{i=0}^{P_n-1} a_i \delta_{ih_n+t_i}$  where  $t_i \in [-\frac{h_n}{2}, \frac{h_n}{2}]$  which estimates  $m_0$ . Applying a Taylor expansion and setting  $b_i = t_i a_i$ , the authors of [23] are led to solve

$$\min_{(a,b) \in \mathcal{C}_{h_n}^{P_n}} \frac{1}{2} \|y - \Phi_{\mathcal{G}_n} a - \Phi'_{\mathcal{G}_n} b\|^2 + \lambda \|a\|_1 \quad (\mathcal{Q}_\lambda^n(y))$$

$$\min_{(a,b) \in \mathcal{C}_{h_n}^{P_n}} \|a\|_1 \text{ such that } \Phi_{\mathcal{G}_n} a + \Phi'_{\mathcal{G}_n} b = y_0. \quad (\mathcal{Q}_0^n(y_0))$$

which are particular instances of  $(\mathcal{Q}_\lambda(y))$  and  $(\mathcal{Q}_0(y_0))$ . The dual problems are respectively:

$$\inf_{q \in D^n} \left\| \frac{y}{\lambda} - q \right\|^2 \quad (\mathcal{E}_\lambda^n(y))$$

$$\sup_{q \in D^n} \langle y_0, q \rangle \quad (\mathcal{E}_0^n(y_0))$$

$$\text{where } D^n \stackrel{\text{def.}}{=} \left\{ q \in L^2(\mathbb{T}) ; \max_{k \in [0, P_n-1]} (\Phi^* q)(kh_n) + \frac{h_n}{2} |(\Phi^* q)'(kh_n)| \leq 1 \right\}, \quad (75)$$

To study the behavior of the solutions to these problems as  $n$  increases, we aim at applying the results of the previous section, and in particular Lemma 2, in the setting where  $\mathcal{H} = L^2(\mathbb{T})$  and  $(A, B) = (\Phi_{\mathcal{G}_n}, \Phi'_{\mathcal{G}_n})$ .

### 5.1. The positive Beurling LASSO

The situation with the continuous basis pursuit on thin grids is quite similar to the situation of the LASSO. Still, the non-negativity constraint on the components  $(a_k)_{0 \leq k \leq P_n}$  passes to the limit, and the appropriate limit model is the *positive Beurling LASSO*,

$$\begin{aligned} \min_{m \in \mathcal{M}^+(\mathbb{T})} \frac{1}{2} \|y - \Phi m\|^2 + \lambda m(\mathbb{T}), & \quad (\mathcal{Q}_\lambda^\infty(y)) \\ \text{and } \min_{m \in \mathcal{M}^+(\mathbb{T})} m(\mathbb{T}) \text{ such that } \Phi m = y_0, & \quad (\mathcal{Q}_0^\infty(y_0)) \end{aligned}$$

where  $\mathcal{M}^+(\mathbb{T})$  refers to the space of positive Radon measures. The indicator function of positive measures plus the total mass may be encoded in the quantity:

$$m(\mathbb{T}) + \iota_{\mathcal{M}^+(\mathbb{T})}(m) = \sup \left\{ \int_{\mathbb{T}} \psi(t) dm(t) ; \psi \in C(\mathbb{T}) \text{ and } \max_{t \in \mathbb{T}} \psi(t) \leq 1 \right\}. \quad (76)$$

As a result, the characterization of optimality, the notions of minimal norm certificates and extended support may be adapted from Section 3.2 in a straightforward manner, replacing condition  $\|\eta\|_\infty \leq 1$  by  $\sup_{t \in \mathbb{T}} \mu(t) \leq 1$  where  $\mu = \Phi^* q$  for  $q \in L^2(\mathbb{T})$ . For instance, up to the addition of a constant, the dual problems to  $(\mathcal{Q}_\lambda^\infty(y))$  and  $(\mathcal{Q}_0^\infty(y_0))$  are respectively:

$$\begin{aligned} \inf_{q \in D^\infty} \left\| \frac{y}{\lambda} - p \right\|^2 & \quad (\mathcal{E}_\lambda^\infty(y)) \\ \sup_{q \in D^\infty} \langle y_0, q \rangle & \quad (\mathcal{E}_0^\infty(y_0)) \end{aligned}$$

$$\text{where } D^\infty \stackrel{\text{def.}}{=} \left\{ q \in L^2(\mathbb{T}) ; \max_{t \in \mathbb{T}} (\Phi^* q)(t) \leq 1 \right\}. \quad (77)$$

### 5.2. The limit problem for thin grids

To consider the limit of  $(\mathcal{Q}_\lambda^n(y))$ , let us recall that we obtain a measure from the vector  $(a, b) \in \mathcal{C}_{h_n}^{P_n}$  by setting

$$m = \sum_{i=0}^{P_n-1} a_i \delta_{ih_n + b_i/a_i} \quad (78)$$

with the convention that  $b_i/a_i = 0$  if  $a_i = 0$ . It should be noticed that  $b_i/a_i \in [-\frac{h_n}{2}, \frac{h_n}{2}]$ .

We rely again on the notion on  $\Gamma$ -convergence to express the convergence of  $(\mathcal{Q}_\lambda^n(y))$  towards  $(\mathcal{Q}_\lambda^\infty(y))$ . As before, we may restrict the problems to  $X_+ \stackrel{\text{def.}}{=} \{m \in \mathcal{M}^+(\mathbb{T}) ; \lambda |m|(\mathbb{T}) \leq \frac{1}{2} \|y\|^2\}$  which is metrizable for the weak\* topology.

**Definition 6.** We say that the Problem  $(\mathcal{Q}_\lambda^n(y))$   $\Gamma$ -converges towards Problem  $(\mathcal{Q}_\lambda^\infty(y))$  if, for all  $m \in X_+$ , the following conditions hold

- (Liminf inequality) for any sequence of measures  $(m^n)_{n \in \mathbb{N}} \in X_+^{\mathbb{N}}$  of the form (78) with  $(a^{(n)}, b^{(n)}) \in \mathcal{C}_{h_n}^{P_n}$  such that  $m^n$  weakly\* converges towards  $m$ ,

$$\liminf_{n \rightarrow +\infty} \left( \lambda \|a^{(n)}\|_1 + \frac{1}{2} \|\Phi_{\mathcal{G}_n} a^{(n)} + \Phi'_{\mathcal{G}_n} b^{(n)} - y\|^2 \right) \geq \lambda m(\mathbb{T}) + \frac{1}{2} \|\Phi m - y\|^2.$$

- (Limsup inequality) there exists a sequence of measures  $(m^n)_{n \in \mathbb{N}} \in X_+^{\mathbb{N}}$  of the form (78) with  $(a^{(n)}, b^{(n)}) \in \mathcal{C}_{h_n}^{P_n}$  such that  $m^n$  weakly\* converges towards  $m$  and

$$\limsup_{n \rightarrow +\infty} \left( \lambda \|a^{(n)}\|_1 + \frac{1}{2} \|\Phi_{\mathcal{G}_n} a^{(n)} + \Phi'_{\mathcal{G}_n} b^{(n)} - y\|^2 \right) \leq \lambda m(\mathbb{T}) + \frac{1}{2} \|\Phi m - y\|^2.$$

**Proposition 11.** The Problem  $(\mathcal{Q}_\lambda^n(y))$   $\Gamma$ -converges towards  $(\mathcal{Q}_\lambda^\infty(y))$ , and

$$\lim_{n \rightarrow +\infty} \inf (\mathcal{Q}_\lambda^n(y)) = \inf (\mathcal{Q}_\lambda^\infty(y)). \quad (79)$$

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

In particular, if the solution  $m_\lambda^\infty$  to  $(\mathcal{Q}_\lambda^\infty(y))$  is unique, the whole sequence  $(m_\lambda^n)_{n \in \mathbb{N}}$  converges towards  $m_\lambda^\infty$ .

*Proof.* The proof is the same as for Proposition 4 with minor adaptations, observing that  $\|a^{(n)}\|_1 = m_n(\mathbb{T})$ . For the liminf inequality, let  $(m^n)_{n \in \mathbb{N}}$  be of the form (78) which weakly\* converges towards  $m$ . We notice that  $\Phi'_{\mathcal{G}_n} b^{(n)} = \Phi'(\sum_{i=0}^{P_n-1} b_i^{(n)} \delta_{ih_n})$ , and

$$\left| \sum_{i=0}^{P_n-1} b_i^{(n)} \delta_{ih_n} \right| (\mathbb{T}) \leq \frac{h_n}{2} \left( \sum_{i=0}^{P_n-1} a_i^{(n)} \right) \leq \frac{h_n}{2\lambda} \left( \frac{1}{2} \|y\|^2 + 1 \right) \rightarrow 0,$$

so that  $\Phi'_{\mathcal{G}_n} b^{(n)}$  (strongly) converges towards 0 in  $L^2(\mathbb{T})$ . Moreover,  $\Phi_{\mathcal{G}_n} a^{(n)} = \Phi(\sum_{i=0}^{P_n-1} a_i^{(n)} \delta_{ih_n})$  and for all  $\psi \in C(\mathbb{T})$ ,

$$\begin{aligned} \left| \left\langle \sum_{i=0}^{P_n-1} a_i^{(n)} \delta_{ih_n + b_i/a_i} - \sum_{i=0}^{P_n-1} a_i \delta_{ih_n}, \psi \right\rangle \right| &= \left| \sum_{i=0}^{P_n-1} a_i^{(n)} (\psi(ih_n + b_i/a_i) - \psi(ih_n)) \right| \\ &\leq \sum_{i=0}^{P_n-1} a_i^{(n)} \omega_\psi \left( \frac{h_n}{2} \right) \rightarrow 0 \end{aligned}$$

where  $\omega_\psi : t \mapsto \sup_{|x'-x| \leq t} |\psi(x) - \psi(x')|$  is the modulus of continuity of  $\psi$ . As a result,  $\sum_{i=0}^{P_n-1} a_i^{(n)} \delta_{ih_n} - m^n \xrightarrow{*} 0$  and  $\sum_{i=0}^{P_n-1} a_i^{(n)} \delta_{ih_n}$  weakly\* converges to  $m$ . Hence,  $\Phi_{\mathcal{G}_n} a^{(n)}$  weakly converges towards  $\Phi m$  in  $L^2(\mathbb{T})$ . To sum up,  $\Phi_{\mathcal{G}_n} a^{(n)} + \Phi'_{\mathcal{G}_n} b^{(n)} - y$  weakly converges towards  $\Phi m - y$  and we conclude as before.

For the limsup inequality, the only difference is in the construction of  $m^n$  for the limsup inequality: it is sufficient to choose  $a_k^{(n)} = m([kh_n, (k+1)h_n))$  and  $b_k^{(n)} = 0$  for all  $k \in \llbracket 0, P_n - 1 \rrbracket$ .  $\square$

## 5.3. Asymptotics of the support: generalities

Though Proposition 11 states the convergence of the solutions of  $(\mathcal{Q}_\lambda^n(y))$  towards those of  $(\mathcal{Q}_\lambda^\infty(y))$ , it does not describe the supports of the solutions. We now study the convergence of those supports using dual certificates and the optimality conditions (Proposition 8). In this context, a dual certificate is determined by a function  $\mu = \Phi^*q \in C(\mathbb{T})$  where  $q \in L^2(\mathbb{T})$ , and

$$I^{(\mathbf{r})} = \left\{ i \in \llbracket 0, P_n - 1 \rrbracket ; b_i > -\frac{h_n}{2}a_i \right\} \subset \left\{ i \in \llbracket 0, P_n - 1 \rrbracket ; \left( \mu + \frac{h_n}{2}\mu' \right) (ih_n) = 1 \right\},$$

$$I^{(1)} = \left\{ i \in \llbracket 0, P_n - 1 \rrbracket ; b_i < \frac{h_n}{2}a_i \right\} \subset \left\{ i \in \llbracket 0, P_n - 1 \rrbracket ; \left( \mu - \frac{h_n}{2}\mu' \right) (ih_n) = 1 \right\}.$$

To sum up, we shall exploit the following observations

- if  $(\mu + \frac{h_n}{2}\mu') (ih_n) = 1$  but  $(\mu - \frac{h_n}{2}\mu') (ih_n) < 1$ , a spike may appear at  $ih_n + \frac{h_n}{2}$ ,
- if  $(\mu - \frac{h_n}{2}\mu') (ih_n) = 1$  but  $(\mu + \frac{h_n}{2}\mu') (ih_n) < 1$ , a spike may appear at  $ih_n - \frac{h_n}{2}$ ,
- if  $(\mu + \frac{h_n}{2}\mu') (ih_n) = 1$  and  $(\mu - \frac{h_n}{2}\mu') (ih_n) = 1$ , a spike may appear anywhere in the interval  $[ih_n - \frac{h_n}{2}, ih_n + \frac{h_n}{2}]$ .

The following lemma is central in our analysis. We consider a sequence of functions  $(\mu^n)_{n \in \mathbb{N}}$  and for  $0 < r < \frac{1}{2} \min_{\nu \neq \nu'} |x_\nu - x_{\nu'}|$ ,  $\nu \in \{1, \dots, N\}$ , we study:

$$S_{n,\nu}^{(\mathbf{r})}(r) \stackrel{\text{def.}}{=} \left\{ t \in \mathcal{G}_n \cap (x_\nu - r, x_\nu + r) ; \left( \mu^n + \frac{h_n}{2}\mu^{n'} \right) (t) = 1 \right\},$$

$$S_{n,\nu}^{(1)}(r) \stackrel{\text{def.}}{=} \left\{ t \in \mathcal{G}_n \cap (x_\nu - r, x_\nu + r) ; \left( \mu^n - \frac{h_n}{2}\mu^{n'} \right) (t) = 1 \right\}.$$

**Lemma 3.** *Let  $(x_1, \dots, x_N) \in \mathbb{T}^N$  pairwise distinct, and let  $\{\mu^n\}_{n \in \mathbb{N}} \in (C^3(\mathbb{T}))^{\mathbb{N}}$  be a sequence of functions which converges uniformly towards some  $\mu^\infty$  (and similarly for the derivatives) such that for all  $\nu \in \{1, \dots, N\}$ ,  $\mu^\infty(x_\nu) = 1$  and for all  $t \in \mathbb{T} \setminus \{x_1, \dots, x_N\}$ ,  $\mu^\infty(t) < 1$ .*

(i) *Then*

$$\limsup_{n \rightarrow +\infty} \left\{ t \in \mathcal{G}_n ; \mu^n(t) + \frac{h_n}{2}|\mu^{n'}(t)| = 1 \right\} \subset \{x_1, \dots, x_N\}. \quad (80)$$

*In particular there exists  $n_0 \in \mathbb{N}$  such that for  $n \geq n_0$*

$$\left\{ t \in \mathcal{G}_n ; \mu^n(t) + \frac{h_n}{2}|\mu^{n'}(t)| = 1 \right\} = \bigcup_{\nu=1}^N (S_{n,\nu}^{(\mathbf{r})}(r) \cup S_{n,\nu}^{(1)}(r)) \subset \bigcup_{\nu=1}^N (x_\nu - r, x_\nu + r).$$

*Assume moreover that for all  $n \in \mathbb{N}$  and all  $t \in \mathcal{G}_n$ ,  $\mu^n(t) + \frac{h_n}{2}|\mu^{n'}(t)| \leq 1$ . For each  $\nu \in \{1, \dots, N\}$ :*

(ii) *If  $(\mu^\infty)''(x_\nu) \neq 0$ , then there exists  $n_0 \in \mathbb{N}$  such that for  $n \geq n_0$ , each set  $S_{n,\nu}^{(\mathbf{r})}(r)$  and  $S_{n,\nu}^{(1)}(r)$  is of the form  $\emptyset$ ,  $\{ih_n\}$ , or  $\{ih_n, (i+1)h_n\}$ , and if both sets are nonempty:*

$$\max S_{n,\nu}^{(\mathbf{r})}(r) \leq \min S_{n,\nu}^{(1)}(r).$$



- (iii) If  $(\mu^\infty)^{(3)}(x_\nu) \neq 0$ , then there exists  $n_0 \in \mathbb{N}$  such that for each  $n \geq n_0$ ,  $S_{n,\nu}^{(\mathbf{r})}(r) = \emptyset$  or  $S_{n,\nu}^{(1)}(r) = \emptyset$ .
- (iv) If  $(\mu^\infty)^{(4)}(x_\nu) \neq 0$ , the set of  $n \in \mathbb{N}$  such that  $S_{n,\nu}^{(\mathbf{r})}(r) = \{(i_n - 1)h_n, i_n h_n\}$  and  $S_{n,\nu}^{(1)}(r) = \{i_n h_n, (i_n + 1)h_n\}$  (with the same  $i_n \in \llbracket 0, P_n - 1 \rrbracket$ ) is finite.

*Proof.* Observe that both  $\mu^n + \frac{h_n}{2}\mu^{n'}$  and  $\mu^n - \frac{h_n}{2}\mu^{n'}$  converge uniformly towards  $\mu^\infty$  as  $n \rightarrow +\infty$  (and similarly for the derivatives).

- (i) For all  $\tilde{r} \in (0, r)$ , by compactness,  $\sup\{\mu^\infty(t); t \in \mathbb{T} \setminus \bigcup(x_\nu - \tilde{r}, x_\nu + \tilde{r})\} < 1$ . Thus by uniform convergence there exists  $n_0 \in \mathbb{N}$  such that for all  $n \geq n_0$ ,  $(\mu^n \pm \frac{h_n}{2}\mu^{n'}) < 1$  on  $\mathbb{T} \setminus \bigcup_{\nu=1}^N(x_\nu - \tilde{r}, x_\nu + \tilde{r})$ , and the first claim is proved.
- (ii) If moreover  $(\mu^\infty)''(x_\nu) \neq 0$ , it is in fact negative. Choosing  $\tilde{r} \in (0, r)$  small enough and then  $n$  large enough, we may assume that  $\mu^{n''} < -k_0$  in  $(x_\nu - \tilde{r}, x_\nu + \tilde{r})$ , for some  $k_0 > 0$ , and by (80) that  $S_{n,\nu}^{(\mathbf{r})}(r) \cup S_{n,\nu}^{(1)}(r) \subset (x_\nu - \tilde{r}, x_\nu + \tilde{r})$ . By uniform convergence,  $\mu^{n''} + \frac{h_n}{2}|\mu^{n(3)}| < -\frac{k_0}{2}$  for  $n$  large enough, so that both functions  $\mu^n + \frac{h_n}{2}\mu^{n'}$  and  $\mu^n - \frac{h_n}{2}\mu^{n'}$  are strictly concave in  $(x_\nu - \tilde{r}, x_\nu + \tilde{r})$ . This implies that  $S_{n,\nu}^{(\mathbf{r})}(r)$  (resp.  $S_{n,\nu}^{(1)}(r)$ ) is of the form  $\emptyset$ ,  $\{ih_n\}$ , or  $\{ih_n, (i+1)h_n\}$ . Observe also that  $\mu^n + \frac{h_n}{2}\mu^{n'} - (\mu^n - \frac{h_n}{2}\mu^{n'}) = h_n\mu^{n'}$ . Since the function  $\mu^{n'}$  is strictly decreasing in  $(x_\nu - \tilde{r}, x_\nu + \tilde{r})$ , it vanishes at most once. If  $S_{n,\nu}^{(\mathbf{r})}(r) \neq \emptyset$  and  $S_{n,\nu}^{(1)}(r) \neq \emptyset$ , it must change sign in  $(x_\nu - \tilde{r}, x_\nu + \tilde{r})$  and thus it vanishes exactly once, at some  $\xi \in (x_\nu - \tilde{r}, x_\nu + \tilde{r})$ . Then for  $t \in (x_\nu - \tilde{r}, \xi)$ ,

$$(\mu^n - \frac{h_n}{2}\mu^{n'})(t) = (\mu^n + \frac{h_n}{2}\mu^{n'})(t) - h_n(\mu^{n'})'(t) \leq 1 - h_n(\mu^{n'})'(t) < 1$$

so that  $\min S_{n,\nu}^{(1)}(r) \geq \xi$ . Similarly  $\max S_{n,\nu}^{(\mathbf{r})}(r) \leq \xi$ .

- (iii) By contradiction, assume that the set of  $n' \in \mathbb{N}$  such that  $S_{n',\nu}^{(\mathbf{r})}(r) \neq \emptyset$  and  $S_{n',\nu}^{(1)}(r) \neq \emptyset$  is infinite. We may extract a subsequence  $n = n'(m)$  such that there exists  $i_n, j_n \in \llbracket 0, P_n - 1 \rrbracket$  (denoted hereafter  $i, j$ ) with  $ih_n \in S_{n,\nu}^{(\mathbf{r})}(r_m)$ ,  $jh_n \in S_{n,\nu}^{(1)}(r_m)$ . Combining the Taylor expansions of  $\mu^n$  and  $(\mu^n)'$  around  $ih_n$  (resp.  $jh_n$ ), we get

$$\begin{aligned} 1 &\geq \mu^n((i+1)h_n) - \frac{h_n}{2}\mu^{n'}((i+1)h_n) \\ &= \underbrace{\mu^n(ih_n) + h_n\mu^{n'}(ih_n)(1 - \frac{1}{2})}_{=1} + \underbrace{h_n^2\mu^{n''}(ih_n)\left(\frac{1}{2!} - \frac{1}{2}\right)}_{=0} + h_n^3\mu^{n(3)}(ih_n)\alpha_3 \\ &\quad + h_n^4 \int_0^1 \mu^{n(4)}(ih_n + th_n) \left( \frac{(1-t)^3}{3!} - \frac{(1-t)^2}{2! \times 2} \right) dt, \text{ and} \\ 1 &\geq \mu^n((j-1)h_n) + \frac{h_n}{2}\mu^{n'}((j-1)h_n) \\ &= \underbrace{\mu^n(jh_n) - h_n\mu^{n'}(jh_n)(1 - \frac{1}{2})}_{=1} + \underbrace{h_n^2\mu^{n''}(jh_n)\left(\frac{1}{2!} - \frac{1}{2}\right)}_{=0} - h_n^3\mu^{n(3)}(jh_n)\alpha_3 \\ &\quad + h_n^4 \int_0^1 \mu^{n(4)}(jh_n - th_n) \left( \frac{(1-t)^3}{3!} - \frac{(1-t)^2}{2! \times 2} \right) dt \end{aligned}$$

where  $\alpha_k$  is defined in (96). Now, let  $n \rightarrow +\infty$ . By (80),  $ih_n \rightarrow x_\nu$  and  $jh_n \rightarrow x_\nu$ , and using the uniform convergence of  $(\mu^n)^{(k)}$  towards  $(\mu^\infty)^{(k)}$ , dividing by  $h_n^3$ , we obtain respectively  $0 \geq -(\mu^\infty)^{(3)}(x_\nu) \times \frac{1}{12}$  and  $0 \geq (\mu^\infty)^{(3)}(x_\nu) \times \frac{1}{12}$ , thus  $(\mu^\infty)^{(3)}(x_\nu) = 0$ .

(iv) Assume by contradiction, that the mentioned set is infinite. For such  $n$ , a Taylor expansion at  $ih_n$  yields (we write  $i$  for  $i_n$ ):

$$\begin{aligned} 1 &= \mu^n((i+1)h_n) - \frac{h_n}{2} \mu^{n'}((i+1)h_n) \\ &= \underbrace{\mu^n(ih_n) + \frac{h_n}{2} \mu^{n'}(ih_n)}_{=1} + \gamma_3 h_n^3 \mu^{n(3)}(ih_n) + \gamma_4 h_n^4 \mu^{n(4)}(ih_n) \\ &\quad + h_n^5 \int_0^1 \mu^{n(5)}(ih_n + th_n) \left( \frac{(1-t)^4}{4!} - \frac{(1-t)^3}{3! \times 2} \right) dt, \text{ and} \\ 1 &= \mu^n((i-1)h_n) + \frac{h_n}{2} \mu^{n'}((i-1)h_n) \\ &= \underbrace{\mu^n(ih_n) - \frac{h_n}{2} \mu^{n'}(ih_n)}_{=1} - \gamma_3 h_n^3 \mu^{n(3)}(ih_n) + \gamma_4 h_n^4 \mu^{n(4)}(ih_n) \\ &\quad + h_n^5 \int_0^1 \mu^{n(5)}(ih_n + th_n) \left( \frac{(1-t)^4}{4!} - \frac{(1-t)^3}{3! \times 2} \right) dt, \end{aligned}$$

with  $\gamma_k = \frac{1}{k!} - \frac{1}{(k-1)! \times 2}$ . Summing both equalities, dividing by  $h_n^4$  and taking the limit  $n \rightarrow +\infty$  yields  $(\mu^\infty)^{(4)}(x_\nu) = 0$ .

□

This other lemma focusses on the limit of the sets  $D^n$  defined in (75).

**Lemma 4.** *As  $n \rightarrow +\infty$ , the sets  $D^n$  converge towards  $D^\infty$  defined in (77) (in the sense of set convergence).*

*Proof.* We observe that  $E^n \subset D^n \subset F^n$ , where

$$\begin{aligned} E^n &\stackrel{\text{def.}}{=} \left\{ q \in L^2(\mathbb{T}) ; \max_{t \in \mathbb{T}} (\Phi^* q)(t) + \frac{h_n}{2} |(\Phi^* q)'(t)| \leq 1 \right\}, \\ F^n &\stackrel{\text{def.}}{=} \left\{ q \in L^2(\mathbb{T}) ; \max_{k \in \llbracket 0, P_n-1 \rrbracket} \Phi^* q(kh_n) \leq 1 \right\} \end{aligned}$$

so that it suffices to prove that  $E^n$  and  $F^n$  converge towards  $D^\infty$ . On the one hand, it is clear that  $D^\infty = \bigcap_{n \in \mathbb{N}} F^n$ , and the sequence  $F^n$  is non-increasing. On the other hand, it is possible to check that  $D^\infty = \overline{\bigcup_{n \in \mathbb{N}} E^n}$ , and the sequence  $E^n$  is non-decreasing. As a consequence, the claimed set convergences hold. □

5.4. Asymptotics of the support for fixed  $\lambda > 0$ 

Let us recall that the dual problem to  $(\mathcal{Q}_\lambda^n(y))$  is the projection onto the closed convex set

$$D^n \stackrel{\text{def.}}{=} \left\{ q \in L^2(\mathbb{T}) ; (\Phi^*q)(ih_n) + \frac{h_n}{2}|(\Phi^*q)'(ih_n)| \leq 1 \right\}.$$

Since the set convergence of  $D^n$  (see Lemma 4) implies the convergence of the projections onto  $D^n$  (see [29], or [21] for a direct proof in a similar context), we obtain:

**Proposition 12.** *Let  $q_\lambda^n$  (resp.  $q_\lambda^\infty$ ) be a solution of  $(\mathcal{E}_\lambda^n(y))$  (resp.  $(\mathcal{E}_\lambda^\infty(y))$ ), and  $\mu_\lambda^n = \Phi^*q_\lambda^n$  (resp.  $\mu_\lambda^\infty = \Phi^*q_\lambda^\infty$ ). Then*

$$\begin{aligned} \lim_{n \rightarrow +\infty} q_\lambda^n &= q_\lambda^\infty \text{ for the } L^2(\mathbb{T}) \text{ strong topology,} \\ \lim_{n \rightarrow +\infty} \mu_\lambda^{n(k)} &= \mu_\lambda^{\infty(k)} \text{ in the sense of the uniform convergence, for all } k \in \mathbb{N}. \end{aligned}$$

The following proposition states that in the generic case, one may observe up to two pairs of spikes for each spike of the solution of the positive Beurling-lasso. As before,  $r$  is chosen such that  $0 < r < \frac{1}{2} \min_{\nu \neq \nu'} |x_\nu - x_{\nu'}|$ .

**Proposition 13.** *Let  $\lambda > 0$ , and assume that there exists a solution to  $(\mathcal{Q}_\lambda^\infty(y))$  which is a sum of a finite number of (positive) Dirac masses:  $m_\lambda^\infty = \sum_{\nu=1}^N \alpha_\nu \delta_{x_\nu}$  where  $\alpha_\nu > 0$ . Assume that  $\mu_\lambda^\infty$  satisfies  $|\mu_\lambda^\infty(t)| < 1$  for all  $t \in \mathbb{T} \setminus \{x_1, \dots, x_N\}$ .*

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

$$\limsup_{n \rightarrow +\infty} (\text{supp } m_\lambda^n) \subset \{x_1, \dots, x_N\}.$$

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

$$\lim_{n \rightarrow +\infty} (\text{supp}(m_\lambda^n)) = \{x_1, \dots, x_N\}. \quad (81)$$

*If, additionally,  $(\mu_\lambda^\infty)''(x_\nu) \neq 0$  for some  $\nu \in \{1, \dots, N\}$ , then for all  $n$  large enough, the restriction of  $m_\lambda^n$  to  $(x_\nu - r, x_\nu + r)$  is a sum of Dirac masses whose configuration is given in Table 1, and if  $(\mu_\lambda^\infty)^{(3)}(x_\nu) \neq 0$ , then only the cases indicated with (\*) may appear.*

*Proof.* By Proposition 12, we know that the dual certificates  $\mu_\lambda^n$  converge towards  $\mu_\lambda^\infty$ . By Lemma 3 and the optimality conditions, we have thus  $\limsup_{n \rightarrow +\infty} (\text{supp}(m_\lambda^n)) \subset \{x_1, \dots, x_N\}$ . If  $m_\lambda^\infty$  is the unique solution, assume by contradiction that  $\liminf (\text{supp}(m_\lambda^n)) \subsetneq \{x_1, \dots, x_N\}$ . Then there is some  $\nu$ , some  $\varepsilon > 0$  such that (up to a subsequence)  $(\text{supp}(m_\lambda^n)) \cap (x_\nu - \varepsilon, x_\nu + \varepsilon) = \emptyset$ . This contradicts the  $\Gamma$ -convergence result (Prop. 11) which ensures that  $m_\lambda^n$  converges towards  $m_\lambda^\infty$  for the weak\* topology. As a result  $\lim_{n \rightarrow +\infty} (\text{supp}(m_\lambda^n)) = \{x_1, \dots, x_N\}$ .

If  $(\mu_\lambda^\infty)''(x_\nu) \neq 0$ , Lemma 3 ensures that the sets  $S_{n,\nu}^{(r)}(r)$  and  $S_{n,\nu}^{(l)}(r)$  are of the form  $\emptyset$ ,  $\{ih_n\}$ , or  $\{ih_n, (i+1)h_n\}$ . Moreover, since  $\lim_{n \rightarrow +\infty} (\text{supp } m_\lambda^n) = \{x_1, \dots, x_N\}$

Number of Dirac masses	Saturations of the certificates ( $S_{n,\nu}^{(r)}$ / $S_{n,\nu}^{(l)}$ )	Possible Dirac Locations
<b>One</b>	$\{ih_n\}/\emptyset$ or $\emptyset/\{ih_n\}$  $\{ih_n\}/\{ih_n\}$	$ih_n + \varepsilon_n \frac{h_n}{2}$ , with $\varepsilon_n \in \{-1, 1\}$ (*)  $ih_n + t_i$ , with $-\frac{h_n}{2} \leq t_i \leq \frac{h_n}{2}$
<b>Two</b>	$\{(i-1)h_n, ih_n\}/\emptyset$ or $\emptyset/\{ih_n, (i+1)h_n\}$  $\{ih_n\}/\{jh_n\}$  $\{ih_n\}/\{jh_n, (i+1)h_n\}$ or $\{(i-1)h_n, ih_n\}/\{ih_n\}$	$\left((i-\varepsilon_n)h_n + \varepsilon_n \frac{h_n}{2}, ih_n + \varepsilon_n \frac{h_n}{2}\right)$ , with $\varepsilon_n \in \{-1, 1\}$ (*)  $\left(ih_n + \frac{h_n}{2}, jh_n - \frac{h_n}{2}\right)$ , $i < j$  $\left((i-\varepsilon_n)h_n + \varepsilon_n \frac{h_n}{2}, ih_n + t_i\right)$ , $\varepsilon_n \in \{-1, 1\}$ , $-\frac{h_n}{2} \leq t_i \leq \frac{h_n}{2}$
<b>Three</b>	$\{ih_n\}/\{jh_n, (j+1)h_n\}$  $\{(i-1)h_n, ih_n\}/\{jh_n\}$  $\{(i-1)h_n, ih_n\}/\{ih_n, (i+1)h_n\}$	$\left(ih_n + \frac{h_n}{2}, jh_n - \frac{h_n}{2}, (j+1)h_n - \frac{h_n}{2}\right)$ , with $i < j$  $\left((i-1)h_n + \frac{h_n}{2}, ih_n + \frac{h_n}{2}, jh_n - \frac{h_n}{2}\right)$ , with $i < j$  $\left((i-1)h_n + \frac{h_n}{2}, ih_n + t_i, (i+1)h_n - \frac{h_n}{2}\right)$ , $-\frac{h_n}{2} \leq t_i \leq \frac{h_n}{2}$
<b>Four</b>	$\{(i-1)h_n, ih_n\}/\{jh_n, (j+1)h_n\}$	$\left((i-1)h_n + \frac{h_n}{2}, ih_n + \frac{h_n}{2}, jh_n - \frac{h_n}{2}, (j+1)h_n - \frac{h_n}{2}\right)$ , $i < j$

Table 1: Number of Dirac masses that may appear if  $(\mu_\lambda^\infty)''(x_\nu) \neq 0$ . For the sake of the simplicity of the table, and since we focus on the saturations of dual certificates, we regard sums like  $\delta_{ih_n+h_n/2} + \delta_{(i+1)h_n-h_n/2}$  as “two” Dirac masses.

we must have  $S_{n,\nu}^{(r)}(r) \neq \emptyset$  or  $S_{n,\nu}^{(l)}(r) \neq \emptyset$ . Using the fact that  $\max S_{n,\nu}^{(r)}(r) \leq \min S_{n,\nu}^{(l)}(r)$ , one may check that the only possible saturation points of  $\mu_\lambda^n + \frac{h_n}{2}\mu_\lambda^{n'}$  and  $\mu_\lambda^n - \frac{h_n}{2}\mu_\lambda^{n'}$  are given in Table 1. The optimality conditions of Proposition 8 imply that  $m_\lambda^n$  is at most a sum of Dirac masses at those locations.

If  $(\mu_\lambda^\infty)^{(3)}(x_\nu) \neq 0$  the third point of Lemma 3 implies that for  $n$  large enough,  $S_{n,\nu}^{(r)}(r) = \emptyset$  or  $S_{n,\nu}^{(l)}(r) = \emptyset$  (but not both). Hence there are at most two (successive) saturations, produced either by  $\mu_\lambda^n + \frac{h_n}{2}\mu_\lambda^{n'}$  or by  $\mu_\lambda^n - \frac{h_n}{2}\mu_\lambda^{n'}$ .  $\square$

*Remark 10.* Proposition 13 states that the support of the C-BP on thin grids actually depends on the properties of the dual certificate  $\mu_\lambda^\infty$  of the (positive) Beurling LASSO. The condition  $(\mu_\lambda^\infty)''(x_\nu) \neq 0$  seems to be overwhelming, if not generic, and it is ensured for instance if  $\lambda$  is small and the Non-Degenerate Source Condition holds (see [21]). As for the condition  $(\mu_\lambda^\infty)^{(3)}(x_\nu) \neq 0$ , it also seems to be generic, as there is nothing to impose  $(\mu_\lambda^\infty)^{(3)}(x_\nu) = 0$  in the positive Beurling LASSO. As a result, in practice, *one does not observe all the configurations given in Table 1, and only the cases indicated with (\*) appear*, the case of two spikes being again overwhelming.

This means that when approximating the positive Beurling LASSO with the Continuous Basis-Pursuit, one generally sees two spikes instead of one, and those spikes are at successive half-grid points:  $(ih + \frac{h}{2}, (i+1)h + \frac{h}{2})$  or  $(ih - \frac{h}{2}, (i+1)h - \frac{h}{2})$ .

## 5.5. Asymptotic of the low noise support

Now, we focus on the behavior of the Continuous Basis Pursuit at low noise. As for the LASSO, this analysis is more difficult in whole generality, since it involves the minimal norm solutions of nonlinear problems, in which it is difficult to pass to the limit. Therefore, we are led to assume that  $\{x_{0,1}, \dots, x_{0,N}\} \subset \mathcal{G}_n$ , and the measure now reads  $m_0 = \sum_{i=0}^{P_n-1} a_{0,i} \delta_{ih_n}$ .

The following property ensures that  $m_0$  is a solution to  $(\mathcal{Q}_0^n(y_0))$  for each  $n$  large enough.

**Lemma 5.** *Assume that there exists a function  $\mu \in \text{Im } \Phi^*$ , such that for all  $t \in \mathbb{T} \setminus \{x_{0,1}, \dots, x_{0,N}\}$ ,  $\mu(t) < 1$  and*

$$\forall \nu \in \{1, \dots, N\}, \mu(x_{0,\nu}) = 1, \mu''(x_{0,\nu}) \neq 0, \mu^{(3)}(x_{0,\nu}) = 0, \mu^{(4)}(x_{0,\nu}) > 0. \quad (82)$$

*Then, for all  $n$  large enough,  $\mu$  is a dual certificate for  $\bar{m}_0 = \sum_{\nu=1}^N \alpha_{0,\nu} \delta_{x_{0,\nu}}$  for  $(\mathcal{Q}_0^n(y_0))$ , and  $m_0$  is a solution to  $(\mathcal{Q}_0^n(y_0))$ . Moreover, if  $\Gamma_{x_0}$  has full rank, this solution is unique.*

*Remark 11.* The condition  $\mu^{(3)}(x_{0,\nu}) = 0$  is natural since our aim is to build a certificate which is valid for all  $n$ , hence Lemma 3 applies with  $S_{n,\nu}^{(r)}(r) \neq \emptyset$  and  $S_{n,\nu}^{(1)}(r) \neq \emptyset$ .

*Proof.* Let  $\nu \in \{1, \dots, N\}$  and  $r_\nu \in (0, r)$  such that  $\mu''(t) < 0$ , and  $\mu^{(4)}(t) > 0$  in  $(x_{0,\nu} - r_\nu, x_{0,\nu} + r_\nu)$ . We shall prove that  $\mu(kh_n) + \frac{h_n}{2} |\mu'(kh_n)| < 1$  for all  $k$  such that  $kh_n \in (x_{0,\nu} - r_\nu, x_{0,\nu} + r_\nu) \setminus \{x_{0,\nu}\}$ . To simplify the notation, we assume without loss of generality that  $x_{0,\nu} = 0$  and we write  $\tilde{r} = r_\nu$ . The variations of  $\mu$  and its derivatives are given by the table below:

$t$	$-\tilde{r}$	$0$	$\tilde{r}$
$\mu^{(4)}$		+	
$\mu^{(3)}$		0	→
$\mu''$	$\mu''(-\tilde{r}) < 0$	$\mu''(0)$	$\mu''(\tilde{r}) < 0$
$\mu'$		0	→
$\mu$		1	→

Let us observe that the function  $\theta : t \mapsto \mu(t) - \frac{t}{2} \mu'(t)$  is (strictly) decreasing in  $[0, \tilde{r})$ , since

$$\forall t \in (0, \tilde{r}), \theta'(t) = \frac{1}{2} (\mu'(t) - t\mu''(t)) = \frac{1}{2} \int_0^t \underbrace{(\mu''(u) - \mu''(t))}_{<0} du < 0. \quad (83)$$

Hence, for all  $k$  such that  $kh_n \in (0, \tilde{r})$ ,

$$\mu(kh_n) - \frac{h_n}{2}\mu'(kh_n) = \underbrace{\mu(kh_n) - \frac{kh_n}{2}\mu'(kh_n)}_{=\theta(kh_n) < \theta(0)=1} + \underbrace{\frac{(k-1)h_n}{2}\mu'(kh_n)}_{<0} < 1. \quad (84)$$

On the other hand,  $\theta$  is (strictly) increasing on  $(-\tilde{r}, 0]$  since

$$\forall t \in (-\tilde{r}, 0), \quad \theta'(t) = \frac{1}{2}(\mu'(t) - t\mu''(t)) = \frac{1}{2} \int_0^t \underbrace{(\mu''(u) - \mu''(t))}_{<0} du > 0. \quad (85)$$

As a consequence, for all  $k$  such that  $kh_n \in (-\tilde{r}, 0)$ ,

$$\mu(kh_n) + \frac{h_n}{2}\mu'(kh_n) = \underbrace{\mu(kh_n) - \frac{kh_n}{2}\mu'(kh_n)}_{=\theta(kh_n) < \theta(0)=1} + \underbrace{\frac{(k+1)h_n}{2}\mu'(kh_n)}_{\leq 0} < 1. \quad (86)$$

Thus we see that  $\mu(kh_n) + \frac{h_n}{2}|\mu'(kh_n)| < 1$  for all  $kh_n \in (-\tilde{r}, \tilde{r}) \setminus \{0\}$ , and we proceed similarly on all the intervals of the form  $(x_{0,\nu} - r_\nu, x_{0,\nu} + r_\nu)$ . By a compactness argument, there exists a constant  $\beta < 1$  such that  $\mu(t) \leq \beta$  for all  $t \in \mathbb{T} \setminus \bigcup_{\nu=1}^N (x_{0,\nu} - r_\nu, x_{0,\nu} + r_\nu)$ . For  $n$  large enough, the inequality  $\frac{h_n}{2}(\sup_{t \in \mathbb{T}} |\mu'(t)|) < 1 - \beta$  holds, and we see that  $\mu(kh_n) + \frac{h_n}{2}|\mu'(kh_n)| < 1$  for all  $t \in \mathbb{T} \setminus \bigcup_{\nu=1}^N (x_{0,\nu} - r_\nu, x_{0,\nu} + r_\nu)$ .

As a conclusion, we see that  $\mu$  is a valid certificate for  $(a_0, 0)$  (see the optimality conditions of Proposition 8), thus  $(a_0, 0)$  is a solution of  $(\mathcal{Q}_0^n(y_0))$ .  $\square$

Now, we consider the limit of the minimal norm solutions of  $(\mathcal{E}_0^n(y_0))$ . In general, they do not converge towards the minimal norm solution of  $(\mathcal{E}_0^\infty(y_0))$ , and we are led to introduce a new variational problem to carry the study further.

**Definition 7** (Third derivative precertificate). *Given  $m_0 \in \mathcal{M}(\mathbb{T})$ , we define the third derivative precertificate as  $\mu_T \stackrel{\text{def.}}{=} \Phi^* q_T$  where*

$$q_T \stackrel{\text{def.}}{=} \underset{q \in L^2(\mathbb{T})}{\operatorname{argmin}} \left\{ \|q\|_2; \forall i \in \{1, \dots, N\}, (\Phi^* q)(x_i) = 1, \right. \\ \left. (\Phi^* q)'(x_i) = 0 \text{ and } (\Phi^* q)^{(3)}(x_i) = 0 \right\}, \quad (87)$$

whenever the above set is not empty.

It is clear that the set defined in (87) is a closed convex set. It is nonempty for instance if the conditions of Lemma 5 hold. Note that  $q_T$  corresponds to a quadratic minimization under linear constraint, and can hence be computed by solving a linear system,

$$q_T = \begin{pmatrix} \Gamma_{x_0}^* \\ \Phi_{x_0}^{(3)*} \end{pmatrix}^+ \begin{pmatrix} \mathbf{1}_N \\ 0 \\ 0 \end{pmatrix} = \Gamma_{x_0}^{+,*} \begin{pmatrix} \mathbf{1}_N \\ 0 \end{pmatrix} - \tilde{\Pi} \Phi_{x_0}^{(3)*} (\Phi_{x_0}^{(3)*} \tilde{\Pi} \Phi_{x_0}^{(3)*})^{-1} \Phi_{x_0}^{(3)*} \Gamma_{x_0}^{+,*} \begin{pmatrix} \mathbf{1}_N \\ 0 \end{pmatrix} \quad (88)$$

$$= p_V^\infty - \tilde{\Pi} \Phi_{x_0}^{(3)*} (\Phi_{x_0}^{(3)*} \tilde{\Pi} \Phi_{x_0}^{(3)*})^{-1} \Phi_{x_0}^{(3)*} p_V^\infty, \quad (89)$$

where  $\tilde{\Pi}$  is the orthogonal projector onto  $(\text{Im } \Gamma_{x_0})^\perp$ , and  $\Gamma_{x_0} = \begin{pmatrix} \Phi_{x_0} & \Phi'_{x_0} \end{pmatrix}$ .

**Definition 8** (Twice Non-Degenerate Source Condition). *We say that  $m_0$  satisfies the Twice Non-Degenerate Source Condition (TNDSC) if  $q_T$  in (87) is well defined and if it satisfies, for  $\mu_T = \Phi^* q_T$ ,*

$$\begin{aligned} \forall t \in \mathbb{T} \setminus \{x_{0,1}, \dots, x_{0,N}\}, \quad \mu_T(t) < 1, \\ \forall \nu \in \{1, \dots, N\}, \quad \mu_T''(x_{0,\nu}) < 0 \quad \text{and} \quad \mu_T^{(4)}(x_{0,\nu}) > 0. \end{aligned}$$

Observe that if the Twice Non-Degenerate Source Condition holds, the hypotheses of Lemma 5 are satisfied and  $m_0$  is a solution to  $(\mathcal{Q}_0^n(y_0))$  for  $n$  large enough. In fact the associated minimal norm certificates (which thus exist) converge towards  $\mu_T$ .

**Proposition 14.** *Let  $m_0 \in \mathcal{M}(\mathbb{T})$  satisfy the Twice Non-Degenerate Source Condition (and  $\mu_T$  the corresponding Third derivative (pre)certificate). Let  $q_0^n$  be the minimal norm solution of  $(\mathcal{E}_0^n(y_0))$ , and  $\mu_0^n = \Phi^* q_0^n$ . Then,*

$$\lim_{n \rightarrow +\infty} q_0^n = q_T \text{ for the } L^2(\mathbb{T}) \text{ strong topology,} \quad (90)$$

$$\lim_{n \rightarrow +\infty} \mu_0^{n(k)} = \mu_T^{(k)} \text{ in the sense of the uniform convergence, for all } k \in \mathbb{N}. \quad (91)$$

*Proof.* As mentioned above, the Twice Non-Degenerate Source Condition implies that  $\mu_T$  is a function admissible for Lemma 5, hence a certificate for  $(\mathcal{Q}_0^n(y_0))$ . As a result,  $\|q_0^n\|_2 \leq \|q_T\|_2$  and the sequence  $(q_0^n)_{n \in \mathbb{N}}$  is bounded in  $L^2(\mathbb{T})$ . We may extract a subsequence  $q_0^{n'}$  which weakly converges towards some  $\tilde{q} \in L^2(\mathbb{T})$ , and then  $\|\tilde{q}\|_2 \leq \liminf_{n' \rightarrow +\infty} \|q_0^{n'}\|_2 \leq \|q_T\|_2$ . Since  $\Phi^*$  and  $\Phi^{(k),*}$  are compact (see Lemma (6) in Appendix), we obtain that  $\mu_0^{n'(k)} = (\Phi^* q_0^{n'})^{(k)}$  converges toward  $\tilde{\mu} = \Phi^* \tilde{q}$  for the (strong) topology of the uniform convergence. We immediately obtain that  $\tilde{\mu}(t) \leq 1$  for all  $t \in \mathbb{T}$ , and  $\tilde{\mu}(x_{0,\nu}) = 1$ ,  $\tilde{\mu}(x_{0,\nu}) = 0$  for all  $\nu \in \{1, \dots, N\}$ .

Moreover, applying Lemma 3 to  $\Phi^* q_0^n$  (observing that  $x_\nu \in S_{n,\nu}^{(r)}(r) \cap S_{n,\nu}^{(1)}(r)$ ), we get  $\tilde{\mu}^{(3)}(x_{0,\nu}) = 0$ . As a result,  $\tilde{q}$  is admissible for (87), hence  $\|q_T\|_2 \leq \|\tilde{q}\|_2$ . Thus in fact  $\|q_T\|_2 = \|\tilde{q}\|_2$  and  $q_T = \tilde{q}$ . Since the limit of the extracted subsequence does not depend on the choice of the subsequence, in fact the whole sequence converges. Moreover, the convergence is strong in  $L^2(\mathbb{T})$  since  $\lim_{n \rightarrow +\infty} \|q_0^n\|_2 = \|q_T\|_2$ .  $\square$

As a consequence of the above convergence result, the third derivative precertificate controls the extended support on thin grids.

**Proposition 15.** *Let  $m_0 \in \mathcal{M}(\mathbb{T})$  (with  $\{x_{0,1}, \dots, x_{0,N}\} \subset \mathcal{G}_n$ ) such that the Twice Non Degenerate Source Condition holds. Then, for  $n$  large enough,  $m_0$  is a solution to  $(\mathcal{Q}_0^n(y_0))$  and its extended support is given by:*

$$\text{ext}_n^{(r)}(m_0) = \bigcup_{\nu=1}^N S_{n,\nu}^{(r)}(r), \quad \text{and} \quad \text{ext}_n^{(1)}(m_0) = \bigcup_{\nu=1}^N S_{n,\nu}^{(1)}(r), \quad (92)$$

where

- $S_{n,\nu}^{(\mathbf{r})}(r)$  is equal to  $\{x_{0,\nu}\}$  or  $\{x_{0,\nu} - h_n, x_{0,\nu}\}$ ,
- $S_{n,\nu}^{(1)}(r)$  is equal to  $\{x_{0,\nu}\}$  or  $\{x_{0,\nu}, x_{0,\nu} + h_n\}$ .

Moreover, one cannot have simultaneously  $S_{n,\nu}^{(\mathbf{r})}(r) = \{x_{0,\nu} - h_n, x_{0,\nu}\}$  and  $S_{n,\nu}^{(1)}(r) = \{x_{0,\nu}, x_{0,\nu} + h_n\}$ .

*Proof.* By Lemma 5,  $m_0$  is a solution to  $(\mathcal{Q}_0^n(y_0))$  and  $\mu_T$  is a solution to  $(\mathcal{E}_0^n(y_0))$ . Applying Lemma 3 to  $\mu_0^n$ ,  $\mu_T$ , we see that  $S_{n,\nu}^{(\mathbf{r})}(r)$  is of the form  $\emptyset$ ,  $\{ih_n\}$  or  $\{(i-1)h_n, ih_n\}$ , and that  $S_{n,\nu}^{(1)}(r)$  is of the form  $\emptyset$ ,  $\{jh_n\}$  or  $\{jh_n, (j+1)h_n\}$ , with  $i \leq j$ . On the other hand, by the extremality relations between  $\mu_0^n$  (solution of  $(\mathcal{E}_0^n(y_0))$ ) and  $m_0^n$  (solution of  $(\mathcal{Q}_0^n(y_0))$ ),  $x_{0,\nu} \in S_{n,\nu}^{(\mathbf{r})}(r)$  and  $x_{0,\nu} \in S_{n,\nu}^{(1)}(r)$ . As a consequence  $S_{n,\nu}^{(\mathbf{r})}(r)$  is equal to  $\{x_{0,\nu}\}$  or  $\{x_{0,\nu} - h_n, x_{0,\nu}\}$ , and  $S_{n,\nu}^{(1)}(r)$  is equal to  $\{x_{0,\nu}\}$  or  $\{x_{0,\nu}, x_{0,\nu} + h_n\}$ .

Now, since  $\mu_T^4(0) \neq 0$ , the fourth point of Lemma 3 ensures that for  $n$  large enough, one cannot have simultaneously  $S_{n,\nu}^{(\mathbf{r})}(r) = \{x_{0,\nu} - h_n, x_{0,\nu}\}$  and  $S_{n,\nu}^{(1)}(r) = \{x_{0,\nu}, x_{0,\nu} + h_n\}$ .  $\square$

*Remark 12.* As Proposition 15 shows, for each original spike, at most one pair of spikes appears at low noise : the original spike slightly shifted and either the immediate left neighbor shifted by  $+h_n/2$  or the immediate right neighbor shifted by  $-h_n/2$ .

We are now in position to provide a sufficient condition for the spikes to appear in pair, with a prediction on the location of the neighbor.

**Theorem 4.** *Assume that the operator  $\begin{pmatrix} \Phi_{x_0} & \Phi'_{x_0} & \Phi_{x_0}^{(3)} \end{pmatrix}$  has full rank and that the Twice Non Degenerate Source condition (Definiton 8) holds. Moreover, assume that all the components of the natural shift*

$$\rho \stackrel{\text{def.}}{=} (\Phi_{x_0}^{(3)*} \tilde{\Pi} \Phi_{x_0}^{(3)})^{-1} \Phi_{x_0}^{(3)*} \Gamma_{x_0}^{+,*} \begin{pmatrix} \mathbf{1}_N \\ 0 \end{pmatrix} \quad (93)$$

are nonzero. Then, for  $n$  large enough, and all  $\nu \in \{1, \dots, N\}$ ,

$$\text{If } \rho_\nu > 0, \text{ then } S_{n,\nu}^{(\mathbf{r})}(r) = \{x_{0,\nu} - h_n, x_{0,\nu}\}, \quad \text{and } S_{n,\nu}^{(1)}(r) = \{x_{0,\nu}\}, \quad (94)$$

$$\text{If } \rho_\nu < 0, \text{ then } S_{n,\nu}^{(1)}(r) = \{x_{0,\nu}\}, \quad \text{and } S_{n,\nu}^{(\mathbf{r})}(r) = \{x_{0,\nu}, x_{0,\nu} + h_n\}, \quad (95)$$

so that the extended support of  $m_0$  on the grid  $\mathcal{G}_n$  has the form

$$\begin{aligned} \text{ext}_n^{(\mathbf{r})}(m_0) &= \{x_{0,1}, \dots, x_{0,N}\} \cup \{x_{0,\nu} - h_n ; \nu \in \llbracket 1, N \rrbracket \text{ and } \rho_\nu > 0\} \\ \text{ext}_n^{(1)}(m_0) &= \{x_{0,1}, \dots, x_{0,N}\} \cup \{x_{0,\nu} + h_n ; \nu \in \llbracket 1, N \rrbracket \text{ and } \rho_\nu < 0\}. \end{aligned}$$

**Corollary 2.** *Under the hypotheses of Theorem 4, for  $n$  large enough, there exists constants  $C_n^{(1)} > 0$ ,  $C_n^{(2)} > 0$  such that for  $\lambda \leq C_n^{(1)} \min_{1 \leq \nu \leq N} |\alpha_{0,\nu}|$ , and for all  $w \in L^2(\mathbb{T})$  such that  $\|w\|_2 \leq C_n^{(2)} \lambda$ , the solution to  $(\mathcal{Q}_\lambda^n(y))$  is unique, and reads  $m_\lambda = \sum_{\nu=1}^N (\alpha_{\lambda,\nu} \delta_{x_{0,\nu} + t_\nu} + \beta_{\lambda,\nu} \delta_{x_{0,\nu} + \varepsilon_\nu h_n})$ , where*

$$-h_n/2 < t_\nu < h_n/2 \quad \text{and} \quad \varepsilon \stackrel{\text{def.}}{=} -\text{sign}(\rho).$$



*Proof of Theorem 4.* We proceed as in the proof of Theorem 2 by building a good candidate for  $\mu_0^n$  and using Lemma 2. To comply with the notations of Lemma 2, let us write  $\sum_{\nu=1}^N \alpha_{0,i} \delta_{x_0,\nu} = \sum_{k=0}^{P_n-1} a_{0,k} \delta_{kh_n}$ , and  $I^{(\mathbf{r})} \stackrel{\text{def.}}{=} I^{(1)} \stackrel{\text{def.}}{=} I \stackrel{\text{def.}}{=} \{i \in \llbracket 0, P_n - 1 \rrbracket ; a_{0,i} \neq 0\}$ .

For any choice of shift  $(\varepsilon_i)_{i \in I} \in \{-1, +1\}^N$ , we set  $J^{(\mathbf{r})} \stackrel{\text{def.}}{=} I^{(\mathbf{r})} \cup \{i + \varepsilon_i ; i \in I \text{ and } \varepsilon_i = -1\}$  and  $J^{(1)} \stackrel{\text{def.}}{=} I^{(1)} \cup \{i + \varepsilon_i ; i \in I \text{ and } \varepsilon_i = +1\}$ . Since  $|x_{0,\nu} - x_{0,\nu'}| > 2h_n$  for  $\nu' \neq \nu$  and  $n$  large enough, we have  $\text{Card } J^{(\mathbf{r})} + \text{Card } J^{(1)} = 3 \times \text{Card } I = 3N$ . We shall find a choice of  $\varepsilon$  such that  $u_j > 0$  for all  $j \in J^{(\mathbf{r})} \setminus I$ , and  $v_j > 0$  for all  $j \in J^{(1)} \setminus I$ , where

$$\begin{pmatrix} u_{J^{(\mathbf{r})}} \\ v_{J^{(1)}} \end{pmatrix} \stackrel{\text{def.}}{=} -(\hat{\mathcal{A}}_h^* \hat{\mathcal{A}}_h)^{-1} \begin{pmatrix} \mathbb{1}_{J^{(\mathbf{r})}} \\ \mathbb{1}_{J^{(1)}} \end{pmatrix}, \quad \hat{\mathcal{A}}_h \stackrel{\text{def.}}{=} \begin{pmatrix} (A + \frac{h}{2}B)_{J^{(\mathbf{r})}} & (A - \frac{h}{2}B)_{J^{(1)}} \end{pmatrix}$$

$$A \stackrel{\text{def.}}{=} \Phi_{\mathcal{G}_n} \quad \text{and} \quad B \stackrel{\text{def.}}{=} \Phi'_{\mathcal{G}_n}.$$

In this particular case where  $I^{(\mathbf{r})} = I^{(1)} = I$ , all  $j$  in  $(J^{(\mathbf{r})} \setminus I) \cup (J^{(1)} \setminus I)$  may be uniquely written as  $j = i + \varepsilon_i$  for some  $i \in I$ , where  $\varepsilon_i \in \{-1, +1\}$ . We may swap the columns of  $\hat{\mathcal{A}}_h$  so as to reformulate the condition  $\begin{pmatrix} u_{J^{(\mathbf{r})}} \\ v_{J^{(1)}} \end{pmatrix} = -(\hat{\mathcal{A}}_h^* \hat{\mathcal{A}}_h)^{-1} \mathbb{1}_{3N}$  into

$$\begin{pmatrix} \tilde{u}_I \\ \tilde{v}_I \\ \tilde{t}_I \end{pmatrix} = -(\bar{\mathcal{A}}_h^* \bar{\mathcal{A}}_h)^{-1} \begin{pmatrix} \mathbb{1}_N \\ \mathbb{1}_N \\ \mathbb{1}_N \end{pmatrix},$$

where  $\bar{\mathcal{A}}_h \stackrel{\text{def.}}{=} \begin{pmatrix} A_I + \frac{h_n}{2} B_I \text{diag}(\varepsilon) & A_I - \frac{h_n}{2} B_I \text{diag}(\varepsilon) & A_{I+\varepsilon} - \frac{h_n}{2} B_{I+\varepsilon} \text{diag}(\varepsilon) \end{pmatrix}$  and  $\tilde{t}_i > 0$  for all  $i \in I$ . But a Taylor expansion yields

$$A_{I+\varepsilon} - \frac{h_n}{2} B_{I+\varepsilon} \text{diag}(\varepsilon) = \underbrace{\Phi_{x_0}}_{=A_I} + \frac{h_n}{2} \underbrace{\Phi'_{x_0} \text{diag}(\varepsilon)}_{=B_I \text{diag}(\varepsilon)} + (h_n)^3 \gamma_3 \Phi_{x_0}^{(3)} \text{diag}(\varepsilon) + o(h_n^3),$$

where we defined

$$\gamma_k \stackrel{\text{def.}}{=} \frac{1}{k!} - \frac{1}{(k-1)! \times 2}. \quad (96)$$

Hence, we may apply Lemma 9 to  $\Phi_{x_0}$ ,  $\Phi'_{x_0} \text{diag}(\varepsilon)$  and  $\gamma_3 \Phi_{x_0}^{(3)} \text{diag}(\varepsilon)$  so as to obtain

$$\tilde{t}_I = -\frac{1}{\gamma_3 h_n^3} \text{diag}(\varepsilon) \rho + o\left(\frac{1}{h_n^3}\right).$$

Therefore it is sufficient to choose  $\varepsilon = -\text{sign}(\rho)$  to make all the components of  $\tilde{t}_I$  nonnegative.

With that choice of  $\varepsilon$ , it remains to prove that

$$\max \left[ \begin{pmatrix} (A + \frac{h}{2}B)_{J^{(\mathbf{r})}^c}^* \\ (A - \frac{h}{2}B)_{J^{(1)}^c}^* \end{pmatrix} \hat{\mathcal{A}}_h \begin{pmatrix} u_{J^{(\mathbf{r})}} \\ v_{J^{(1)}} \end{pmatrix} \right] < 1.$$

Let us write  $\tilde{q}_n \stackrel{\text{def.}}{=} \hat{\mathcal{A}}_h \begin{pmatrix} u_{J^{(r)}} \\ v_{J^{(l)}} \end{pmatrix}$ . Since  $\begin{pmatrix} u_{J^{(r)}} \\ v_{J^{(l)}} \end{pmatrix} = -(\hat{\mathcal{A}}_h^* \hat{\mathcal{A}}_h)^{-1} \mathbf{1}_{3N}$ , we get  $\tilde{q}_n = \hat{\mathcal{A}}_h^{+,*} \mathbf{1}_{3N}$ , and applying Lemma 9 to  $\Phi_{x_0}$ ,  $\Phi'_{x_0} \text{diag}(\varepsilon)$  and  $\gamma_3 \Phi_{x_0}^{(3)} \text{diag}(\varepsilon)$ , we see that  $\tilde{q}_n$  converges towards  $q_T$  (using (88)).

By construction of  $\tilde{q}_n$ ,

$$\begin{aligned} \forall j \in J^{(r)} \setminus I, \quad (\Phi^* \tilde{q}_n + \frac{h_n}{2} (\Phi^* \tilde{q}_n)'(jh_n)) &= 1, \\ \text{and } \forall j \in J^{(l)} \setminus I, \quad (\Phi^* \tilde{q}_n - \frac{h_n}{2} (\Phi^* \tilde{q}_n)'(jh_n)) &= 1, \end{aligned} \quad (97)$$

which may be summarized as

$$\forall i \in I, (\Phi^* \tilde{q}_n - \varepsilon_i \frac{h_n}{2} (\Phi^* \tilde{q}_n)'((i + \varepsilon_i)h_n)) = 1.$$

Arguing as in the proof of point (iv) in Lemma 3 (replacing “ $1 = \dots$ ” with “ $1 \geq \dots$ ” and using that  $\mu_T^{(4)}(x_{0,\nu}) > 0$ ), we may prove that for  $n$  large enough,  $(\Phi^* \tilde{q}_n + \varepsilon_i \frac{h_n}{2} (\Phi^* \tilde{q}_n)'((i - \varepsilon_i)h_n)) < 1$ .

Then, by the same argument of compactness and local concavity as in point (ii) of Lemma 3, we observe that

$$\begin{aligned} \left\{ k \in \llbracket 0, P_n - 1 \rrbracket ; (\Phi^* \tilde{q}_n + \frac{h_n}{2} (\Phi^* \tilde{q}_n)'(kh_n)) \geq 1 \right\} &\subset J^{(r)}, \\ \left\{ k \in \llbracket 0, P_n - 1 \rrbracket ; (\Phi^* \tilde{q}_n - \frac{h_n}{2} (\Phi^* \tilde{q}_n)'(kh_n)) \geq 1 \right\} &\subset J^{(l)}, \end{aligned}$$

and those inclusions are in fact equalities. That precisely means that

$$\max \left[ \left( \begin{pmatrix} (A + \frac{h}{2}B)_{J^{(r)c}}^* \\ (A - \frac{h}{2}B)_{J^{(l)c}}^* \end{pmatrix} \tilde{q}_n \right) \right] < 1.$$

Hence, by Lemma 2,  $\Phi^* \tilde{q}_n$  is the minimal norm certificate  $\mu_0^n$  and  $(J^{(r)}h_n, J^{(l)}h_n)$  is the extended support. This concludes the proof.  $\square$

### 5.6. Asymptotics of the constants

Again, we may examine the asymptotic behavior of the constants given in Corollary 2. Those constants stem from Theorem 3 which is itself a variant of Theorem 1 for the LASSO.

Replacing the constants  $c_1, \dots, c_3$  of the proof of Theorem 1 with the corresponding

expressions for the C-BP, and using Lemma 9 we get

$$c_{1,n} = \|R_{I^{(r)} \cup I^{(l)}} \hat{\mathcal{A}}_h^+\| \sim \frac{1}{(h_n)^3} \left\| \begin{pmatrix} (\Phi_{x_0}^{(3),*} \tilde{\Pi} \Phi_{x_0}^{(3)})^{-1} \Phi_{x_0}^{(3),*} \tilde{\Pi} \\ 0 \end{pmatrix} \right\|_{\infty,2} \quad (98)$$

$$c_{2,n} = \left\| \begin{pmatrix} \tilde{u}_I \\ \tilde{v}_I \end{pmatrix} \right\| \sim \frac{1}{(h_n)^3} \left\| \begin{pmatrix} \rho \\ 0 \end{pmatrix} \right\|_{\infty} \quad (99)$$

$$c_{3,n} = \left( \|R_{(J^{(r)} \setminus I^{(r)}) \cup (J^{(l)} \setminus I^{(l)})} \hat{\mathcal{A}}_h^+\|_{\infty,2} \right)^{-1} \left( \min_{i \in I} \tilde{t}_i \right) \sim \frac{\min_i \left| \frac{1}{\gamma_3} \rho_i \right|}{\|(\Phi_{x_0}^{(3),*} \tilde{\Pi} \Phi_{x_0}^{(3)})^{-1} \Phi_{x_0}^{(3),*} \tilde{\Pi}\|_{\infty,2}} \quad (100)$$

where  $\gamma_k$  is defined in (96). As for  $c_{4,n}$  and  $c_{5,n}$ , like in the case of the LASSO, their expression lead to a pessimistic bound for the low noise regime, and we are led to make finer majorizations.

**Proposition 16.** *The constants  $C_n^{(1)}, C_n^{(2)}$  in Corollary 2 can be chosen as  $C_n^{(1)} = O(h_n^3)$  and  $C_n^{(2)} = O(1)$ , and one has*

$$\left\| \begin{pmatrix} \alpha_\lambda \\ \beta_\lambda \end{pmatrix} - \begin{pmatrix} \alpha_0 \\ 0 \end{pmatrix} \right\|_{\infty} = O\left(\frac{w}{h_n^3}, \frac{\lambda}{h_n^3}\right). \quad (101)$$

*Proof.* The proof of (101) follows from (98) and (99). Using the reformulation  $(\tilde{\mathcal{Q}}_\lambda(y))$  of the C-BP as a (positive) LASSO, we have to ensure that (23) holds, or more precisely,

$$\max \left[ \begin{pmatrix} (A^* + \frac{h_n}{2} B^*)_{(J^{(r)})^c} \\ (A^* - \frac{h_n}{2} B^*)_{(J^{(l)})^c} \end{pmatrix} (y - Aa - Bb) \right] < \lambda$$

where  $A \stackrel{\text{def.}}{=} \Phi_{\mathcal{G}_n}$ ,  $B \stackrel{\text{def.}}{=} \Phi'_{\mathcal{G}_n}$ . Let  $\hat{\mathcal{A}}_h \stackrel{\text{def.}}{=} \begin{pmatrix} (A + \frac{h}{2} B)_{J^{(r)}} & (A - \frac{h}{2} B)_{J^{(l)}} \end{pmatrix}$ ,  $\tilde{\Pi}$  be the orthogonal projector onto  $\ker \hat{\mathcal{A}}_h^* = (\text{Im } \hat{\mathcal{A}}_h)^\perp$ , and  $\omega = \Phi^* \tilde{\Pi} w$ . Since

$$y - Aa - Bb = w - \hat{\mathcal{A}}_h (\hat{\mathcal{A}}_h^* \hat{\mathcal{A}}_h)^{-1} \hat{\mathcal{A}}_h^* w + \lambda \hat{\mathcal{A}}_h (\hat{\mathcal{A}}_h^* \hat{\mathcal{A}}_h)^{-1} \mathbf{1}_{3N} = \tilde{\Pi} w + \lambda \hat{\mathcal{A}}_h^{+,*} \mathbf{1}_{3N},$$

we are led to check that

$$(\omega + \lambda \mu_0^n)(jh_n) + \frac{h_n}{2} (\omega + \lambda \mu_0^n)'(jh_n) < \lambda \quad \text{for all } j \in (J^{(r)})^c, \quad (102)$$

$$(\omega + \lambda \mu_0^n)(jh_n) - \frac{h_n}{2} (\omega - \lambda \mu_0^n)'(jh_n) < \lambda \quad \text{for all } j \in (J^{(l)})^c, \quad (103)$$

where  $\mu_0^n \stackrel{\text{def.}}{=} \Phi^* (\hat{\mathcal{A}}_h^* \hat{\mathcal{A}}_h)^{-1} \mathbf{1}_{3N}$  yields the minimal norm certificate

$$\begin{pmatrix} (\mu_0^n + \frac{h_n}{2} \mu_0^n)(\mathcal{G}_n) \\ (\mu_0^n - \frac{h_n}{2} \mu_0^n)(\mathcal{G}_n) \end{pmatrix} = \begin{pmatrix} (A + \frac{h}{2} B)^* \\ (A - \frac{h}{2} B)^* \end{pmatrix} (\hat{\mathcal{A}}_h^* \hat{\mathcal{A}}_h)^{-1} \mathbf{1}_{3N}.$$

Given  $0 < r < \frac{1}{2} \min_{\nu \neq \nu'} |x_{0,\nu} - x_{0,\nu'}|$ , let  $N(r) \stackrel{\text{def.}}{=} \bigcup_{\nu} (x_{0,\nu} - r, x_{0,\nu} + r)$  be a neighborhood of the  $x_{0,\nu}$ 's. By the Twice Non-Degenerate Source condition, we may choose  $r > 0$ , such that

$$-\tilde{k}_1 \stackrel{\text{def.}}{=} \sup_{t \in N(r)} \mu_T''(t) < 0, \quad \text{and} \quad \tilde{k}_2 \stackrel{\text{def.}}{=} \inf_{t \in N(r)} \mu_T^{(4)}(t) > 0.$$

By compactness,  $\tilde{k}_3 \stackrel{\text{def.}}{=} \sup_{t \in \mathbb{T} \setminus N(r)} \mu_T(t) < 1$ .

Let us recall that  $\mu_0^n \rightarrow \mu_T$  in the sense of the uniform convergence (and similarly for the derivatives). As a result, for  $n \in \mathbb{N}$  large enough,

$$\sup_{t \in N(r)} (\mu_0^n)''(t) < -\frac{\tilde{k}_1}{2} < 0, \quad \inf_{t \in N(r)} (\mu_0^n)^{(4)}(t) > \frac{\tilde{k}_2}{2} > 0, \quad \sup_{t \in \mathbb{T} \setminus N(r)} \mu_0^n(t) < \frac{1 + \tilde{k}_3}{2} < 1, \quad (104)$$

$$\frac{h_n}{2} \|(\mu_0^n)^{(3)}\|_\infty \leq \frac{\tilde{k}_1}{8}, \quad \text{and} \quad \frac{h_n}{2} \|(\mu_0^n)'\|_\infty \leq \frac{1 - \tilde{k}_3}{6}.$$

Now, we assume that  $\frac{\|w\|_2}{\lambda}$  is small enough, so that

$$\|(\Phi^{(k)})^*\|_{\infty,2} \frac{\|w\|_2}{\lambda} < \frac{\tilde{k}_1}{8}, \quad \text{for } k \in \{2, 3\}, \quad \|(\Phi^{(4)})^*\|_{\infty,2} \frac{\|w\|_2}{\lambda} < \frac{\tilde{k}_2}{4},$$

$$\text{and} \quad \|(\Phi^{(k)})^*\|_{\infty,2} \frac{\|w\|_2}{\lambda} < \frac{1 - \tilde{k}_3}{6}, \quad \text{for } k \in \{0, 1\}, \quad (105)$$

Then, using the fact that  $|\omega^{(k)}(t)| \leq \|(\Phi^{(k)})^*\|_{\infty,2} \|w\|_2$  and  $h_n \leq 1$ , we obtain

$$\sup_{t \in \mathbb{T} \setminus N(r)} \left( \frac{\omega}{\lambda} + \mu_0^n + \frac{h_n}{2} \left| \left( \frac{\omega}{\lambda} + \mu_0^n \right)' \right| \right) (t) < 1.$$

Thus it remains to prove that for each  $\nu \in \{1, \dots, N\}$ ,

$$\left( \frac{\omega}{\lambda} + \mu_0^n + \frac{h_n}{2} \left( \frac{\omega}{\lambda} + \mu_0^n \right)' \right) (t) < 1 \quad \text{for } t \in (x_{0,\nu} - r, x_{0,\nu} + r) \setminus S_{n,\nu}^{(r)}(r), \quad (106)$$

$$\text{and} \quad \left( \frac{\omega}{\lambda} + \mu_0^n - \frac{h_n}{2} \left( \frac{\omega}{\lambda} + \mu_0^n \right)' \right) (t) < 1 \quad \text{for } t \in (x_{0,\nu} - r, x_{0,\nu} + r) \setminus S_{n,\nu}^{(l)}(r). \quad (107)$$

We only deal with the case  $S_{n,\nu}^{(r)}(r) = \{x_{0,\nu}\}$ ,  $S_{n,\nu}^{(l)}(r) = \{x_{0,\nu}, x_{0,\nu} + h_n\}$ , the symmetric case being similar. Let  $f \stackrel{\text{def.}}{=} \frac{1}{\lambda} \omega(\cdot - x_{0,\nu}) + \mu_0^n(\cdot - x_{0,\nu})$ . By definition of  $\tilde{\Pi}$ ,  $\omega(x_{0,\nu}) = \omega'(x_{0,\nu}) = \omega(x_{0,\nu} + h_n) - \frac{h_n}{2} \omega'(x_{0,\nu} + h_n) = 0$ , so that

$$f(0) = 1, \quad f'(0) = 1, \quad \text{and} \quad f(h_n) - \frac{h_n}{2} f'(h_n) = 1. \quad (108)$$

Moreover, from Eq. (104) to (105), and letting  $k_1 = \frac{\tilde{k}_1}{8}$ ,  $k_2 = \frac{\tilde{k}_2}{4}$ , we deduce that

$$\forall t \in (-r, r), \quad f''(t) + \frac{h}{2} |f^{(3)}(t)| < -k_1 < 0, \quad \text{and} \quad f^{(4)}(t) > k_2 > 0, \quad (109)$$

so that the strict concavity of  $f - \frac{h_n}{2} f'$  implies that  $(f - \frac{h_n}{2} f')(t) < 1$  for  $t \in (-r, -h_n) \cup (0, r)$ .

It remains to prove that  $(f + \frac{h_n}{2} f')(t) < 1$  for  $t \in (-r, r) \setminus (-h_n, 0]$ . A Taylor expansion of  $f$  and  $f'$  yields (writing as usual  $\gamma_k = \frac{1}{k!} - \frac{1}{(k-1)! \times 2}$ )

$$\underbrace{1 - (f(h) - \frac{h_n}{2} f(h_n))}_{=0} = \underbrace{1 - f(0) - \frac{h_n}{2} f'(0)}_{=0} - h_n^3 \gamma_3 f^{(3)}(0) - h_n^4 \gamma_4 f^{(4)}(0) + R_1(h_n),$$

$$1 - (f(-h_n) + \frac{h_n}{2} f'(-h_n)) = \underbrace{1 - f(0) + \frac{h_n}{2} f'(0)}_{=0} + h_n^3 \gamma_3 f^{(3)}(0) - h_n^4 \gamma_4 f^{(4)}(0) + R_2(h_n).$$

Adding both equations we get

$$1 - (f(-h_n) + \frac{h_n}{2}f'(-h_n)) = -2h_n^4\gamma_4f^{(4)}(0) + (R_1 + R_2)(h_n) \quad (110)$$

$$\geq 2(-\gamma_4)k_2h_n^4 + (R_1 + R_2)(h_n), \quad (111)$$

where

$$R_1 + R_2(h) = h_n^5 \int_0^1 (-f^{(5)}(sh_n) + f^{(5)}(-sh_n)) \left( \frac{(1-s)^4}{4!} - \frac{(1-s)^3}{2 \times 3!} \right) ds, \quad (112)$$

$$\text{with } \|f^{(5)}\|_\infty \leq \left\| \frac{1}{\lambda} \omega^{(5)} \right\|_\infty + \|(\mu_0^n)^{(5)}\|_\infty = O(1). \quad (113)$$

Hence,

$$1 - (f(-h_n) + \frac{h_n}{2}f'(-h_n)) \geq \underbrace{2(-\gamma_4)k_2}_{>0} h_n^4 + O(h_n^5) > 0. \quad (114)$$

Moreover, by the strict concavity of  $f + \frac{h_n}{2}f'$ , we also deduce that  $(f + \frac{h_n}{2}f')(t) < 1$  for  $t \in (-r, -h_n] \cup (0, r)$ , thus we get the local inequalities (106) and (107), hence the global inequalities (102) and (103).

To conclude, the constants in the condition on  $\frac{\|w\|_2}{\lambda}$  are  $O(1)$ , and gathering the asymptotics for  $c_{1,n}, c_{2,n}, c_{3,n}$  we obtain  $C_n^{(1)} = O(h_n^3)$ ,  $C_n^{(2)} = O(1)$ .  $\square$

## 6. Numerical illustrations

In this section, we illustrate the usefulness of our analysis to gain a precise understanding of the recovery performance of  $\ell^1$ -type methods (LASSO and C-BP) for both deconvolution and compressed sensing problems. The code to reproduce these numerical experiments is available online $\ddagger$ .

### 6.1. Convergence of pre-certificates

In this section and in Section 6.2, we consider the deconvolution problems in the case where  $\varphi$  is an ideal filter, i.e. whose Fourier coefficients

$$\forall k \in \mathbb{Z}, \quad \hat{\varphi}(k) \stackrel{\text{def.}}{=} \int_{\mathbb{T}} \varphi(t) e^{-2i\pi kt} dt$$

satisfy  $\hat{\varphi}(k) = 1$  if  $k \in \{-f_c, \dots, f_c\}$  and  $\hat{\varphi}(k) = 0$  otherwise. This allows us to implement exactly the  $\Phi$  operator appearing in the LASSO and C-BP problem since  $\text{Im}(\Phi)$  is a finite dimensional space of dimension  $Q = 2f_c + 1$ , i.e. it can be represented using a matrix of size  $(Q, P)$  when evaluated on a grid of  $P$  points. In Figures 1 and 2 we used  $f_c = 10$ .

$\ddagger$  <https://github.com/gpeyre/2015-IP-lasso-cbp/>

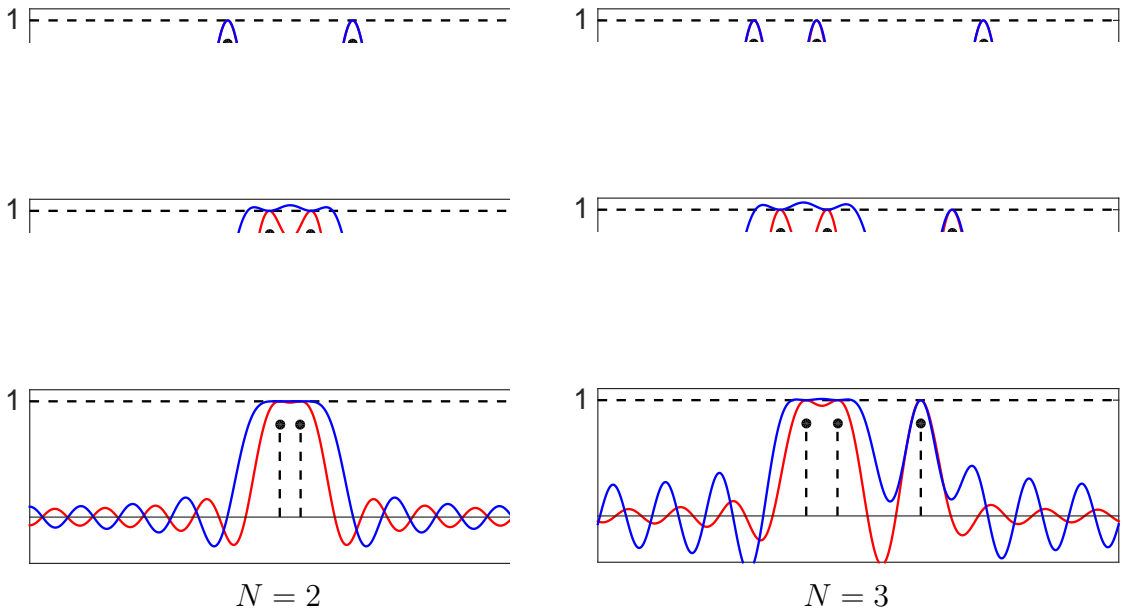


Figure 1: Display of  $\eta_V^\infty$  (red) and  $\mu_T$  (blue) pre-certificate for different input positive measures  $m_0$  (showed as black dots to symbolize the position of the Diracs).

Figure 1 illustrates for the case of two ( $N = 2$ ) and three ( $N = 3$ ) spikes the behavior of the vanishing pre-certificate  $\eta_V^\infty$  (see Definition 3) useful to analyze LASSO/BLASSO problems and of the pre-certificate  $\mu_T$  (see Definition 7) useful to analyze C-BP problems.

We first notice that for all the (positive) input measures (i.e. whatever the spacing between the Diracs),  $\eta_V^\infty$  is always a non-degenerate certificate (in the sense of Proposition 3), meaning that one actually has  $\eta_V^\infty = \eta_0^\infty$  (where the minimal norm certificate  $\eta_0^\infty$  is defined in (37)). This empirical finding is the subject of another recent work on the asymptotic of sparse recovery of positive measures when the spacing between the Diracs tends to zero [16]. Since  $\eta_0^\infty$  is non-degenerate, one can thus apply Theorem 2 to analyze the extended support of the LASSO (see below Section 6.2 for a numerical illustration).

For the C-BP problem, the situation is however more contrasted. We observe that when the Dirac masses are separated enough (first row) then the pre-certificate  $\mu_T$  is a valid certificate, meaning the the Twice Non-Degenerate Source Condition (see Definition 8) holds. This means that Theorem 4 can be applied to analyze the extended support of C-BP (see Section 6.2 below for a numerical illustration). But when the Dirac masses are too close (second and third rows), one has  $\|\mu_T\|_\infty > 1$ , so that one cannot ensure the support stability of the C-BP solution with our result.

## 6.2. Extended support for deconvolution

We still consider the case of an ideal low pass filter. Figure 2 displays the evolution, as a function of  $\lambda$  (in abscissa) of the solutions  $a_\lambda$  of  $(\mathcal{P}_\lambda^n(y))$  and of  $(a_\lambda, b_\lambda)$  of  $(\mathcal{Q}_\lambda^n(y))$ . We consider here the case of an input measure with two nearby Diracs (displayed as

red/blue dots in the upper-left part of the Figure) and when there is no noise (i.e.  $w = 0$ ). Each 1-D curve (either plain or dashed) corresponds to a different variable, e.g.  $(a_\lambda)_i$ , for some index  $i$  (only non-zero).

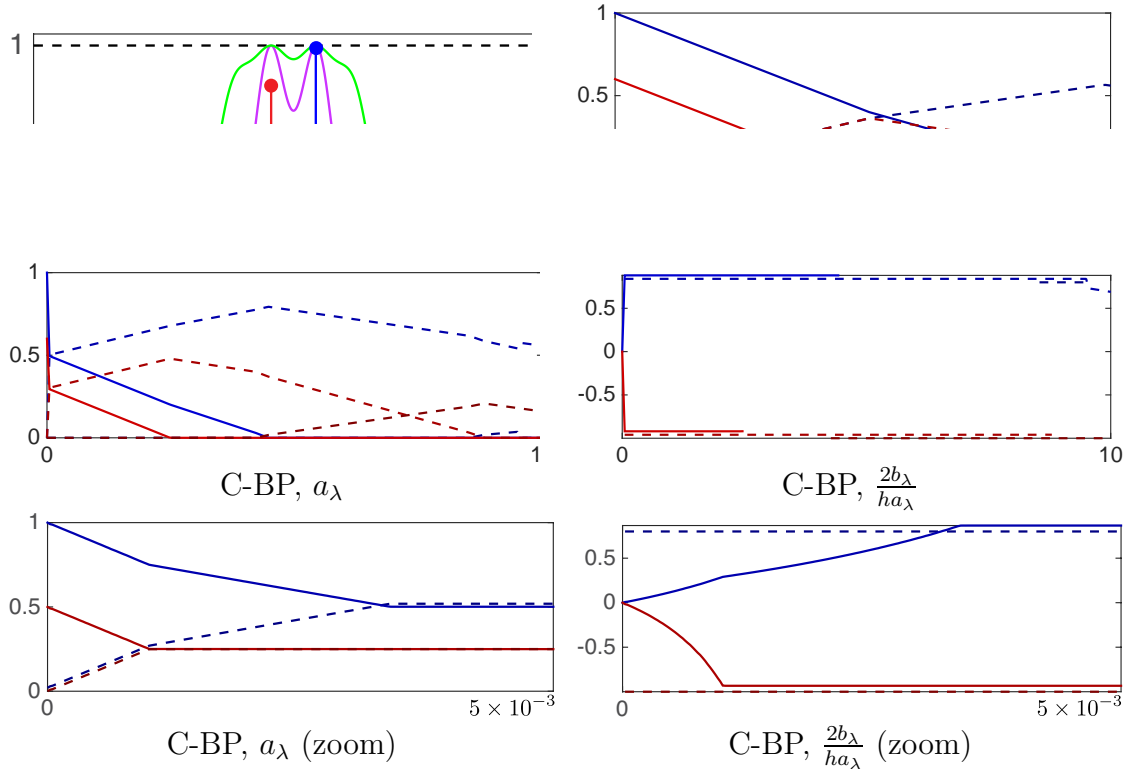


Figure 2: Display of the evolution as a function of  $\lambda$  of the solutions of the LASSO and C-BP problems. Note that dashed curves have been (artificially) slightly shifted to avoid that they overlap with the plain curve.

The solutions path  $\lambda \mapsto a_\lambda$  (for LASSO) and  $\lambda \mapsto (a_\lambda, b_\lambda)$  (for C-BP) are continuous and piecewise affine, which is to be expected since the regularizations ( $\ell^1$  and  $\ell^1$  under conic constraints) are polyhedral. The upper-left plot in the figure displays the precertificate  $\eta_V^\infty$  (in magenta, see Definition 3) and  $\mu_T$  (in green, see Definition 7). This shows graphically that these two precertificates are non-degenerate (according to Definitions 2 and 8) so that the results of Theorems 2 and 4 hold, hence precisely describing the evolution of the solution on the extended support when  $\lambda$  is small. On these graphs, this corresponds to the first segment of the corresponding piecewise affine paths.

The behavior for BP agrees with our analysis. As predicted by Theorem 2, there exists a range of values  $0 < \lambda < \lambda_0$  on which the solution is exactly supported on the extended support  $J$ , which is composed of four spikes (the plain curve corresponds to the support  $I$  and the dashed curve corresponds to  $J \setminus I$ ). Also, as predicted by Proposition 7 in the case  $w = 0$ , we verify that  $\lambda_0 = O(h_n)$  and that the Lipschitz constant of  $\lambda \mapsto a_\lambda$  is of order  $O(1/h_n)$ .

In sharp contrast, the behavior for C-BP is less regular, since the range  $0 < \lambda < \lambda_0$  on which the solution is supported on the extended support is shorter, as it can be clearly seen on the zoom for very small values of  $\lambda$ . This is in agreement with Proposition 16 which shows that  $\lambda_0$  is of the order of  $O(h_n^3)$  and that the Lipschitz constant of  $\lambda \mapsto (a_\lambda, b_\lambda)$  is of order  $O(1/h_n^3)$ . On this range of small  $\lambda$ , as predicted by Theorem 4, the support of the solutions (which correspond to the extended support  $J$  described in Theorem 4) is composed of one pair of neighboring spikes for each original spike. For indices on the support  $i \in I$ , one has  $|(b_\lambda)_i|/(a_\lambda)_i < h/2$  (the constraint is non-saturating, and the spike moves “freely” inside  $(ih - \frac{h}{2}, ih + \frac{h}{2})$ ) while for indices on the extended part  $i \in J \setminus I$ , one has  $|(b_\lambda)_i|/(a_\lambda)_i = h/2$  (the constraint is saturating, the spikes are fixed at half-grid points). Another part of the path is interesting, for  $\lambda$  not so small (say  $\lambda > \lambda_1$ ), which is in fact the prominent regime in the non-zoomed figure. For this range of  $\lambda$ , there is still a pair of spikes for each original spike, but this time both spikes saturate, on same side. This observation should be related to Proposition 13 and Remark 10 which predict that, in the case where  $\mu_\lambda^{\infty(3)}(x_{\lambda,\nu}) \neq 0$ , the C-BP yields either one spike or a pair of spikes with the same shift (the latter case is in fact overwhelming).

### 6.3. Extended support for compressed sensing

To show the usefulness of our “abstract” support analysis of the LASSO problem (Section 4), we illustrate its use to analyze the performance of  $\ell^1$  recovery in a compressed sensing setup. 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 [8] 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  $|\text{supp}(a_0)| = s$ ), that there are roughly three regimes:

- If  $s < s_0 \stackrel{\text{def.}}{=} \frac{Q}{2 \log(P)}$ , 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  $\eta_F$  (as defined in (27)) is a valid certificate,  $\|\eta_F\|_\infty \leq 1$ . This is shown for instance in [36, 20].
- If  $s < s_1 \stackrel{\text{def.}}{=} \frac{Q}{2 \log(P/Q)}$ , then  $a_0$  is with “high probability” the unique solution of  $(\mathcal{P}_0(y_0))$ , but the support is not stable, meaning that  $\eta_F$  is not a valid certificate.

This phenomena is precisely analyzed in [11, 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 3 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 curve shows, as a function of  $s$  (in abscissa), the probability (with respect to a random draw of  $\Phi$  and  $a_0$  a  $s$ -sparse vector) of the event that  $a_0$  is identifiable (plain curve) and of the event that  $\eta_F$  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  $\lambda$  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  $\ell^1$ -type variational problems in the low-noise regime. This includes in particular the LASSO and the C-BP problems. 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 the deconvolution problem, 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 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.

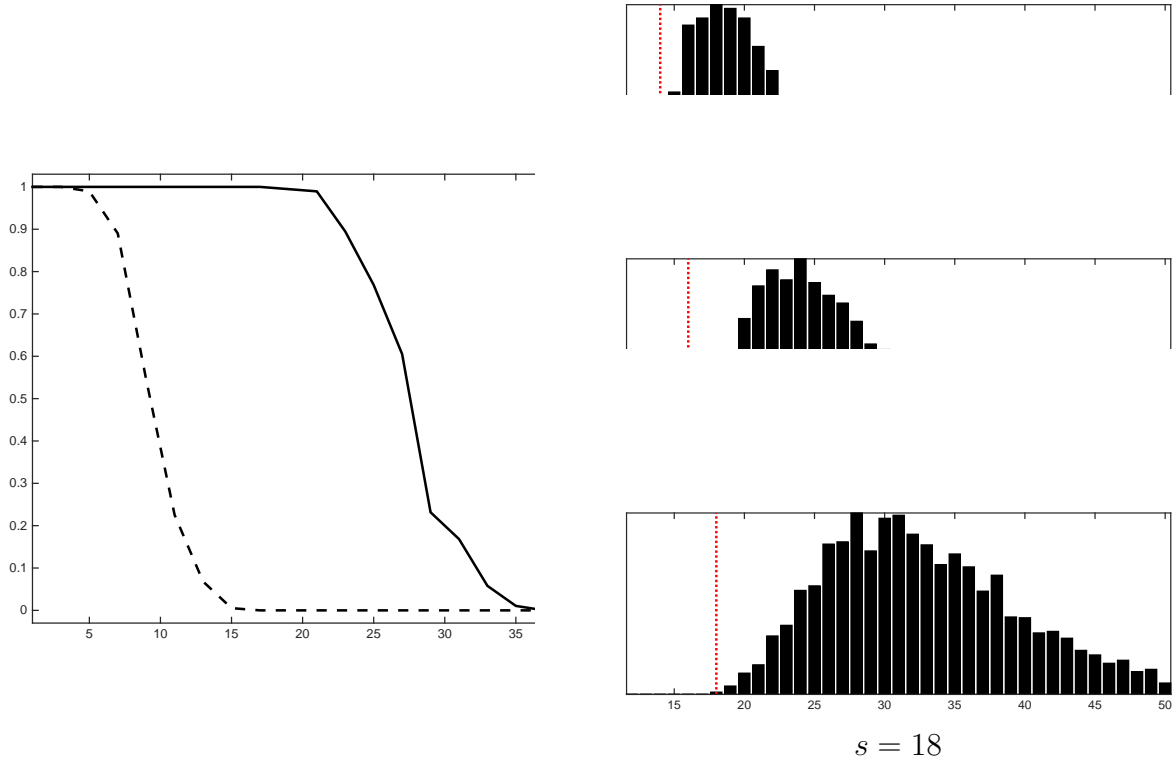


Figure 3: *Left*: probability as a function of  $s$  of the event that  $a_0$  is identifiable (plain curve) and of the event that its support is stable (dashed curve). *Right*: for several value of  $s$ , display of histogram of repartition of the sizes  $|J|$  of the extended support  $J$ .

### 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 6.** *Let  $K \in \mathbb{N}^*$  and assume that  $\varphi \in C^K(\mathbb{T} \times \mathbb{T})$ . Then for all  $p \in L^2(\mathbb{T})$ ,  $\Phi^*p \in C^K(\mathbb{T})$  and for all  $k \in \{0, 1, \dots, K\}$*

$$\forall y \in \mathbb{T}, (\Phi^*p)^{(k)}(y) = \int_{\mathbb{T}} (\partial_2)^k \varphi(x, y) p(x) dx. \quad (\text{A.1})$$

Moreover, the adjoint operator

$$\Phi^{(k),*} : \begin{array}{l} L^2(\mathbb{T}) \longrightarrow C(\mathbb{T}) \\ p \longmapsto \int_{\mathbb{T}} (\partial_2)^k \varphi(x, y) p(x) dx \end{array}, \quad (\text{A.2})$$

is compact for all  $k \in \{0, 1, \dots, K\}$ .

*Proof.* The first part of the lemma is a standard application of the Lebesgue dominated convergence theorem.

As for the second part, it is a consequence of the Ascoli-Arzelà theorem. Indeed, let  $B_{L^2}(0, 1) = \{p \in L^2(\mathbb{T}); \|p\|_2 \leq 1\}$ , and  $\mathcal{F} = \{\Phi^*p; p \in B_{L^2}(0, 1)\}$ . Then  $\mathcal{F} \subset C(\mathbb{T})$  is bounded since

$$|\Phi^{(k),*}p(x)| \leq \sqrt{\int_{\mathbb{T}} ((\partial_2)^k \varphi(x, y))^2 dx} \sqrt{\int_{\mathbb{T}} (p(x))^2 dx} \leq \sqrt{\|(\partial_2)^k \varphi\|_\infty}$$

Moreover, it is equicontinuous since

$$|\Phi^{(k),*}p(x) - \Phi^{(k),*}p(x')| = \left| \int_{\mathbb{T}} ((\partial_2)^k \varphi(x, y) - (\partial_2)^k \varphi(x', y)) p(x) \right| \quad (\text{A.3})$$

$$\leq \sqrt{\omega_{(\partial_2)^k \varphi}(|x - x'|, 0)}, \quad (\text{A.4})$$

where  $\omega_{(\partial_2)^k \varphi}$  is the modulus of continuity of  $(\partial_2)^k \varphi$ . Thus Ascoli-Arzelà's theorem ensures that  $\Phi^{(k),*}B_{L^2}(0, 1)$  is relatively compact, hence the result.  $\square$

An interesting consequence of the above lemma is the following. Given any bounded sequence  $\{p_n\}_{n \in \mathbb{N}}$  in  $L^2(\mathbb{T})$ , we may extract a subsequence  $\{p_{n'}\}_{n' \in \mathbb{N}}$  which converges weakly towards some  $\tilde{p} \in L^2(\mathbb{T})$ . 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 7.** *Let  $A: \mathbb{R}^N \rightarrow L^2(\mathbb{T})$ ,  $B: \mathbb{R}^N \rightarrow \mathbb{R}^N$  be linear operators such that  $A$  has full rank and  $B$  is invertible. Then  $(AB)^+ = B^{-1}A^+$ .*

*Proof.* It is sufficient to write

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

$\square$

**Lemma 8.** *Let  $A, B, B_h: \mathbb{R}^N \rightarrow L^2(\mathbb{T})$  be linear operators such that  $B_h = B + O(h)$  for  $h > 0$ , and that  $\begin{pmatrix} A & B \end{pmatrix}$  has full rank. Let  $\Pi$  be the orthogonal projector onto  $(\text{Im } A)^\perp$ , and let*

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

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

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

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

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

*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

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

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

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

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

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

Eventually, by Lemma 7,  $\begin{pmatrix} A & A + hB_h \end{pmatrix}^+ = \begin{pmatrix} I_N & -\frac{1}{h} I_N \\ 0 & \frac{1}{h} I_N \end{pmatrix} \begin{pmatrix} A^* A & A^* B_h \\ B_h^* A & B_h^* B_h \end{pmatrix}^{-1} \begin{pmatrix} A^* \\ B_h^* \end{pmatrix}$ .

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

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

□

**Lemma 9.** *Let  $A, B, C, C_h: \mathbb{R}^N \rightarrow L^2(\mathbb{T})$  be linear operators such that  $C_h = C + o(1)$  for  $h > 0$ , and that  $\begin{pmatrix} A & B & C \end{pmatrix}$  has full rank. Let  $\tilde{\Pi}$  be the orthogonal projector onto  $(\text{Im} \begin{pmatrix} A & B \end{pmatrix})^\perp$ , and let*

$$G_h \stackrel{\text{def.}}{=} \begin{pmatrix} (A + \frac{h}{2} B)^* \\ (A - \frac{h}{2} B)^* \\ (A + \frac{h}{2} B + h^3 C_h)^* \end{pmatrix} \begin{pmatrix} A + \frac{h}{2} B & A - \frac{h}{2} B & A + \frac{h}{2} B + h^3 C_h \end{pmatrix}.$$

Then for  $h > 0$  small enough,  $G_h$  and  $C^* \tilde{\Pi} C$  are invertible, and

$$G_h^{-1} \begin{pmatrix} \mathbf{1}_N \\ \mathbf{1}_N \\ \mathbf{1}_N \end{pmatrix} = -\frac{1}{h^3} \begin{pmatrix} -I_N \\ 0 \\ I_N \end{pmatrix} (C^* \tilde{\Pi} C)^{-1} C^* \begin{pmatrix} A & B \end{pmatrix}^{+,*} \begin{pmatrix} \mathbf{1}_N \\ 0 \end{pmatrix} + o\left(\frac{1}{h^3}\right)$$

$$\begin{pmatrix} A + \frac{h}{2} B & A - \frac{h}{2} B & A + \frac{h}{2} B + h^3 C_h \end{pmatrix}^+ = \frac{1}{h^3} \begin{pmatrix} -(C^* \tilde{\Pi} C)^{-1} C^* \tilde{\Pi} \\ 0 \\ (C^* \tilde{\Pi} C)^{-1} C^* \tilde{\Pi} \end{pmatrix} + o\left(\frac{1}{h^3}\right),$$

but

$$\begin{pmatrix} A^* + \frac{h}{2} B^* \\ A^* - \frac{h}{2} B^* \\ A^* + \frac{h}{2} B^* + h^3 C_h^* \end{pmatrix}^+ \begin{pmatrix} \mathbf{1}_N \\ \mathbf{1}_N \\ \mathbf{1}_N \end{pmatrix} = \begin{pmatrix} A^* \\ B^* \end{pmatrix}^+ \begin{pmatrix} \mathbf{1}_N \\ 0 \end{pmatrix} - \tilde{\Pi} C (C^* \tilde{\Pi} C)^{-1} C^* \begin{pmatrix} A^* \\ B^* \end{pmatrix}^+ \begin{pmatrix} \mathbf{1}_N \\ 0 \end{pmatrix} + o(1).$$

*Proof.* Observe that

$$\begin{pmatrix} A + \frac{h}{2} B & A - \frac{h}{2} B & A + \frac{h}{2} B + h^3 C_h \end{pmatrix} = \begin{pmatrix} A & B & C_h \end{pmatrix} \text{diag} \left( 1, \frac{h}{2}, h^3 \right) \begin{pmatrix} I_N & I_N & I_N \\ I_N & -I_N & I_N \\ 0 & 0 & I_N \end{pmatrix}$$

As a result, for  $h > 0$  small enough  $G_h$  is invertible and

$$G_h^{-1} = \begin{pmatrix} \frac{1}{2}I_N & \frac{1}{2}I_N & -I_N \\ \frac{1}{2}I_N & -\frac{1}{2}I_N & 0 \\ 0 & 0 & I_N \end{pmatrix} \text{diag} \left( 1, \frac{2}{h}, \frac{1}{h^3} \right) \begin{pmatrix} A^*A & A^*B & A^*C_h \\ B^*A & B^*B & B^*C_h \\ C_h^*A & C_h^*B & C_h^*C_h \end{pmatrix}^{-1} \\ \times \text{diag} \left( 1, \frac{2}{h}, \frac{1}{h^3} \right) \begin{pmatrix} \frac{1}{2}I_N & \frac{1}{2}I_N & 0 \\ \frac{1}{2}I_N & -\frac{1}{2}I_N & 0 \\ -I_N & 0 & I_N \end{pmatrix}$$

the middle matrix being invertible from the full rank assumption on  $(A \ B \ C)$ .

Moreover, writing  $\tilde{\Gamma} \stackrel{\text{def.}}{=} (A \ B)$  and  $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \stackrel{\text{def.}}{=} \begin{pmatrix} \tilde{\Gamma}^*\tilde{\Gamma} & \tilde{\Gamma}^*C_h \\ C_h^*\tilde{\Gamma} & C_h^*C_h \end{pmatrix}$ , we obtain

$$\begin{pmatrix} A^*A & A^*B & A^*C_h \\ B^*A & B^*B & B^*C_h \\ C_h^*A & C_h^*B & C_h^*C_h \end{pmatrix}^{-1} = \begin{pmatrix} u & -a^{-1}bS^{-1} \\ -S^{-1}ca^{-1} & S^{-1} \end{pmatrix}$$

where  $u \stackrel{\text{def.}}{=} a^{-1} + a^{-1}bS^{-1}ca^{-1}$ ,  $S \stackrel{\text{def.}}{=} d - ca^{-1}b = C_h^*\tilde{\Pi}C_h$ ,  $a^{-1}bS^{-1} = (\tilde{\Gamma}^*\tilde{\Gamma})^{-1}\tilde{\Gamma}^*C_h(C_h^*\tilde{\Pi}C_h)^{-1}$ ,  $S^{-1}ca^{-1} = (C_h^*\tilde{\Pi}C_h)^{-1}C_h^*\tilde{\Gamma}(\tilde{\Gamma}^*\tilde{\Gamma})^{-1}$ , and  $\tilde{\Pi}$  is the orthogonal projector onto  $(\text{Im } \tilde{\Gamma})^\perp$ . Thus

$$G_h^{-1} \begin{pmatrix} \mathbf{1}_N \\ \mathbf{1}_N \\ \mathbf{1}_N \end{pmatrix} = \begin{pmatrix} \frac{1}{2}I_N & \frac{1}{2}I_N & -I_N \\ \frac{1}{2}I_N & -\frac{1}{2}I_N & 0 \\ 0 & 0 & I_N \end{pmatrix} \text{diag} \left( 1, \frac{2}{h}, \frac{1}{h^3} \right) \begin{pmatrix} u & -a^{-1}bS^{-1} \\ -S^{-1}ca^{-1} & S^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{1}_N \\ 0 \\ 0 \end{pmatrix} \\ = \frac{1}{h^3} \begin{pmatrix} 0 & 0 & -I_N \\ 0 & 0 & 0 \\ 0 & 0 & I_N \end{pmatrix} \begin{pmatrix} u \begin{pmatrix} \mathbf{1}_N \\ 0 \end{pmatrix} \\ -S^{-1}ca^{-1} \begin{pmatrix} \mathbf{1}_N \\ 0 \end{pmatrix} \end{pmatrix} + o\left(\frac{1}{h^3}\right) \\ = -\frac{1}{h^3} \begin{pmatrix} -I_N \\ 0 \\ I_N \end{pmatrix} (C^*\tilde{\Pi}C)^{-1}C^*\tilde{\Gamma}^{+,*} \begin{pmatrix} \mathbf{1}_N \\ 0 \end{pmatrix} + o\left(\frac{1}{h^3}\right).$$

Eventually, one has

$$\begin{aligned}
& \left( A + \frac{\hbar}{2}B \quad A - \frac{\hbar}{2}B \quad A + \frac{\hbar}{2}B + \hbar^3 C_h \right)^+ \\
&= \begin{pmatrix} \frac{1}{2}I_N & \frac{1}{2}I_N & -I_N \\ \frac{1}{2}I_N & -\frac{1}{2}I_N & 0 \\ 0 & 0 & I_N \end{pmatrix} \text{diag}\left(1, \frac{2}{\hbar}, \frac{1}{\hbar^3}\right) \left( \tilde{\Gamma} \quad C_h \right)^+ \\
&= \frac{1}{\hbar^3} \begin{pmatrix} 0 & 0 & -I_N \\ 0 & 0 & 0 \\ 0 & 0 & I_N \end{pmatrix} \begin{pmatrix} u & -a^{-1}bS^{-1} \\ -S^{-1}ca^{-1} & S^{-1} \end{pmatrix} \left( \tilde{\Gamma}^* \quad C_h^* \right) + o\left(\frac{1}{\hbar^3}\right) \\
&= \frac{1}{\hbar^3} \begin{pmatrix} -(C^* \tilde{\Pi} C)^{-1} C^* \tilde{\Pi} \\ 0 \\ (C^* \tilde{\Pi} C)^{-1} C^* \tilde{\Pi} \end{pmatrix} + o\left(\frac{1}{\hbar^3}\right),
\end{aligned}$$

and

$$\begin{aligned}
& \left( A + \frac{\hbar}{2}B \quad A - \frac{\hbar}{2}B \quad A + \frac{\hbar}{2}B + \hbar^3 C_h \right)^{+,*} \begin{pmatrix} \mathbf{1}_N \\ \mathbf{1}_N \\ \mathbf{1}_N \end{pmatrix} \\
&= \begin{pmatrix} \tilde{\Gamma}^* \\ C_h^* \end{pmatrix}^+ \text{diag}\left(1, \frac{2}{\hbar}, \frac{1}{\hbar^3}\right) \begin{pmatrix} \frac{1}{2}I_N & \frac{1}{2}I_N & 0 \\ \frac{1}{2}I_N & -\frac{1}{2}I_N & -I_N \\ 0 & 0 & I_N \end{pmatrix} \begin{pmatrix} \mathbf{1}_N \\ \mathbf{1}_N \\ \mathbf{1}_N \end{pmatrix} \\
&= \left[ \tilde{\Gamma}^{+,*} - \tilde{\Pi} C_h (C_h^* \tilde{\Pi} C_h)^{-1} C_h^* \tilde{\Gamma}^{+,*} \right] \begin{pmatrix} \mathbf{1}_N \\ 0 \end{pmatrix} \\
&= \tilde{\Gamma}^{+,*} \begin{pmatrix} \mathbf{1}_N \\ 0 \end{pmatrix} - \tilde{\Pi} C (C^* \tilde{\Pi} C)^{-1} C^* \tilde{\Gamma}^{+,*} \begin{pmatrix} \mathbf{1}_N \\ 0 \end{pmatrix} + o(1)
\end{aligned}$$

□

## 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] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. New York: Springer-Verlag, 2011.
- [4] 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.

- [5] T. Blu et al. “Sparse Sampling of Signal Innovations”. In: *IEEE Signal Processing Magazine* 25.2 (2008), pp. 31–40.
- [6] 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.
- [7] M. Burger and S. Osher. “Convergence rates of convex variational regularization”. In: *Inverse Problems* 20.5 (2004), pp. 1411–1421.
- [8] E. Candes and M. Wakin. “An Introduction to Compressive Sensing”. In: *IEEE Signal Processing Magazine* 25.2 (2008), pp. 21–30.
- [9] 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.
- [10] 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.
- [11] V. Chandrasekaran et al. “The Convex Geometry of Linear Inverse Problems”. In: *Foundations of Computational Mathematics* 12.6 (2012), pp. 805–849.
- [12] 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.
- [13] J. F. Claerbout and F. Muir. “Robust modeling with erratic data”. In: *Geophysics* 38.5 (1973), pp. 826–844.
- [14] G. Dal Maso. *An Introduction to  $\Gamma$ -convergence*. Vol. 8. Progress in nonlinear differential equations and their applications. Boston, MA: Birkhauser, 1993.
- [15] 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.
- [16] Q. Denoyelle, V. Duval, and G. Peyré. “Asymptotic of Sparse Recovery for Positive Measures”. In: *Preprint HAL* (2015).
- [17] D. L. Donoho. “Superresolution via sparsity constraints”. In: *SIAM J. Math. Anal.* 23.5 (1992), pp. 1309–1331.
- [18] C. Dossal. “A necessary and sufficient condition for exact recovery by  $\ell^1$  minimization”. In: *Comptes Rendus Mathématique* 350.1 (2011), pp. 117–120.
- [19] C. Dossal and S. Mallat. “Sparse spike deconvolution with minimum scale”. In: *Proceedings of SPARS*. Nov. 2005, pp. 123–126.
- [20] C. Dossal et al. “Sharp Support Recovery from Noisy Random Measurements by  $L_1$  minimization”. In: *Applied and Computational Harmonic Analysis* 33.1 (2012), pp. 24–43.
- [21] V. Duval and G. Peyré. “Exact Support Recovery for Sparse Spikes Deconvolution”. In: *to appear in Foundation of Computational Mathematics* (2015).



- [22] C. Ekanadham, D. Tranchina, and E.P. Simoncelli. “A unified framework and method for automatic neural spike identification”. In: *Journal of Neuroscience Methods* 222 (2014), pp. 47–55.
- [23] C. Ekanadham, D. Tranchina, and E.P. Simoncelli. “Recovery of Sparse Translation-Invariant Signals With Continuous Basis Pursuit”. In: *Signal Processing, IEEE Transactions on* 59.10 (2011), pp. 4735–4744.
- [24] 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.
- [25] A. Florescu et al. *A Constrained Optimization Approach For Complex Sparse Perturbed Models*. Tech. rep. Preprint hal-00783298, 2013.
- [26] J.J. Fuchs. “On sparse representations in arbitrary redundant bases”. In: *IEEE Transactions on Information Theory* 50.6 (2004), pp. 1341–1344.
- [27] M. Grasmair, O. Scherzer, and M. Haltmeier. “Necessary and sufficient conditions for linear convergence of  $\ell_1$ -regularization”. In: *Communications on Pure and Applied Mathematics* 64.2 (2011), pp. 161–182.
- [28] 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.
- [29] R. T. Rockafellar, R. J-B. Wets, and M. Wets. *Variational analysis*. Grundlehren der mathematischen Wissenschaften. Berlin, Heidelberg, New York: Springer, 1998.
- [30] G. Still. “Discretization in semi-infinite programming: the rate of convergence”. In: *Mathematical Programming* 91.1 (2001), pp. 53–69.
- [31] G. Tang, B. Narayan Bhaskar, and B. Recht. “Near Minimax Line Spectral Estimation”. In: *CoRR* abs/1303.4348 (2013).
- [32] G. Tang, B. Narayan Bhaskar, and B. Recht. “Sparse recovery over continuous dictionaries-just discretize”. In: *Proc. ACSSC*. IEEE, 2013, pp. 1043–1047.
- [33] 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.
- [34] S. Vaiter, G. Peyré, and J. Fadili. *Model Consistency of Partly Smooth Regularizers*. Tech. rep. Preprint Hal-00987293, 2014.
- [35] S. Vaiter et al. “Model Selection with Low Complexity Priors”. In: *to appear in Information and Inference* (2015).
- [36] M. J. Wainwright. “Sharp Thresholds for High-Dimensional and Noisy Sparsity Recovery Using  $\ell_1$ -Constrained Quadratic Programming (Lasso)”. In: *IEEE Transactions on Information Theory* 55.5 (2009), pp. 2183–2202.
- [37] P. Zhao and B. Yu. “On Model Selection Consistency of Lasso”. In: *J. Mach. Learn. Res.* 7 (Dec. 2006), pp. 2541–2563.